Боян Димитров   Николай Янев

ВЕРОЯТНОСТИ
И
СТАТИСТИКА

СОФТЕХ

София, 2007

В предлагания учебник са въведени основните понятия и факти от теорията на вероятностите и математическата статистика за изграждане на необходимия минимум от знания в тази интензивно развиваща се математическа дисциплина с многобройни и важни приложения в другите науки и практиката. Всяка глава съдържа известен брой задачи за упражнения, а в текста са разгледани достатъчно примери, които улесняват усвояването на материала.

Учебникът може да се използва от студентите и преподавателите във всички висши учебни заведения, където се изучават въпроси от теорията на вероятностите и математическата статистика, и може да обслужва учебни програми с различна степен на трудност по двете дисциплини.

© Боян Недялков Димитров
Николай Михайлов Янев

2007

ISBN 978-954-8495-33-2

СОФТЕХ

СЪДЪРЖАНИЕ

ПРИЛОЖЕНИЯ

ПРЕДГОВОР КЪМ ВТОРОТО ИЗДАНИЕ

От написването на този учебник изминаха около осем години. Основното му ядро остава непроменено, тъй като отразява вече установилите се концепции на съвременните курсове по теория на вероятностите и математическата статистика, приети на въоръжение в почти всички водещи университети по света. За разлика от останалите естествени науки математиката представлява последователно изграждаща се структура, в която новополучените резултати естествено попадат в надстройката, без да засягат аксиоматичните основи на структурата. Бурното развитие на теорията на вероятностите и математическата статистика през последните години очертава още по-ясно тази базисна основа, тази първа стъпка към по-нататъшните многобройни разклонения на теорията и нейните приложения.

Това, което отличава настоящия учебник, е опитът да се съчетаят простотата и достъпността на един първоначален курс със строгостта на математическото изложение, базиращо се на основните математически курсове, изучавани от студентите преди това.

Във второто издание са внесени несъществени изменения, отнасящи се само до стиловото и техническо изложение и до отстраняване на допуснатите грешки в първото издание.

Натрупаният опит от изминалите години показва, че учебникът е подходящ предимно за студентите от факултетите по математика и информатика на университетите в София, Пловдив, Велико Търново, Шумен, Бургас и Благоевград, но той може да бъде използван и от студентите и преподавателите във всички висши учебни заведения, където се изучават въпроси от теорията на вероятностите и математическата статистика.

Авторите считат за приятен дълг да изразят своята благодарност към всички колеги, които посочиха допуснати грешки и неточности в първото издание. Искрено се надяваме, че второто издание ще бъде по-добро от първото, за което сме благодарни и на Университетско издателство „Св. Климент Охридски“.

Февруари 1996 г. *Авторите*

ПРЕДГОВОР КЪМ ПЪРВОТО ИЗДАНИЕ

Стохастиката (от гр. *στὸχοσ* — предполагаемо, зависещо от случая, или *στοχαστικὸζ* — умеещ да отгатва) обединява теорията на вероятностите и математическата статистика, както и техните приложения. Тя е математически фундамент на съвременното естествознание, без който е немислимо изграждането на правилен научен мироглед. Нещо повече, знанията по стохастика могат да ни помогнат да се ориентираме в редица чисто житейски ситуации, а изследването и оценяването на много обществено–икономически и социални процеси е невъзможно без широкото привличане на стохастични методи. Както отбелязват някои изследователи, общата култура на една нация съществено зависи от нейната стохастична култура.

Теорията на вероятностите възниква като математическа дисциплина през XVII в. и е свързана с имената на Паскал, Ферма, Галилей, Хюйгенс, Якоб Бернули. Първоначален подтик дават някои задачи, свързани с хазартните игри, решаването на които не е било възможно в рамките на съществуващата тогава математика.Твърде скоро тези блестящи учени забелязват, че възникналите основни понятия на стохастиката могат да опишат много по-широк кръг случайни явления. Нещо повече, те предвиждат фундаменталния характер на младата наука за бъдещото развитие на естествознанието. Наистина прогресът на науката и техниката е неизменно свързан и с развитието на теорията на вероятностите и математическата статистика. Тези успехи на стохастиката са свързани с имената на Моавър, Лаплас, Гаус, Поасон, Лобачевски, Чебишов, Марков, Ляпунов, Борел, Бернщайн, Леви, Фишер, Хинчин, Винер, Гнеденко, Нейман, Пирсън, Фелър. Сред тази блестяща плеяда особена роля се пада на един от най-забележителните математици Андрей Николаевич Колмогоров (1903–1987), с основание наричан „баща на съвременната стохастика“, който през 1933 г. поставя теорията на вероятностите на строга аксиоматична основа и доказва редица фундаментални резултати. С част от тях читателите ще могат да се запознаят и в настоящия учебник.

Заглавието на книгата съответства на наименованието на курса лекции, четен от авторите на студентите от ФМИ при Софийския университет през последните десет години.

Основното предназначение на този курс и съответно на книгата е да въведе читателя в областта на теорията на вероятностите и математическата статистика — екзистенц–минимум за бъдещите математици и информатици в областта на тази интензивно развиваща се математическа дисциплина с многобройни и важни приложения в другите науки и практиката.

Главната цел на авторите е да развият у своите слушатели и читатели т. нар. вероятностна (стохастична) интуиция, т. е. умението да построяват и изследват адекватни математически модели на реални случайни явления. Ето защо наред със строгото аксиоматично изложение на теорията те често прибягват до нестроги „евристични“ обяснения, които, от една страна, посочват произхода на математическите понятия, а, от друга, дават възможност за различни интерпретации на получените математически резултати.

Книгата е естествено разделена на две части: Теория на вероятностите (първа – девета глава) и Математическа статистика (десета – петнадесета глава), написани съответно от Н. М. Янев и Б. Н. Димитров в тясно сътрудничество. Във всяка глава са включени задачи за упражнения, а освен това в текста има доста решени примери, което се надяваме да улесни усвояването на материала. Читателят трябва да реши допълнително и немалък брой задачи, за което препоръчваме съответните Ръководства за упражнения по теория на вероятностите [26] и по математическа статистика [30], които са съобразени с настоящия курс.

Формулите (теоремите, дефинициите, примерите и т. н.) се цитират със съответните номера, към които се добавя отпред номерът на параграфа при цитиране в друг параграф от същата глава. При цитиране извън съответната глава се получава тройна номерация. Така например (3.4.7) означава формула (7) от § 4 на трета глава; теорема 4.5.1 означава теорема 1 от § 5 на четвърта глава; задача 9.6 означава задача 6 от девета глава и т. н.

Авторите считат за приятен дълг да благодарят за полезните съвети на колегите от сектор „Вероятности и статистика“, Лаборатория по компютърна стохастика и Лаборатория по статистически контрол на качеството. Особено ни е приятно да отбележим дейното участие на по-младите колеги Ива Цанкова, Маруся

Славчова, Марияна Белева, Райна Робева, Сахиб Иса, Георги Янев, Борис Ковачев и Иво Гавазки в окончателното оформяне на ръкописа.

Рецензентите Цветан Игнатов и Георги Чобанов направиха ценни препоръки, за които авторите са им много благодарни. Особено признателни сме на Росица Додунекова за добрата ѝ редакторска работа.

Януари 1988 г. *Авторите*

П ъ р в а ч а с т

ТЕОРИЯ НА ВЕРОЯТНОСТИТЕ

П ъ р в а г л а в а

Вероятностни пространства

§ 1. Емпирични основи

Предмет на теорията на вероятностите е математическият анализ на понятието случайност. При това, докато в ежедневните представи, в житейската практика се счита, че случайните явления или събития представляват нещо крайно рядко, което не е в рамките на обикновения ред, при „закономерното“ развитие на събитията, в теорията на вероятностите ще се откажем от тези представи. Случайните събития като основни понятия в теорията на вероятностите имат редица характерни свойства.

Ето какво казва по този повод А. Н. Колмогоров [13]:

„Най-простите закономерности, които формира естествознанието, се състоят в посочването на условията, при които някакво интересуващо ни събитие се сбъдва или не се сбъдва, т.е. това може да бъде изразено по една от следните две схеми:

1) при всяко осъществяване на комплекса (т.е. съвкупността) от условия S се сбъдва събитието A;

2) при осъществяването на комплекса от условия S събитието A не се сбъдва.

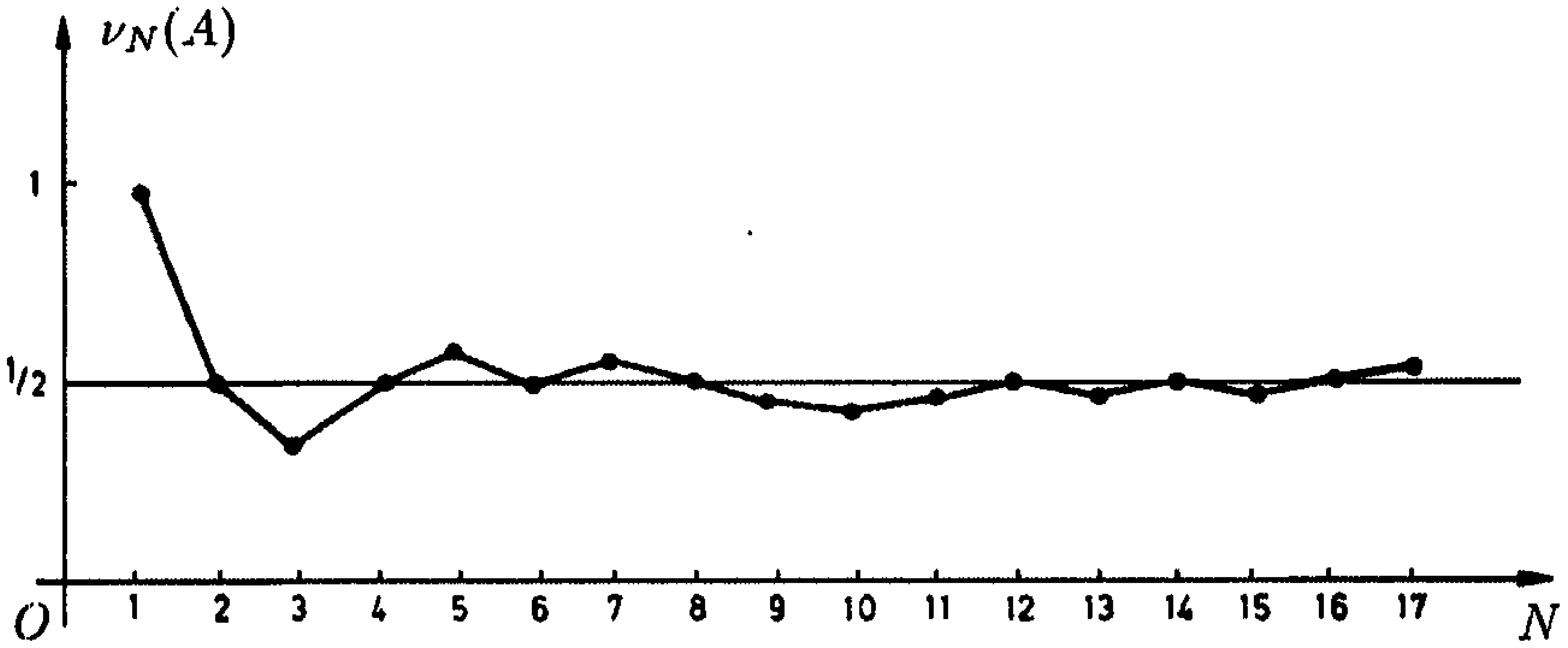
В първия случай събитието A се нарича „сигурно (достоверно)“ или „необходимо“ събитие по отношение на комплекса S, а във втория — „невъзможно“ събитие. Събитие A, което при реализиране на комплекса от условия S понякога се сбъдва, а понякога не се сбъдва по отношение на този комплекс от условия, се нарича „случайно“.

Например при химични реакции без обмен с околната среда (комплекс S) общата маса на реагиралите вещества остава неизменна (събитие A). Друг пример: при температура $t^\circ \geqq 100^\circ$C и атмосферно налягане $P = 760$ mm (комплекс S) водата кипи, т.е. превръща се в пара (сигурно събитие A), и не може да се намира в твърдо състояние (невъзможно събитие A).

В своята практическа дейност човекът на всяка крачка среща случайни явления. Нещо повече, без тях не протича нито един реален процес. Случайни явления например се проявяват при всеки процес на измерване. При това колкото измервателният уред е по-точен, толкова „случайните грешки" при измерване са по-ясно забележими. Като претегляме един и същ предмет на аналитични везни много пъти, винаги получаваме близки, но различни резултати. Примери на случайни явления представляват отказите на различни технически устройства, разсейването на снарядите при стрелба, заразяването от грип, радиоактивното разпадане, космическите лъчи, шумовете при радиоприемане, състоянието на атмосферата, демографските процеси, лотариите, хазартните игри и т.н. Въобще от гледна точка на съвременната атомно-молекулна теория всяко вещество се състои от огромен брой елементарни частици, които се намират в непрекъснато движение и взаимодействия със случаен характер. В този смисъл може да се каже, че теорията на вероятностите като единствена наука, занимаваща се със случайностите, е фундаментален математически апарат на съвременното естествознание.

От гледна точка на познанието простото утвърждаване на случайността на дадено събитие представлява ограничен интерес. Обаче за широк кръг от явления при многократно осъществяване на даден опит (експеримент, комплекс от условия S) броят на случаите, при които се сбъдва (реализира) събитието A (често се означава $\#(A)$), разделен на броя на всички случаи N (т.е. *честотата* на това събитие $\nu(A) = \#(A)/N$), малко се отклонява от (колебае се около, стреми се към) някакво средно число. Това число може да служи за количествена оценка на случайността на това събитие, която бихме могли да наречем вероятност на събитието A (естествено в съответния математически модел). И така, теорията на вероятностите изследва не всички случайни събития, а предимно масовите случайни събития, които се подчиняват на вероятностни (стохастични) закономерности, т.е. предполага се, че могат да се създадат много пъти едни и същи условия (да се реализират експерименти), при които може да се сбъдне или не някакво случайно събитие A.

Например можем много пъти да подхвърляме една и съща монета, като всеки път се осъществява едно от двете събития $A = \{\text{Герб}\}$ и $\overline{A} = \{\text{Лице}\}$ (т.е. отхвърляме възможността монетата да застане на ръба си). На фиг. 1 е дадена графиката на честотата $\nu(A) = \#(A)/N$ за редицата от изходи ГЛЛГГЛГЛЛЛГГЛГЛ, т.е. начупената линия, съединяваща точките $(k, \nu(A))$, която с нарастването на N бързо се приближава към правата $\nu(A) = 1/2$.



Фиг. 1

За да провери този факт експериментално, Бюфон е извършил 4040 хвърляния на монета, от които 2048 са били „герб“, откъдето е получил честотата $\nu(A) = 0,5080$. Пирсън е хвърлял монета 24 000 пъти, при което са се паднали 12 012 герба, т.е. $\nu(A) = 0,5005$. Така в случай на „правилна“ (т.е. симетрична и от хомогенен материал) монета е естествено да приемем вероятността на събитието A (падане на „герб“) за равна на 1/2. До аналогични изводи можем да стигнем, ако разглеждаме статистиката на ражданията, която показва, че наблюдаваната честота на ражданията за момчетата е между 0,51 и 0,52 (колебае се за различни периоди и различни страни).

Устойчивостта на честотите е обективно свойство на масовите случайни явления в реалния свят. Устойчивите честоти ни дават представа какви свойства трябва да има аксиоматично дефинираната вероятност в съответния математически модел.

Теорията на вероятностите е математическа наука, която изучава математически модели на случайните явления и процеси. В нея както във всяка математическа дисциплина се изхожда от някои основни понятия (т.е. математически обекти), които не подлежат на дефиниция, а се основават на знания за действителността. Те се явяват синтез на нашия житейски опит, т.е. дефинират се аксиоматично. От аксиоматична гледна точка събитията са математически обекти, които могат да се комбинират с помощта на логическите операции „и“, „или“, „не“ в съответствие с дадени правила, а вероятността е числова характеристика на събитията от даден клас, чиито свойства по дефиниция са аналогични на свойствата на честотите.

§ 2. Основно пространство и алгебра на събитията

Първично понятие в аксиоматичното изграждане на теорията на вероятностите е понятието *елементарно събитие.* Понякога елементарните събития се интерпретират като възможните изходи (резултати) на един „случаен опит“ (експеримент). Тук понятието „случаен опит“ е спомагателно и служи само като посредник между абстрактния математически модел и реалната действителност. Считаме, че опитът е напълно определен, ако е дадено множеството от неговите възможни резултати.

Съвкупността на елементарните събития (за даден опит) наричаме *основно пространство* или *сигурно* (*достоверно*) *събитие.* По-нататък винаги ще изхождаме от непразно множество $\Omega = \{\omega\}$, елементите на което ще считаме за елементарни събития.

Пример 1 (хвърляне на монета). В този случай основното пространство $\Omega = \{\omega_1, \omega_2\}$ съдържа две елементарни събития $\omega_1 =$ „Герб“ и $\omega_2 =$ „Лице“. За нас е съществено, че резултатите са само два и че при всеки опит се реализира точно един от тях.

Пример 2 (хвърляне на зар). В този случай

$$\Omega = \{1,\ 2,\ 3,\ 4,\ 5,\ 6\}.$$

Пример 3 (хвърляне на зар, докато се падне шестица). Сега основното пространство съдържа изброимо много елементарни събития

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_n, \ldots\},$$

където $\omega_n = \{$да се падне шестица за първи път на n-тия опит$\}$.

Пример 4 (хвърляне на две монети). В този случай основното пространство $\Omega = \{(Л,Л),(Л,Г),(Г,Л),(Г,Г)\}$ съдържа четири елементарни събития.

Пример 5 (стрелба по кръгова мишена). Елементарното събитие е всяка точка ω от кръга и основното пространство от елементарни събития $\Omega = \{\omega\}$ съдържа неизброимо много елементарни събития.

Пример 6 (осцилограф). В този случай

$$\Omega = \{f(x) : a \leqq f(x) \leqq b,\ x \in [c,d]\}$$

съдържа неизброимо много елементарни събития и всяко от тях е някаква ограничена и непрекъсната функция $f(x)$ в краен интервал $[c,d]$.

З а б е л е ж к а. Опитът въобще не дефинира еднозначно основното пространство Ω. Понякога на един опит могат да се съпоставят различни основни пространства.

Пример 7. При хвърляне на два зара можем да разглеждаме следните основни пространства:

$$\Omega_1 = \{\text{ненаредените двойки } (i,k);\quad i,k = 1,\dots,6\},$$

$$\Omega_2 = \{\text{наредените двойки } (i,j);\quad 1 \leqq i \leqq j \leqq 6\},$$

$$\Omega_3 = \{\text{сумата от точките } i+k;\quad i,k = 1,\dots,6\}.$$

Да отбележим, че пространствата имат различен брой елементи $\#(\Omega_1) = 36$, $\#(\Omega_2) = 21$, $\#(\Omega_3) = 11$.

Този пример не бива да ни смущава, тъй като теорията на вероятностите „започва" всъщност едва след като сме фиксирали съответното основно пространство от елементарни събития. Сега вече можем да въведем следващото важно понятие — *събитие.*

Най-общо казано, *всяко събитие е множество от елементарни събития.* Различните събития са различни множества от елементарни събития.

Ако елементарното събитие ω се сбъдне и $\omega \in A \subseteq \Omega$, ще казваме, че се е сбъднало събитието A (или че е настъпило събитието A). Същевременно казваме, че елементарният изход ω е благоприятен за събитието A.

В условията на пример 4 нека:

$$A = \{\text{да се падне поне едно лице}\},$$

$$B = \{\text{да се падне точно един герб}\},$$

$$C = \{\text{да се паднат две лица}\} = \{\text{да не се падне герб}\},$$

$$D = \{\text{да се паднат поне четири герба}\}.$$

Тогава горните събития съдържат следните елементарни събития: $A = \{(Г,Л),(Л,Г),(Л,Л)\}$, $B = \{(Г,Л),(Л,Г)\}$, $C = \{(Л,Л)\}$, а D не съдържа нито едно елементарно събитие.

Последното събитие се нарича ***невъзможно събитие*** и се бележи с $\varnothing$. Това е такова събитие, което не може да се реализира при дадения експеримент.

Тъй като за всяко ω имаме $\omega \in \Omega$, то Ω се сбъдва при всеки експеримент и затова естествено се нарича *сигурно събитие.*

Нека в резултат на експеримента в пример 4 е настъпило елементарното събитие $\omega = (Л,Г)$. Тогава очевидно се сбъдват събитията A и B, но не се сбъдва C. В този случай казваме, че е настъпило *допълнителното* (*противоположното*) на C събитие $\overline{C}$.

И така по дефиниция $\omega \in \overline{C}$ тогава и само тогава, когато $\omega \notin C$.

Очевидно $\overline{\Omega} = \varnothing$ и $\overline{\varnothing} = \Omega$.

Обединение (или *сума*) на две събития A и B наричаме събитието $H = A \cup B$, което се сбъдва само когато се сбъдва или A, или B, или двете събития едновременно, т.е. H съдържа елементарните събития, които принадлежат на поне едно от събитията A или B.

Сечение на събитията A и B наричаме събитие $G = A \cap B$ (бележи се още AB), което се сбъдва само когато се сбъдват и A, и B едновременно, т.е. G съдържа елементарните събития, общи за A и B.

Събитието $A \backslash B = A \cap \overline{B}$ се нарича *разлика* на A и B и настъпва само когато настъпва A, но не настъпва B.

Събитията A и B се наричат ***несъвместими***, ако $A \cap B = \varnothing$. В този случай вместо $A \cup B$ често се използва означението $A + B$.

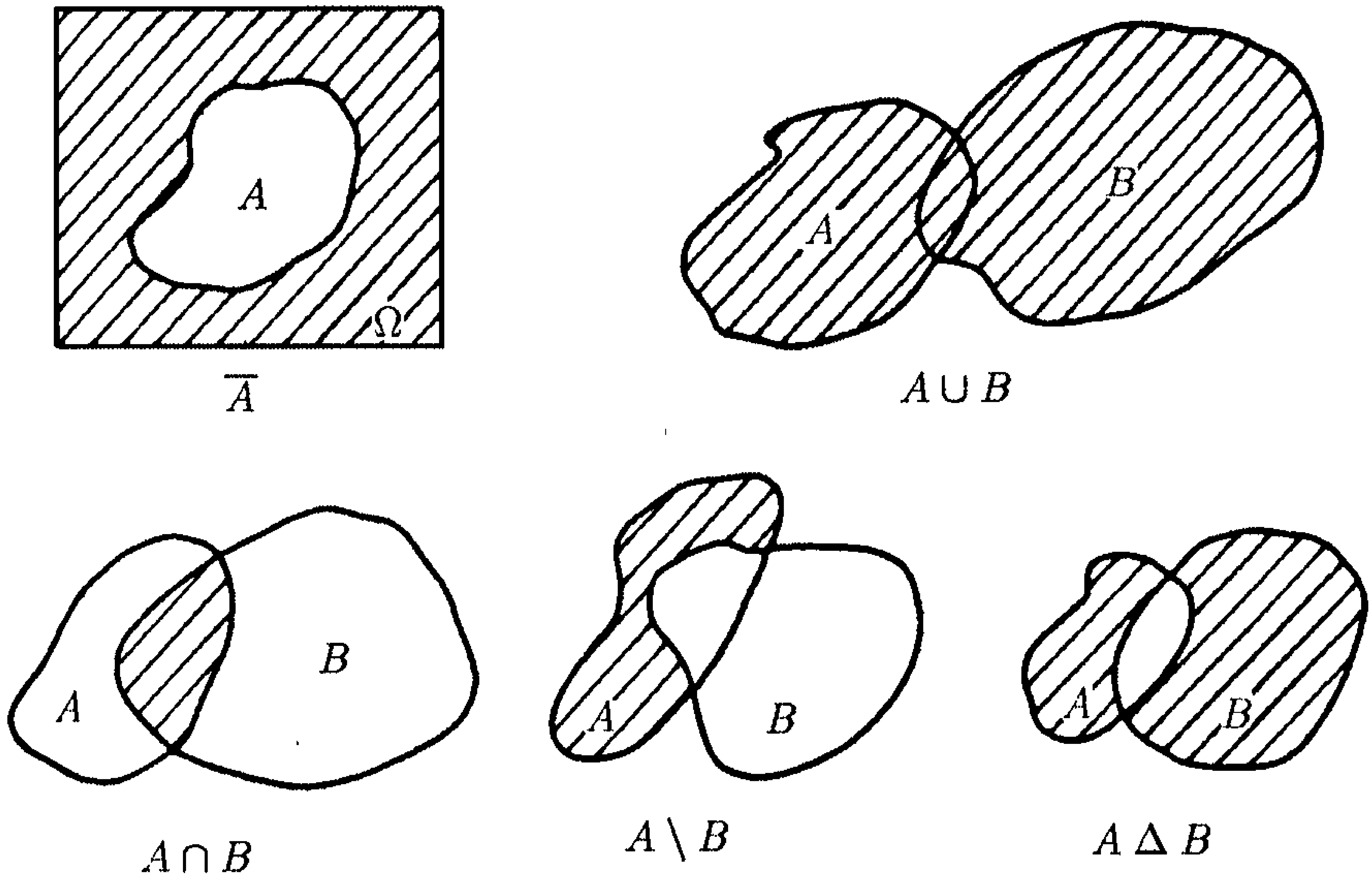
Казваме, че събитието A *влече* събитието B (бележим $A \subset B$ или $B \supset A$), ако $A = A \cap B$ (или равносилно $B = A \cup B$), т.е. ако при сбъдването на събитието A се сбъдва и събитието B. Казва се още, че B ***следва*** от A.

Ако $A \subset B$ и $B \subset A$, казваме, че събитията A и B са ***еквивалентни*** и означаваме $A = B$.

Симетрична разлика $A \Delta B = A\overline{B} + \overline{A}B$ е събитието, което се сбъдва само когато настъпва точно едно от събитията A или B.

За въведените събития от пример 4 не е трудно да се провери, че $B \subset A$, $B \cup C = A$, $A \backslash C = B$, $B \cap C = \varnothing$, така че $A = B + C$.

Така дефинираните операции над събития са илюстрирани на фиг. 2, където резултатът от съответната операция е защрихован.



Фиг. 2

По-нататък с $\mathfrak{B}(\Omega)$ ще означаваме съвкупността от всички подмножества на Ω.

Операциите обединение и сечение на събития се разпространяват за всяко крайно или безкрайно множество от събития $\{A_k\}$, където обединението $\bigcup_k A_k$ означава сбъдването на поне едно от събитията, а сечението $\bigcap_k A_k$ означава едновременното сбъдване на всички събития. Обичайните свойства на операциите над множества се пренасят без изменение върху операциите над събития. Например в сила са законите на де Морган:

$$\overline{\bigcup_k A_k} = \bigcap_k \overline{A_k}, \qquad \overline{\bigcap_k A_k} = \bigcup_k \overline{A_k}. \tag{1}$$

Често се използват означенията

$$\sup_{k \geqq n} A_k = \bigcup_{k \geqq n} A_k, \qquad \inf_{k \geqq n} A_k = \bigcap_{k \geqq n} A_k. \tag{2}$$

Ще докажем (1), като символът за еквивалентност $\Longleftrightarrow$ ще означава „тогава и само тогава, когато“. Наистина

$$\omega \in \overline{\bigcup_k A_k} \Longleftrightarrow \omega \notin \bigcup_k A_k \Longleftrightarrow \omega \notin A_k,\ k = 1,2,\dots$$
$$\Longleftrightarrow \omega \in \overline{A_k},\ k = 1,2,\dots \Longleftrightarrow \omega \in \bigcap_k \overline{A_k}.$$

Второто равенство се доказва аналогично.

Дефиниция 1. *Нека $\mathfrak{F}$ е система от подмножества на Ω (т.е $\mathfrak{F} \subset \mathfrak{B}(\Omega)$) със следните свойства:*

1) $\Omega \in \mathfrak{F}$, $\varnothing \in \mathfrak{F}$;

2) *ако $A \in \mathfrak{F}$, то $\overline{A} \in \mathfrak{F}$;*

3) *ако $A, B \in \mathfrak{F}$, то $A \cup B \in \mathfrak{F}$ и $A \cap B \in \mathfrak{F}$.*

Тогава $\mathfrak{F}$ се нарича булова алгебра (б.а.), а нейните елементи — случайни събития.

От тази дефиниция веднага следва, че ако $A_k \in \mathfrak{F}$, $k = 1,\dots,n$, то $\bigcup_k A_k \in \mathfrak{F}$ и $\bigcap_k A_k \in \mathfrak{F}$.

Това дава основание да казваме, че *булевата алгебра $\mathfrak{F}$ е устойчива относно операциите допълнение, крайни обединения и сечения.*

Дефиниция 2. *Булова алгебра $\mathfrak{F}$, която е устойчива относно изброимите обединения и сечения, се нарича σ-булова алгебра, т.е. от $A_k \in \mathfrak{F}$, $k = 1,2,\dots$, следва $\bigcup_k A_k \in \mathfrak{F}$ и $\bigcap_k A_k \in \mathfrak{F}$.*

Като използваме тъждествата на де Морган (1), лесно виждаме, че в последната дефиниция е достатъчно да поискаме устойчивост относно едната от операциите обединение или сечение.

Най-прости примери на булови алгебри са тривиалната булова алгебра $\mathfrak{F}_0 = \{\varnothing, \Omega\}$ и $\mathfrak{B}(\Omega)$ — множеството от всички подмножества на Ω. В частност, ако $\#(\Omega) = n$, то $\#(\mathfrak{B}(\Omega)) = 2^n$. За всяка булова алгебра от подмножества на Ω имаме $\mathfrak{F}_0 \subset \mathfrak{F} \subset \mathfrak{B}(\Omega)$.

Ако $A \subset \Omega$, то $\mathfrak{F}_1 = \{\varnothing, \Omega, \overline{A}, A\}$ е булова алгебра. Аналогично, ако $A \subset \Omega$ и $B \subset \Omega$, то $\mathfrak{F}_2 = \{\varnothing, \Omega, A, B, \overline{A}, \overline{B}, A \cup B, A \cap B, \overline{A} \cup \overline{B}, \overline{A} \cap \overline{B}, \overline{A} \cup B, \overline{A} \cap B, A \cup \overline{B}, A \cap \overline{B}, (A \cap B) \cup (\overline{A} \cap \overline{B}), (A \cap \overline{B}) \cup (\overline{A} \cap B)\}$ е булова алгебра.

Не е трудно да се провери, че ако $A_k \subset \Omega$, $k = 1,\dots,n$, то класът $\mathfrak{F}_n$, който съдържа $\varnothing, \Omega$, множествата A_k, техните допълнения и всевъзможните обединения и сечения на тези множества, е също булова алгебра. При това $\mathfrak{F}_0 \subset \mathfrak{F}_i \subset \mathfrak{F}_{i+1} \subset \mathfrak{B}(\Omega)$ за $i = 1,\dots,n-1$.

Дефиниция 3. *Двойката* $(\Omega, \mathfrak{F})$, *където* $\mathfrak{F}$ *е булова алгебра от подмножества на* Ω, *се нарича измеримо пространство.*

Да отбележим, че Ω е съвкупността от елементарните събития, докато $\mathfrak{F}$ съдържа случайните събития.

Забележка. Елементарните събития могат да не са случайни събития!

Пример 8. Нека Ω е от пример 2 и A означава събитието „да се падне четно число". Очевидно $\overline{A} = \{$да се падне нечетно число$\}$. Както бе отбелязано по-горе, $\mathfrak{F}_1 = \{\varnothing, \Omega, A, \overline{A}\}$ е булова алгебра. Тогава в измеримото пространство $(\Omega, \mathfrak{F}_1)$ елементарните събития не са елементи на $\mathfrak{F}_1$, т.е. не са събития.

Нека $\{A_n\}$ е редица от събития в σ-алгебрата $\mathfrak{F}$. Тогава събитието

$$(3)\quad A_* = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m = \sup_{n \geqq 1}(\inf_{m \geqq n} A_m) = \lim_{n\to\infty} \uparrow \inf_{m \geqq n} A_m = \liminf A_n$$

$$= \underline{\lim} A_n$$

се нарича *долна граница* на редицата $\{A_n\}$, а събитието

$$(4)\quad A^* = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m = \inf_{n \geqq 1}(\sup_{m \geqq n} A_m) = \lim_{n\to\infty} \downarrow \sup_{m \geqq n} A_m = \limsup A_n$$

$$= \overline{\lim} A_n$$

— *горна граница* на редицата $\{A_n\}$.

Да отбележим, че в (3) и (4) сме използвали означенията (2). Множеството A_* може да се интерпретира като събитие, състоящо се в това, да се сбъднат всички събития на редицата $\{A_n\}$ от известно място нататък. Наистина от (3) следва, че ако $\omega \in A_*$, то съществува такова N, че $\omega \in \bigcap_{m=N}^{\infty} A_m$, т.е. $\omega \in A_m$ за всички $m \geqq N$; следователно ω не принадлежи евентуално само на краен брой събития от редицата $\{A_n\}$.

Ако $\omega \in A^*$, от (4) следва, че $\omega \in \bigcup_{m=n}^{\infty} A_m$ за всяко $n \geqq 1$, т.е. ω принадлежи на безбройно много събития A_n. Следователно множеството A^* може да се интерпретира като събитие, при което се сбъдват безбройно много събития от редицата $\{A_n\}$. Очевидно $A_* \subset A^*$.

Дефиниция 4. *Ако* $A_* = A^*$, *казваме, че редицата* $\{A_n\}$ *има граница* $\lim A_n = \liminf A_n = \limsup A_n$ *при* $n \to \infty$.

Нека например $\{A_n\}$ е монотонно намаляваща редица от събития, т.е. $A_1 \supset A_2 \supset \ldots \supset A_n \supset A_{n+1} \supset \ldots$ (често това се означава с $A_n \downarrow$). Тогава от (3) и (4) следва, че

$$\liminf A_n = \limsup A_n = \bigcap_{n=1}^{\infty} A_n,$$

т.е. редицата $\{A_n\}$ има граница и $\lim\limits_{n\to\infty} A_n = \bigcap\limits_{n=1}^{\infty} A_n$.

Ако $\{A_n\}$ е монотонно растяща редица от събития, т.е. $A_1 \subset A_2 \subset \ldots \subset A_n \subset A_{n+1} \subset \ldots$ (означава се накратко с $A_n \uparrow$), от (3) и (4) получаваме

$$\liminf A_n = \limsup A_n = \bigcup_{n=1}^{\infty} A_n.$$

Следователно редицата $\{A_n\}$ има граница и $\lim\limits_{n\to\infty} A_n = \bigcup\limits_{n=1}^{\infty} A_n$.

По-нататък често ще използваме представянето

$$\bigcup_{k=1}^{n} A_k = \sum_{k=1}^{n} \left[A_k \setminus \bigcup_{i=1}^{k-1} A_i \right], \quad n \geqq 1, \tag{5}$$

където знакът $\sum$ означава обединение на несъвместими събития, а $\bigcup\limits_{i=1}^{0} \equiv \varnothing$. При $n = 1$ формула (5) е очевидна, а при $n = 2$ равенството

$$A_1 \bigcup A_2 = A_1 + (A_2 \setminus A_1) \tag{6}$$

следва непосредствено от дефинициите (вж. фиг. 2).

Да предположим, че (5) е вярно за някакво $n \geqq 2$. Полагаме $A = \bigcup\limits_{k=1}^{n} A_k$, $B = A_{n+1}$. Тогава, като приложим (6) към $A \cup B$, намираме, че (5) е вярно и за $n + 1$.

§ 3. Вероятност. Основни свойства

Нека $(\Omega, \mathfrak{F})$ е измеримо пространство, т.е. Ω е основно пространство на елементарните събития, а $\mathfrak{F}$ — булова алгебра от негови подмножества (които нарекохме случайни събития).

Вероятност **P** в измеримото пространство $(\Omega, \mathfrak{F})$ се нарича всяка числова функция, дефинирана върху събитията от $\mathfrak{F}$, която удовлетворява следните аксиоми:

1°. $\mathbf{P}(A) \geqq 0$ *за всички* $A \in \mathfrak{F}$ (***неотрицателност***);

2°. $\mathbf{P}(\Omega) = 1$ (***нормираност***);

3°. $\mathbf{P}(A + B) = \mathbf{P}(A) + \mathbf{P}(B)$, *когато* $AB = \varnothing$ (***адитивност***);

4°. *За всяка редица от свбития* $\{A_n\}$, *за която* $A_n \downarrow \varnothing$ (*т.е.* $A_1 \supset A_2 \supset \dots$ *и* $\bigcap\limits_{n=1}^{\infty} A_n = \varnothing$), *е в сила* $\lim\limits_{n\to\infty} \mathbf{P}(A_n) = 0$ (***непрекъснатост***).

Тогава тройката $(\Omega, \mathfrak{F}, \mathbf{P})$ се нарича *вероятностно пространство.*

З а б е л е ж к а. От аксиомите **1°**– **4°** се вижда, че вероятността е функция от множества, чиито стойности са реални числа. Такива функции (изображения) е прието да се наричат *мерки.* Примери за мерки са дължината, лицето и обемът. В нашия случай можем да кажем, че *вероятността е неотрицателна, нормирана, адитивна и непреквсната мярка.*

Казваме, че е построен *вероятностен* (*стохастичен*) *модел* на даден експеримент, ако на него му е съпоставено вероятностно пространство. На всеки реален експеримент могат да бъдат съпоставени различни вероятностни пространства. В това ще имаме възможност да се убедим нееднократно.

Една от основните цели на стохастиката е построяване и изследване на адекватни стохастични модели.

Задачата за адекватност (т.е. съответствие) на математическия модел с реалната действителност се третира от математическата статистика, която се разглежда във втората част на книгата.

Следващите елементарни свойства на вероятностите следват непосредствено от аксиомите:

1) $\mathbf{P}(\varnothing) = 0$.

Действително, от $\varnothing \cup \varnothing = \varnothing$ и $\varnothing \cap \varnothing = \varnothing$ по аксиома **3°** намираме $\mathbf{P}(\varnothing) = \mathbf{P}(\varnothing) + \mathbf{P}(\varnothing)$, откъдето следва 1).

2) Ако $A \subset B$, то $\mathbf{P}(A) \leqq \mathbf{P}(B)$ (***монотонност***). Освен това $\mathbf{P}(B \backslash A) = \mathbf{P}(B) - \mathbf{P}(A)$.

Тъй като $B = A + \overline{A}B$, от аксиома **3°** имаме $\mathbf{P}(B) = \mathbf{P}(A) + \mathbf{P}(\overline{A}B)$, а от аксиома **1°** следва $\mathbf{P}(\overline{A}B) \geqq 0$, откъдето получаваме 2).

3) $0 \leqq \mathbf{P}(A) \leqq 1$ за всяко $A \in \mathfrak{F}$.

Свойството следва непосредствено от 2), тъй като $A \subset \Omega$.

4) $\mathbf{P}(\overline{A}) = 1 - \mathbf{P}(A)$ (***допълнителност***).

Това следва от аксиоми **3°** и **2°**, понеже $A\overline{A} = \varnothing$ и $A + \overline{A} = \Omega$.

5) $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(AB)$ (***силна адитивност***).

От представянето $A \cup B = A + (B \backslash AB)$ по аксиома **3°** се получава $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B \backslash AB)$. От свойство 2) имаме $\mathbf{P}(B \backslash AB) = \mathbf{P}(B) - \mathbf{P}(AB)$ и след заместване в предишната формула следва 5).

6) Ако $A_i A_j = \varnothing$ за всички $i \neq j$, то $\mathbf{P}\left(\sum\limits_{k=1}^{n} A_k\right) = \sum\limits_{k=1}^{n} \mathbf{P}(A_k)$ (***крайна адитивност***).

Доказва се по индукция от аксиома **3°**.

7) $\mathbf{P}\left(\bigcup\limits_{k=1}^{n} A_k\right) \leqq \sum\limits_{k=1}^{n} \mathbf{P}(A_k)$ (***полуадитивност***).

От равенство (2.5) получаваме $\mathbf{P}\left(\bigcup\limits_{k=1}^{n} A_k\right) = \mathbf{P}\left(\sum\limits_{k=1}^{n} B_k\right)$, където $B_k = A_k \backslash \bigcup\limits_{i=1}^{k-1} A_i \subset A_k$. Оттук с помощта на 6) и 2) намираме 7).

8) Нека $A_n \in \mathfrak{F}$ и $A_n \uparrow A$ (или $A_n \downarrow A$). Тогава $\mathbf{P}(A_n) \uparrow \mathbf{P}(A)$ (респективно $\mathbf{P}(A_n) \downarrow \mathbf{P}(A)$).

От $A_1 \subset A_2 \subset \ldots \subset A_n \subset \ldots \subset A$ следва, че $B_n = A \backslash A_n \downarrow \varnothing$, откъдето съгласно **4°** е изпълнено $\mathbf{P}(B_n) \downarrow 0$. Но от свойство 2) имаме $\mathbf{P}(B_n) = \mathbf{P}(A) - \mathbf{P}(A_n)$, откъдето $\mathbf{P}(A_n) \uparrow \mathbf{P}(A)$.

За случая $A_n \downarrow A$ (т.е. $A_1 \supset A_2 \supset \ldots \supset A_n \supset \ldots \supset A$) докажете сами, че $\mathbf{P}(A_n) \downarrow \mathbf{P}(A)$.

Аксиомите **3°** и **4°** могат да бъдат заменени със следната аксиома:

3*. *Ако събитията от редицата $\{A_n\}$ са две по две несъвместими (т.е. $A_i A_j = \varnothing, i \neq j$) и $\sum\limits_{n=1}^{\infty} A_n \in \mathfrak{F}$, то* $\mathbf{P}\left(\sum\limits_{n=1}^{\infty} A_n\right) = \sum\limits_{n=1}^{\infty} \mathbf{P}(A_n)$ (σ-***адитивност***).

По-точно в сила е следният резултат:

Теорема 1. *Системата аксиоми* **1°**, **2°**, **3°** *и* **4°** *е еквивалентна на системата аксиоми* **1°**, **2°** *и* **3***

Доказателство. Нека са в сила аксиомите **1°**, **2°**, **3°** и **4°** и нека $\{A_n\}$ е редица от две по две несъвместими събития, за които $\sum\limits_{n=1}^{\infty} A_n \in \mathfrak{F}$. Тогава $B_n = \sum\limits_{k=1}^{n} A_k$, $B = \sum\limits_{k=1}^{\infty} A_k$, $B_n \uparrow B$ и от свойствата 8) и 6) намираме

$$\mathbf{P}(B) = \mathbf{P}\left(\sum_{k=1}^{\infty} A_k\right) = \lim_{n\to\infty} \uparrow \mathbf{P}(B_n) = \lim_{n\to\infty} \sum_{k=1}^{n} \mathbf{P}(A_k) = \sum_{k=1}^{\infty} \mathbf{P}(A_k),$$

което доказва σ-адитивността.

Нека сега са в сила аксиомите **1°**, **2°**, **3*** и нека $A_n \downarrow \varnothing$. От представянето $A_n = \sum\limits_{k=n}^{\infty} A_k \backslash A_{k+1}$ по аксиома **3*** получаваме

$$\mathbf{P}(A_n) = \sum_{k=n}^{\infty} \mathbf{P}(A_k \backslash A_{k+1}).$$

Тъй като редът $\mathbf{P}(A_1) = \sum\limits_{k=1}^{\infty} \mathbf{P}(A_k \backslash A_{k+1})$ е сходящ, то

$$\mathbf{P}(A_n) = \sum_{k=n}^{\infty} \mathbf{P}(A_k \backslash A_{k+1}) \downarrow 0 \qquad \text{при } n \to \infty$$

като остатък от сходящ ред с неотрицателни членове.

Теоремата е доказана.

9) Ако за всяка редица от събития $\{A_n\}$ имаме $\bigcup\limits_{n=1}^{\infty} A_n \in \mathfrak{F}$, то

$$\mathbf{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \leqq \sum_{n=1}^{\infty} \mathbf{P}(A_n) \qquad (\sigma - \textit{\textbf{полуадитивност}}).$$

Следва от свойство 7), като се използва граничен преход.

10) За всяка редица от събития $\{A_n\}$ имаме

$$\mathbf{P}(\liminf A_n) \leqq \liminf \mathbf{P}(A_n) \leqq \limsup \mathbf{P}(A_n) \leqq \mathbf{P}(\limsup A_n).$$

От (2.3), (2.4), 8) и 2) получаваме

$$\mathbf{P}(\liminf A_n) = \lim_{n\to\infty} \mathbf{P}\left(\inf_{m \geqq n} A_m\right) \leqq \liminf \mathbf{P}(A_n),$$

$$\mathbf{P}(\limsup A_n) = \lim_{n\to\infty} \mathbf{P}\left(\sup_{m \geqq n} A_m\right) \geqq \limsup \mathbf{P}(A_n),$$

където $\liminf \mathbf{P}(A_n)$ и $\limsup \mathbf{P}(A_n)$ са съответно най-лявата и най-дясната точка на сгъстяване на редицата $\{\mathbf{P}(A_n)\}$.

Системата аксиоми **1°**, **2°**, **3°** и **4°** (респективно **1°**, **2°**, **3***) дефинира вероятностна мярка върху буловата алгебра $\mathfrak{F}$ от подмножества на основното пространство Ω. Тази аксиоматика е предложена от А. Н. Колмогоров.

Нека $\mathfrak{J}$ е клас от подмножества на Ω, т.е. $\mathfrak{J} \subseteq \mathfrak{B}(\Omega)$. Ще казваме, че $\mathfrak{F}$ е *минимална алгебра* (σ*—алгебра*), *съдържаща* $\mathfrak{J}$, ако за всяка алгебра (σ-алгебра) G, за която $\mathfrak{J} \subseteq G$, следва $\mathfrak{F} \subseteq G$. Минималната алгебра (σ-алгебра), съдържаща даден клас, се нарича също алгебра (σ-алгебра), породена от този клас, и се означава $\mathfrak{F} = \sigma(\mathfrak{J})$. Изобщо алгебрата ($\sigma$-алгебрата), породена от даден клас, е сечение на всички алгебри (σ-алгебри), съдържащи този клас.

Дефиниция 1. *Нека $\Omega = \mathbf{R}^1$, където $\mathbf{R}^1$ е реалната права, а $\mathfrak{J}$ е класът на интервалите от вида $[a, b)$. Минималната σ-алгебра $\mathfrak{B}_1$, породена от $\mathfrak{J}$, се нарича σ-алгебра на бореловите множества в $\mathbf{R}^1$ (или накратко борелова σ-алгебра $\mathfrak{B}_1 = \sigma(\mathfrak{J})$).*

Едноточковите множества $\{a\}$ са борелови, тъй като

$$\{a\} = \bigcap_{n=1}^{\infty} \left[a, a + \frac{1}{n}\right).$$

Интервалите $(a, b) = [a, b) \setminus \{a\}$ са също борелови множества (събития). Въобще всяко отворено множество е борелово, понеже може да се представи като обединение на изброимо много интервали. Всяко затворено множество е също борелово (тъй като неговото допълнение е отворено).

Дефиниция 2. *Нека $\Omega = \mathbf{R}^n$, а $\mathfrak{J}$ е класът на паралелепипедите от вида*

$$\prod_{i=1}^{n} [a_i, b_i) = \{x : x = (x_1, \ldots, x_n),\ a_i \leqq x_i < b_i,\ i = 1, \ldots, n\}.$$

Тогава $\mathfrak{B}_n = \sigma(\mathfrak{J})$ се нарича булова σ-алгебра в $\mathbf{R}^n$.

Следващата теорема, известна като теорема на Каратеодори за продължение на вероятностите, играе важна роля в теорията на вероятностите.

Теорема 2. *Всяка вероятностна мярка, дефинирана върху булова алгебра от подмножества на дадено множество, допуска единствено продължение върху σ-алгебрата, породена от нея.*

Това значи, че за всяко вероятностно пространство $(\Omega, \mathfrak{F}, \mathbf{P})$, където $\mathfrak{F}$ е булова алгебра, може да се построи единствено вероятностно пространство $(\Omega, \mathfrak{F}', \mathbf{P}')$ такова, че $\mathfrak{F}' = \sigma(\mathfrak{F})$ и $\mathbf{P}'(A) = \mathbf{P}(A)$ за всяко $A \in \mathfrak{F}$.

Тази теорема дава възможност да считаме отсега нататък, че във вероятностното пространство $(\Omega, \mathfrak{F}, \mathbf{P})$ съвкупността на събитията $\mathfrak{F}$ е σ-алгебра.

Доказателството на теорема 2 ще приведем в § 8, като преди това ще разгледаме някои частни (но важни) случаи на вероятностни мерки.

§ 4. Класическа вероятност

Дефиниция 1. *Нека $(\Omega, \mathfrak{F})$ е измеримо пространство, където $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n, \ldots\}$ е изброимо множество, а $\mathfrak{F} = \mathfrak{B}(\Omega)$ (т.е. множеството от всички подмножества на Ω). Нека на всяко елементарно събитие $\omega_n \in \Omega$ е съпоставено неотрицателно число $p(\omega_n)$, така че*

$$\sum_{n=1}^{\infty} p(\omega_n) = 1. \tag{1}$$

Тогава на всяко събитие $A \in \mathfrak{F}$ съпоставяме неотрицателното число

$$\mathbf{P}(A) = \sum_{\omega_i \in A} p(\omega_i), \tag{2}$$

където сумирането се извършва по всички елементарни събития, принадлежащи на A.

Ще покажем, че формула (2) дефинира една вероятностна мярка върху $(\Omega, \mathfrak{F})$, т.е. че $(\Omega, \mathfrak{F}, \mathbf{P})$ е вероятностно пространство. За тази цел трябва да проверим верността на аксиомите 1° – 4° (или на **1°**, **2°**, **3***). Наистина **1°** и **2°** очевидно следват от (1) и (2). Нека сега A, $B \in \mathfrak{F}$ и $A \cap B = \varnothing$. Тогава

$$\mathbf{P}(A+B) = \sum_{\omega \in A+B} p(\omega) = \sum_{\omega \in A} p(\omega) + \sum_{\omega \in B} p(\omega) = \mathbf{P}(A) + \mathbf{P}(B),$$

т.е. в сила е аксиома **3°**.

Ако $A_n \in \mathfrak{F}$ и $A_n \downarrow \varnothing$, то $\mathbf{P}(A_n) = \sum\limits_{\omega \in A_n} p(\omega) \downarrow 0$, тъй като при $n \to \infty$ събираемите се изчерпват, което доказва валидността на аксиома **4°**.

Следователно функцията $\mathbf{P}(A)$, дефинирана с (2), е вероятностна мярка и притежава всички свойства, които доказахме в § 3.

Сега ще разгледаме един частен случай от горния пример, когато основното пространство има краен брой елементарни събития, т.е. $\Omega = \{\omega_1, \ldots, \omega_n\}$, и на всяко елементарно събитие е съпоставено едно и също неотрицателно число $p = p(\omega_i), i = 1, \ldots, n$. Съгласно (1) е изпълнено $1 = \sum\limits_{i=1}^{n} p(\omega_i) = np$, откъдето $p = \dfrac{1}{n}$, т.е. всяко елементарно събитие има вероятност, равна на $\dfrac{1}{n} = \dfrac{1}{\#(\Omega)}$. Тогава от (2) получаваме

$$\mathbf{P}(A) = \frac{\#(A)}{\#(\Omega)} = \frac{m}{n}. \tag{3}$$

Формула (3) дефинира *класическа вероятност* (наречена така, понеже първоначално са разглеждани само такива вероятности).

Елементарните събития, които принадлежат на дадено множество A, често се наричат *благоприятни събития* (т.е. тези, при които се сбъдва събитието A). Тогава формула (3) се интерпретира като отношение на броя на благоприятните събития $m = \#(A)$ към броя на всички възможни елементарни събития $n = \#(\Omega)$.

Модели на вероятностни пространства, водещи до класическата дефиниция на вероятността, се използват, когато елементарните събития имат свойството „симетрия“, т.е. намират се в едно и също отношение към условията, дефиниращи характера на случайния опит. Такива са елементарните събития при хвърлянето на зар, спорт тото 2, лотарии и т.н.

Комбинаторика. При пресмятания в схемата на класическата вероятност широко се използва комбинаториката, особено комбинаторните понятия вариации, пермутации и комбинации (които от своя страна биват без повторение или с повторение).

Нека $X = \{x_1, x_2, \ldots, x_n\}$ е крайно множество с n различни елемента. Множеството $\{x_{i_1}, x_{i_2}, \ldots, x_{i_k}\}$, съставено от елементи на X, се нарича извадка от клас k, $k \geqq 1$.

Вариация (без повторение) на n елемента от k-ти клас, където $1 \leqq k \leqq n$, се нарича всяка наредена извадка, в която елементите не се повтарят. Две вариации $\{x_{i_1}, x_{i_2}, \ldots, x_{i_k}\}$ и $\{x_{j_1}, x_{j_2}, \ldots, x_{j_k}\}$ съвпадат само когато $x_{i_m} = x_{j_m}$, $m = 1, 2, \ldots, k$.

Вариациите на n елемента от k-ти клас можем да образуваме, като към всеки елемент присъединим вариациите на останалите $n-1$ елемента от $(k-1)$-ви клас. Следователно за броя на вариациите V_n^k на n елемента от k-ти клас получаваме формулата

$$V_n^k = nV_{n-1}^{k-1}, \tag{4}$$

откъдето последователно се намира

$$V_n^k = n(n-1)\ldots(n-k+1) = \frac{n!}{(n-k)!}. \tag{5}$$

Частният случай на вариация при $k = n$ се нарича *пермутация* на n елемента. Тогава броят на всички пермутации P_n на n елемента се дефинира по формулата

$$P_n = V_n^n = n(n-1)\cdots 2.1 = n!. \tag{6}$$

Комбинация (без повторение) на n елемента от k-ти клас се нарича всяко подмножество $\{x_{i_1}, x_{i_2}, \ldots, x_{i_k}\}$ на множеството X с мощност k (т.е. с брой на елементите k). За разлика от вариациите при комбинациите няма значение наредбата на елементите. Следователно за общия брой C_n^k на комбинациите на n елемента от k-ти клас е в сила формулата

$$C_n^k = \frac{V_n^k}{P_k} = \frac{n!}{k!(n-k)!} = \binom{n}{k}. \tag{7}$$

Сега ще разгледаме *извадки с повторение* (или *с връщане*).

Нека в извадките има N елемента от X, като k_1 от тях са равни на x_1, k_2 — на x_2 и т.н., k_n — на x_n. Очевидно $k_1+k_2+\ldots+k_n = N$.

Извадки, които се различават една от друга само по местата на неравните помежду си елементи, се наричат *пермутации с повторение.* Не е трудно да се съобрази, че за техния брой $P_N(k_1, \ldots, k_n)$ е в сила формулата

$$P_N(k_1, \ldots, k_n) = \frac{N!}{k_1! \ldots k_n!}, \quad \sum_{i=1}^{n} k_i = N. \tag{8}$$

Подобно се дефинират *вариации с повторение* на елементите — наредени m-орки от елементите на съвкупност X с $\#(X) = n$. Броят на вариациите с повторение на n елемента от m-ти клас се означава с W_n^m. Те могат да се интерпретират като извадки с връщане с обем m от съвкупност с n елемента. Отделните извадки се различават помежду си по неповтарящите се елементи или по техните места, т.е. в такава извадка $(\mathfrak{C}_1, \ldots, \mathfrak{C}_m)$ единственото ограничение е $\mathfrak{C}_i \in X$, $i = 1, \ldots, m$. Тогава лесно се съобразява, че

$$W_n^m = n^m. \tag{9}$$

С други думи, множеството на вариациите с повторение на елементите на множеството X от клас m представлява декартовото произведение.

$$X^m = X \times X \times \ldots \times X \qquad (m \text{ пъти}).$$

Комбинациите с повторение на n елемента от m-ти клас са такива извадки, при които всеки елемент на X може да се повтаря в извадката най-много до класа на комбинацията (т.е. до обема на извадката), а отделните комбинации се разграничават една от друга само по различните си елементи без оглед на тяхната наредба. Броя на комбинациите с повторение на n елемента от m-ти клас ще означаваме с K_n^m.

Комбинациите с повторение на n елемента от m-ти клас можем да получим като към елемента x_1 присъединим комбинациите с повторение на елементите $\{x_1, \ldots, x_n\}$ от $(m-1)$-ви клас, към елемента x_2 присъединим комбинациите с повторение на елементите $\{x_2, \ldots, x_n\}$ от $(m-1)$-ви клас, към x_3 — комбинациите с повторение от $(m-1)$-ви клас на $\{x_3, \ldots, x_n\}$ и т.н. Така получаваме формулата

$$K_n^m = K_n^{m-1} + K_{n-1}^{m-1} + \ldots + K_1^{m-1}. \tag{10}$$

Ще докажем по индукция, че

(11) $$K_n^m = \binom{n+m-1}{m} = C_{n+m-1}^m.$$

За $m = 1$ формулата може да се провери директно. Да допуснем, че (11) е вярна при някакво $m > 1$. Тогава от (10) и (11) получаваме

(12) $$K_n^{m+1} = \sum_{i=1}^{n} K_i^m = \sum_{j=0}^{n-1} \binom{j+m}{m}.$$

Непосредствено се проверява представянето

(13) $$C_{j+m}^m = C_{j+m+1}^{m+1} - C_{j+m}^{m+1}, \qquad j \geqq 0,$$

където $C_l^i = 0$ при $i > l$. Като заместим (13) в (12), не е трудно да видим, че

$$K_n^{m+1} = C_{n+m}^{m+1} = \binom{n+m}{m+1},$$

което формула (11) е доказана.

Пример 1 (извадка без връщане). Нека урна съдържа N топки, номерирани с числата $1, 2, \ldots, N$, от които M са бели, а останалите $N - M$ — черни. Извадката без връщане се състои в това, че случайно вадим от урната n топки последователно и без да ги връщаме обратно. Това означава, че за основно пространство на елементарните събития $\Omega = \{\omega\}$ приемаме множеството на всички наредени n-орки

(14) $$\omega = (\mathfrak{C}_1, \mathfrak{C}_2, \ldots, \mathfrak{C}_n),$$

където $1 \leqq \mathfrak{C}_i \leqq N$, $\mathfrak{C}_i \neq \mathfrak{C}_j$, $i \neq j$, $i, j = 1, \ldots, n$, т.е. Ω е съвкупността на вариациите (без повторение) от N елемента n-ти клас всички такива извадки са равновероятни.

Следователно от (5) намираме

(15) $$\#(\Omega) = V_N^n = \frac{N!}{(N-n)!}.$$

Нека A_m е събитието „сред извадените n топки има точно m бели“. Ще покажем, че

(16) $$\#(A_m) = C_n^m V_M^m V_{N-M}^{n-m}.$$

Действително, броят на елементарните събития (14), при които имаме точно m бели топки, се получава като произведение на: C_n^m (на брой) начина, по които могат да бъдат избрани местата за m-те бели топки измежду възможните n места, V_M^m възможности за вариране стойностите на m-те координати измежду M-те „бели“, и аналогично V_{N-M}^{n-m} — за останалите „черни“.

Сега по формулата за класическата вероятност (3) от (15) и (16) получаваме

$$\mathbf{P}(A_m) = \frac{C_n^m V_M^m V_{N-M}^{n-m}}{V_N^n}. \tag{17}$$

Като използваме (7), не е трудно да се види, че (17) може да се изрази в следните еквивалентни форми:

$$\mathbf{P}(A_m) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n} = \frac{C_n^m C_{N-n}^{M-m}}{C_N^M}. \tag{18}$$

Пример 2 (извадка с връщане). Сега от описаната в пример 1 урна се вадят последователно n топки, при което всеки път се записват номерът и цветът на топката, която след това се връща обратно в урната (и урната добре се разбърква). В този случай пространството от елементарните събития се състои от всевъзможните вектори (14), за чиито координати няма други ограничения освен $1 \leqq \mathfrak{C}_i \leqq N$, т.е. Ω е съвкупността на вариациите с повторение на N елемента от n-ти клас.

Следователно от (9) получаваме

$$\#(\Omega) = W_N^n = N^n. \tag{19}$$

Аналогично на (16) се показва, че в този случай

$$\#(A_m) = C_n^m W_M^m W_{N-M}^{n-m}. \tag{20}$$

По този начин от (19) и (20) по формулата за класическата вероятност (3) намираме

$$\mathbf{P}(A_m) = \frac{C_n^m M^m (N-M)^{n-m}}{N^n} = C_n^m \left(1 - \frac{M}{N}\right)^{n-m} \left(\frac{M}{N}\right)^m. \tag{21}$$

По-късно ще видим, че този пример може да бъде интерпретиран по друг начин, а (21) е частен случай на биномните вероятности.

§ 5. Геометрични вероятности

Важен клас модели на вероятностни пространства може да бъ-
де построен с помощта на геометричните вероятности.

Нека $(\Omega, \mathfrak{F})$ е измеримо пространство, където $\Omega = \{\omega\}$ е об-
аст в n-мерното евклидово пространство $\mathbf{R}^n$ с краен n-мерен
бем $\mu(\Omega) < \infty$ (под n-мерен обем разбираме съответната лебего-
а мярка μ), а $\mathfrak{F} = \Omega \cap \mathfrak{B}_n$, където $\mathfrak{B}_n$ е бореловата σ-алгебра в
$\mathbf{R}^n$ (вж. дефиниция 3.2). За всяко $A \in \mathfrak{F}$ дефинираме функцията

$$\mathbf{P}(A) = \frac{\mu(A)}{\mu(\Omega)}. \tag{1}$$

От (1) веднага се вижда, че функцията $\mathbf{P}$ удовлетворява акси-
мите $\mathbf{1}^\circ$ – $\mathbf{4}^\circ$, т.е. тя е вероятност. Тази вероятност е прието да
е нарича *геометрична вероятност.*

Така дефинираното вероятностно пространство $(\Omega, \mathfrak{F}, \mathbf{P})$ служи
ато математически модел на реални задачи, при които някаква
астица се хвърля (попада) случайно в дадена област Ω. Формула
1) показва, че при това положението на частицата е равномерно
азпределено в тази област, т.е. вероятността частицата да по-
адне в дадена област A е пропорционална на n-мерния обем на
ази област и не зависи от положението ѝ.

Пример 1. Пръчка се чупи на две части в случайна точка,
авномерно разпределена по дължината ѝ. Нека A е събитието
по-малката част има дължина, ненадминаваща $\frac{1}{3}$ от дължина-
а на пръчката". Да означим с l дължината на пръчката, т.е.
$(\Omega) = l$. Нека x е разстоянието от точката на счупване до фик-
иран край на пръчката. Тогава

$$A = \left\{x : \{x \leqq \frac{l}{3}\} \cup \{x \geqq \frac{2l}{3}\}\right\} \quad \text{и} \quad \mu(A) = \frac{l}{3} + \frac{l}{3} = \frac{2l}{3},$$

гкъдето по формула (1) получаваме $\mathbf{P}(A) = \frac{2}{3}$.

Пример 2. Във вътрешността на квадрат със страна a по слу-
аен начин е избрана точка. Каква е вероятността разстоянието
г тази точка до центъра на квадрата да не надминава $\frac{a}{4}$?

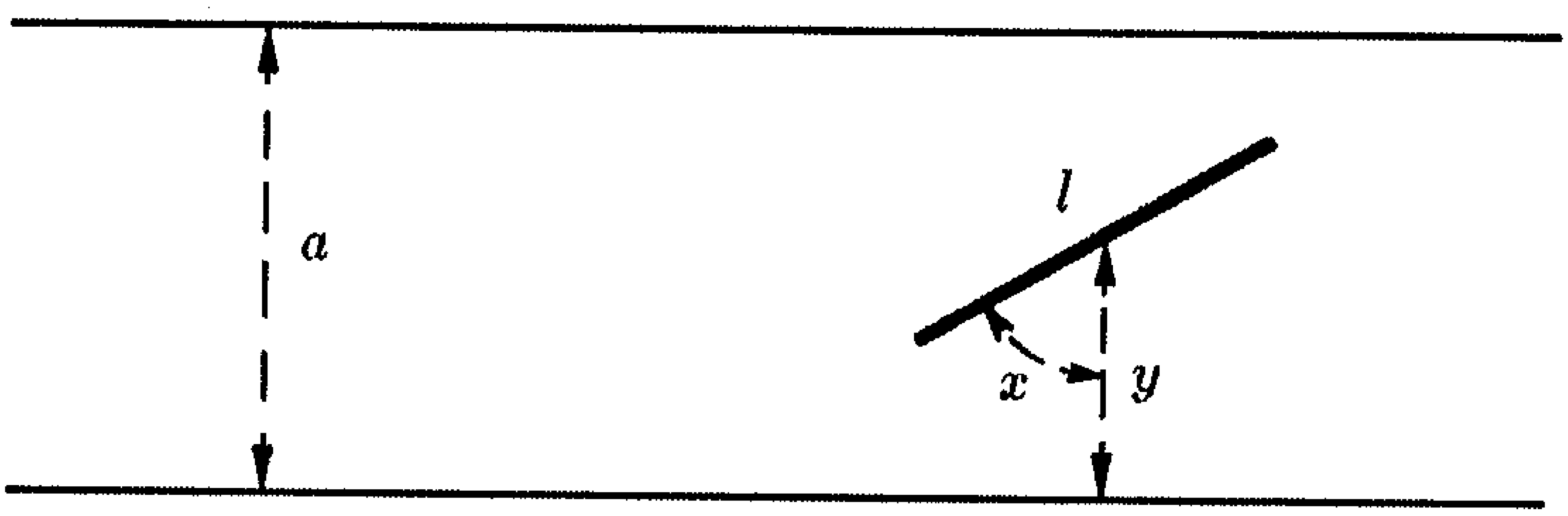
Да означим последното събитие с A. Очевидно то се реализира тогава и само тогава, когато точката лежи в кръг с радиус $\frac{a}{4}$ и център в центъра на квадрата. Тогава

$$\mu(A) = \pi \left(\frac{a}{4}\right)^2 = \frac{\pi a^2}{16}.$$

Тъй като лицето на квадрата е $\mu(\Omega) = a^2$, от формула (1) следва, че $\mathbf{P}(A) = \pi/16$.

Пример 3 (задача на Бюфон). Върху равнина, разчертана с успоредни прави на разстояние a една от друга, се хвърля случайно игла с дължина $l < a$. Каква е вероятността иглата да пресече някоя от успоредните прави?

Да означим с A въпросното събитие, с x — острия ъгъл между иглата и перпендикуляра към успоредните прави, а с y — разстоянието от средата на иглата до най-близката права (фиг. 3).



Фиг. 3

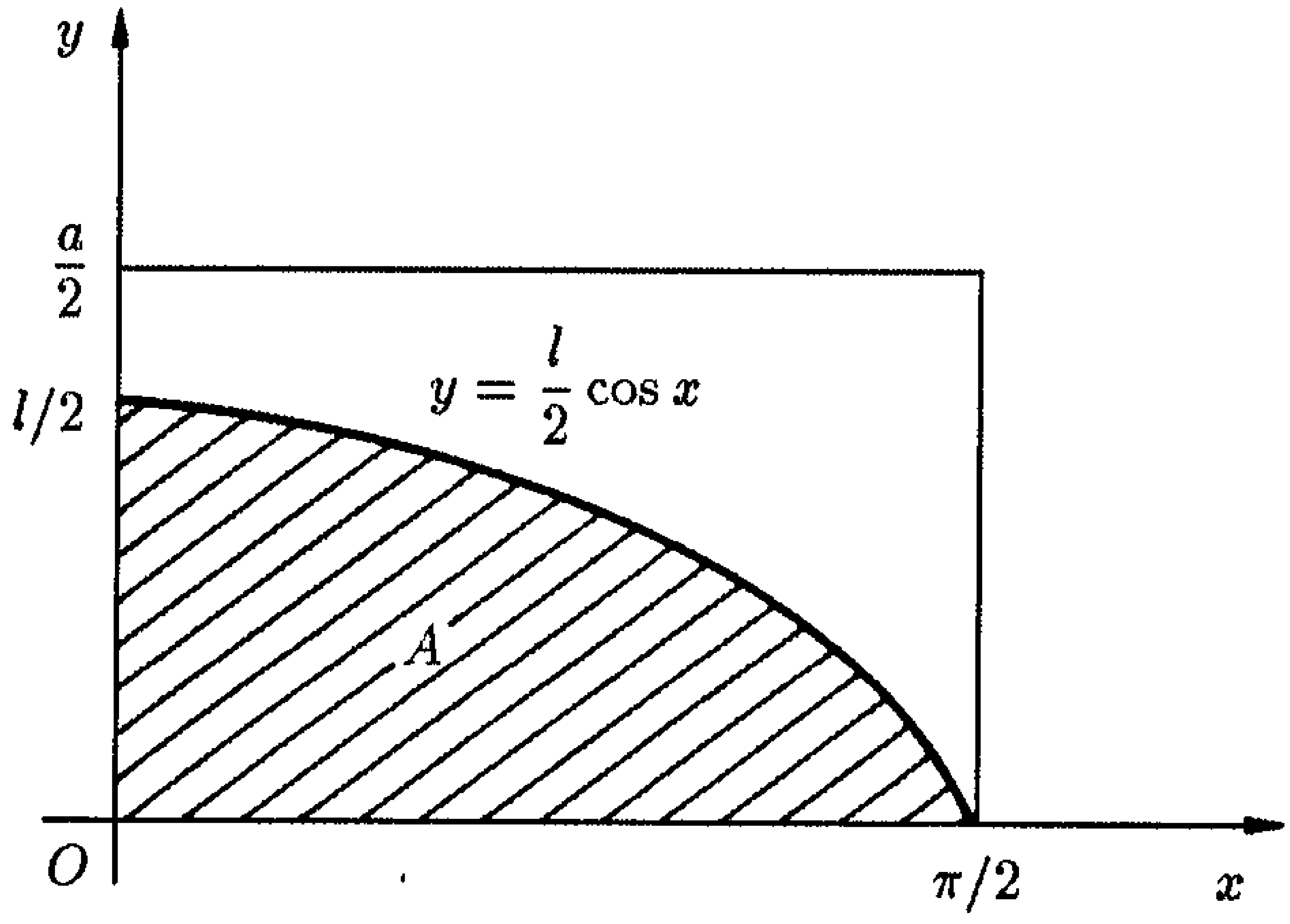
Координатите (x, y) дефинират положението на иглата относно успоредните прави. Тогава (фиг. 4)

$$\Omega = \left\{(x,y):\ 0 \leqq x \leqq \frac{\pi}{2},\ 0 \leqq y \leqq \frac{a}{2}\right\},$$

$$A = \left\{(x,y):\ 0 \leqq x \leqq \frac{\pi}{2},\ 0 \leqq y \leqq \frac{l}{2}\cos x\right\}.$$

Не е трудно да се пресметне, че

$$\mu(A) = \int\limits_0^{\pi/2} \frac{l}{2}\cos x dx = \frac{l}{2}.$$



Фиг. 4

ъй като $\mu(\Omega) = \frac{a\pi}{4}$, по формула (1) получаваме $\mathbf{P}(A) = \frac{2l}{a\pi}$.

Накрая да отбележим, че в някои задачи, свързани с геомет-ични вероятности, е възможно елементарните събития да не са очки, а други геометрични обекти — например отсечки, рав-ини и т.н. Всички тези проблеми се изучават в стохастичната юметрия — сравнително нов, но интензивно развиващ се дял от тохастиката.

§ 6. Условна вероятност. Независимост

Често се налага да търсим вероятност на дадено събитие A, ›гато е известно, че се е осъществило събитие B. Да разгледаме ървоначално няколко примера.

Пример 1. Нека случайният опит е двукратно хвърляне на ıр. Да означим с A събитието „при първото хвърляне на зара › пада единица“, а с B събитието „сумата от точките е по-малка г 4“. Основното пространство на елементарните събития е

$$\Omega = \{(i,j) :\ i,j = 1,\ldots,6\}\ , \qquad \#(\Omega) = 36.$$

Събитията A и B са следните подмножества на Ω:

$$A = \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6)\}, \quad \#(A) = 6,$$
$$B = \{(1,1),(1,2),(2,1)\}, \quad \#(B) = 3.$$

Оттук се вижда, че

$$A \cap B = \{(1,1),(1,2)\}, \quad \#(A \cap B) = 2.$$

Следователно

$$\mathbf{P}(A) = \frac{6}{36}, \ \mathbf{P}(B) = \frac{3}{36}, \ \mathbf{P}(A \cap B) = \frac{2}{36}.$$

При условие, че се е сбъднало събитието B, събитието A може да настъпи само когато се сбъднат елементарните събития (1,1) и (1,2). Тогава по формулата за класическата вероятност за вероятността на A при условие B получаваме

$$\mathbf{P}(A|B) = \frac{2}{3} = \frac{\frac{2}{36}}{\frac{3}{36}} = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}.$$

Пример 2. Нека в условията на дефинираната в § 4 класическа вероятност $\#(\Omega) = n$. Нека A и B са случайни събития, като $\#(A) = m$, $\#(B) = l$, $\#(A \cap B) = r$. Тогава

$$\mathbf{P}(A) = \frac{m}{n}, \ \mathbf{P}(B) = \frac{l}{n}, \ \mathbf{P}(A \cap B) = \frac{r}{n}.$$

Ако е настъпило събитието B, т.е. сбъднало се е едно от l-те негови елементарни събития, събитието A ще се сбъдне тогава и само тогава, когато се сбъдне някое от r-те елементарни събития, принадлежащи на $A \cap B$. Следователно

$$\mathbf{P}(A|B) = \frac{r}{l} = \frac{\frac{r}{n}}{\frac{l}{n}} = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}.$$

Тези два примера обосновават следната

Дефиниция 1. *Нека* $\mathbf{P}(B) > 0$. *Условна вероятност* $\mathbf{P}(A|B)$ *на събитието* A *при условие, че се е сбъднало събитието* B *(или накратко: при условие* B*), се нарича отношението*

(1) $$\mathbf{P}(A|B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}.$$

За условната вероятност $\mathbf{P}(A|B)$ понякога се използва и озна-
чението $\mathbf{P}_B(A)$.

Нека събитието B е фиксирано, а $A \in \mathfrak{F}$ от дадено вероятнос-
тно пространство $(\Omega, \mathfrak{F}, \mathbf{P})$. Тогава условната вероятност $\mathbf{P}_B(A)$, разглеждана като функция на събитията $A \in \mathfrak{F}$, дефинира ново вероятностно пространство $(\Omega, \mathfrak{F}, \mathbf{P}_B)$. За да докажем това, трябва да проверим валидността на аксиомите $\mathbf{1}^\circ - \mathbf{4}^\circ$ за $\mathbf{P}_B$.

Действително, от (1) получаваме

$$\mathbf{P}_B(A) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} \geqq 0; \quad \mathbf{P}_B(\Omega) = \frac{\mathbf{P}(\Omega B)}{\mathbf{P}(B)} = 1.$$

Ако $A_1 \cap A_2 = \varnothing$, то $(A_1 B) \cap (A_2 B) = \varnothing$ и

$$\mathbf{P}_B(A_1 + A_2) = \frac{\mathbf{P}(A_1 B + A_2 B)}{\mathbf{P}(B)} = \frac{\mathbf{P}(A_1 B)}{\mathbf{P}(B)} + \frac{\mathbf{P}(A_2 B)}{\mathbf{P}(B)}$$
$$= \mathbf{P}_B(A_1) + \mathbf{P}_B(A_2).$$

Накрая от $A_n \downarrow \varnothing$ следва $BA_n \downarrow \varnothing$, откъдето

$$\mathbf{P}_B(A_n) = \frac{\mathbf{P}(BA_n)}{\mathbf{P}(B)} \downarrow 0.$$

Очевидно (1) е еквивалентно на представянето

$$\mathbf{P}(AB) = \mathbf{P}(B)\mathbf{P}(A|B), \tag{2}$$

известно като *теорема за умножение на вероятностите.*

На практика (2) дава възможност за пресмятане вероятността на сечението на две събития (едно сложно събитие), ако се знаят условната вероятност на едното събитие спрямо другото и вероятността на събитието, стоящо в „условието“. Формула (2) лесно се обобщава по индукция. Ако $\mathbf{P}(A_1 \dots A_{n-1}) > 0$, то

$$\mathbf{P}(A_1 \dots A_n) = \mathbf{P}(A_1)\mathbf{P}_{A_1}(A_2) \dots \mathbf{P}_{A_1 A_2 \dots A_{n-1}}(A_n). \tag{3}$$

З а б е л е ж к а. Условната вероятност $\mathbf{P}(A|B)$ не е дефинирана, когато $\mathbf{P}(B) = 0$. Удобно е в такива случаи да се приеме, че $\mathbf{P}(A|B) = \mathbf{P}(A)$. Но както и да е дефинирана тази вероятност, тя винаги трябва да е число между 0 и 1. Затова теоремата за умножение на вероятностите (2) винаги е вярна, дори когато $\mathbf{P}(B) = 0$ (тогава вляво на (2) стои $\mathbf{P}(AB) = 0$).

Теоремата за умножение на вероятностите (2) има два варианта:

$$\mathbf{P}(AB) = \mathbf{P}(A|B)\mathbf{P}(B) \quad \text{и} \quad \mathbf{P}(AB) = \mathbf{P}(B|A)\mathbf{P}(A).$$

Аналогично нейното обобщение (3) има $n!$ еквивалентни представяния, съответни на възможните пермутации при подреждането в сечението $A_1A_2\ldots A_n$.

Дефиниция 2. *Събитията A и B се наричат независими, ако е в сила равенството*

$$\mathbf{P}(A\cap B)=\mathbf{P}(A)\mathbf{P}(B). \tag{4}$$

Ако (4) не е изпълнено, събитията A и B се наричат зависими.

Ако събитията A и B са независими и $\mathbf{P}(B)>0$, тогава

$$\mathbf{P}(A|B)=\mathbf{P}(A). \tag{5}$$

Аналогично, ако $\mathbf{P}(A)>0$, то

$$\mathbf{P}(B|A)=\mathbf{P}(B). \tag{6}$$

Доказателствата на (5) и (6) следват непосредствено от (1) и (4) и показват, че при независимост на събитията условните вероятности съвпадат с безусловните.

Понятието независимост (както ще видим по-нататък) има основно значение в теорията на вероятностите. Обикновено независимостта на A и B, която се нарича стохастична (за разлика от причинната независимост на реалните събития), не се установява с (4), а се постулира въз основа на някакви външни обективни съображения. С помощта на (4) само се пресмята вероятността $\mathbf{P}(AB)$, като се знаят вероятностите $\mathbf{P}(A)$ и $\mathbf{P}(B)$ на двете независими събития.

При установяване независимостта на събитията A и B често се използва следният принцип: Събитията A и B, чиито реални прототипи $\tilde{A}$ и $\tilde{B}$ са причинно-независими, са независими в стохастичен (теоретико-вероятностен) смисъл.

Разбира се, този принцип не е теорема, тъй като не е формулиран в рамките на математическия модел. Естествено от стохастичната независимост на събитията A и B не следва причинната независимост на техните реални прототипи $\tilde{A}$ и $\tilde{B}$.

Следващият пример показва, че независимостта може да изчезне при малка промяна на вероятностния модел.

Пример 3. От обикновено тесте с 52 карти (т.е. състоящо се от 13 различни карти, всяка в четири цвята) случайно се избира една карта. Да въведем събитията $A=\{\text{да изтеглим асо}\}$,

B = {да изтеглим пика}. Тогава AB = {да изтеглим асо пика}. Тъй като

$$\mathbf{P}(A)=4/52=1/13,\ \mathbf{P}(B)=13/52=1/4,\ \mathbf{P}(AB)=1/52=\mathbf{P}(A)\mathbf{P}(B),$$

събитията A и B са независими.

Ако същото тесте съдържа и джокер, събитията A и B ще станат зависими, защото $\mathbf{P}(A) = 4/53$, $\mathbf{P}(B) = 13/53$, $\mathbf{P}(AB) = 1/53$ и следователно $\mathbf{P}(AB) \neq \mathbf{P}(A)\mathbf{P}(B)$.

Теорема 1. *Ако събитията A и B са независими, то независими са и двойките $A, \overline{B}$; $\overline{A}, B$; $\overline{A}, \overline{B}$.*

Доказателство. Тъй като $A = A\overline{B} + AB$, съгласно аксиома 3° имаме $\mathbf{P}(A\overline{B}) = \mathbf{P}(A) - \mathbf{P}(AB)$. Оттук, като използваме (4) и свойствата 4) от § 3, получаваме

$$\mathbf{P}(A\overline{B}) = \mathbf{P}(A) - \mathbf{P}(A)\mathbf{P}(B) = \mathbf{P}(A)[1 - \mathbf{P}(B)] = \mathbf{P}(A)\mathbf{P}(\overline{B}),$$

т.е. събитията A и $\overline{B}$ са също независими. Аналогично се доказва независимостта на $\overline{A}$ и B. Независимостта на $\overline{A}$ и $\overline{B}$ следва от представянето $\overline{B} = \overline{A}\overline{B} + A\overline{B}$ и равенствата

$$\mathbf{P}(\overline{A}\overline{B}) = \mathbf{P}(\overline{B}) - \mathbf{P}(A\overline{B}) = \mathbf{P}(\overline{B}) - \mathbf{P}(A)\mathbf{P}(\overline{B}) = [1 - \mathbf{P}(A)]\mathbf{P}(\overline{B})$$
$$= \mathbf{P}(\overline{A})\mathbf{P}(\overline{B}).$$

Дефиниция 3. *Събитията $A_1, A_2, \dots, A_n$ се наричат независими (или независими в съвкупност), ако за кои да са k от тях, $2 \leqq k \leqq n$, е в сила*

$$\mathbf{P}(A_{i_1}, A_{i_2}, \dots, A_{i_k}) = \prod_{m=1}^{k} \mathbf{P}(A_{i_m}). \tag{7}$$

Ако (7) е изпълнено само при $k = 2$, събитията $A_1, A_2, \dots, A_n$ се наричат независими две по две.

От независимост в съвкупност очевидно следва независимост две по две. Обратното не е вярно, както се вижда от следващия класически

Пример на Бернщайн. Случайният експеримент е хвърляне на правилен тетраедър, стените на който са оцветени съответно бяло, зелено, червено и трикольор (бяло, зелено, червено). Нека A е събитието тетраедърът да е паднал на стена, съдържаща бял цвят, B — съдържаща зелен, C — съдържаща червен цвят Тогава

$$\mathbf{P}(A) = \mathbf{P}(B) = \mathbf{P}(C) = \frac{2}{4} = \frac{1}{2},$$

$$\mathbf{P}(AB) = \frac{1}{4} = \mathbf{P}(A)\mathbf{P}(B) = \frac{1}{2}.\frac{1}{2},$$

$$\mathbf{P}(AC) = \frac{1}{4} = \mathbf{P}(A)\mathbf{P}(C) = \frac{1}{2}.\frac{1}{2},$$

$$\mathbf{P}(BC) = \frac{1}{4} = \mathbf{P}(B)\mathbf{P}(C) = \frac{1}{2}.\frac{1}{2},$$

т.е. събитията A, B и C са две по две независими, но не са независими в съвкупност, защото

$$\mathbf{P}(ABC) = 1/4 \neq \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C) = \left(\frac{1}{2}\right)^3 = \frac{1}{8}.$$

§ 7. Формула за пълната вероятност и формула на Бейс

Без да уговаряме това специално, по-нататък винаги ще считаме, че е зададено вероятностно пространство $(\Omega, \mathfrak{F}, \mathbf{P})$.

Дефиниция 1. *Събитията $\{H_n\}$ образуват пълна група от събития (или разлагане), ако $H_i \cap H_j = \varnothing$, $i \neq j$ (т.е. събитията са две по две несъвместими) и $\sum\limits_{n=1}^{\infty} H_n = \Omega$.*

Теорема 1 (формула за пълната вероятност). *Ако $\{H_n\}$ образуват пълна група от събития и $\mathbf{P}(H_n) > 0$ за $n = 1, 2, \ldots$, то за всяко събитие A е изпълнено*

$$\mathbf{P}(A) = \sum_{n=1}^{\infty} \mathbf{P}(H_n)\mathbf{P}(A|H_n). \tag{1}$$

Доказателство. Тъй като $\sum\limits_{n=1}^{\infty} H_n = \Omega$, то

$$A = A \cap \Omega = A \cap \sum_{n=1}^{\infty} H_n = \sum_{n=1}^{\infty} AH_n.$$

Следователно по аксиома **3*** получаваме

$$(2)\qquad \mathbf{P}(A)=\sum_{n=1}^{\infty}\mathbf{P}(AH_n).$$

От теоремата 6.2 за умножение на вероятностите имаме

$$\mathbf{P}(AH_n)=\mathbf{P}(H_n)\mathbf{P}(A|H_n),$$

оето, заместено в (2), дава (1).

Събитията на пълната група $\{H_n\}$ имат характер на предпо-ожения относно това, кое от тях ще се случи. Затова понякога и наричат *хипотези.*

Пример 1. От N изпитни билета n са „щастливи" (т.е. по-лес-и). Студентите теглят последователно билети. Кой има по-го-яма вероятност да изтегли „щастлив" билет: този, който тегли ърви, или този, който тегли втори?

Нека A е събитието „първият студент е изтеглил „щастлив" илет". Очевидно $\mathbf{P}(A)=n/N$. Да означим с B събитието „вто-ият студент е изтеглил „щастлив" билет". Възможни са две ипотези: A — първият студент е изтеглил „щастлив" билет, $\overline{A}$ - първият студент не е изтеглил „щастлив" билет. По формула 1)

$$\begin{aligned}\mathbf{P}(B)&=\mathbf{P}(A)\mathbf{P}(B|A)+\mathbf{P}(\overline{A})\mathbf{P}(B|\overline{A})\\&=\frac{n}{N}.\frac{n-1}{N-1}+\frac{N-n}{N}.\frac{n}{N-1}=\frac{n}{N},\end{aligned}$$

.е. двамата студенти имат еднакъв шанс да изтеглят „щастлив" илет.

Пример 2. Дадени са n еднакви урни, като урна с номер i съ-ържа N_i топки, от които m_i са бели. От случайно избрана урна извадена топка. Каква е вероятността тази топка да е бяла?

Нека H_i е събитието да бъде избрана урната с номер i, при $=1,2,\dots,n$, а A — да бъде извадена бяла топка. Тогава

$$\mathbf{P}(H_i)=\frac{1}{n},\quad \mathbf{P}(A|H_i)=\frac{m_i}{N_i}$$

по формулата за пълната вероятност намираме

$$\mathbf{P}(A)=\sum_{i=1}^{n}\mathbf{P}(H_i)\mathbf{P}(A|H_i)=\frac{1}{n}\sum_{i=1}^{n}\frac{m_i}{N_i}.$$

Теорема 2 (формула на Бейс). *При условията на теорема* 1, *ако* $\mathbf{P}(A) > 0$, *в сила са формулите*

$$\mathbf{P}(H_k|A) = \frac{\mathbf{P}(H_k)\mathbf{P}(A|H_k)}{\sum\limits_{i=1}^{\infty} \mathbf{P}(H_i)\mathbf{P}(A|H_i)} \quad \text{за } k = 1, 2, \ldots \tag{3}$$

Доказателство. По теоремата за умножение на вероятностите (6.2) намираме $\mathbf{P}(H_k A) = \mathbf{P}(H_k)\mathbf{P}(A|H_k) = \mathbf{P}(A)\mathbf{P}(H_k|A)$, откъдето

$$\mathbf{P}(H_k|A) = \mathbf{P}(H_k)\mathbf{P}(A|H_k)/\mathbf{P}(A).$$

Като приложим към знаменателя $\mathbf{P}(A)$ формулата за пълната вероятност (1), получаваме (3).

Формулата на Бейс допуска следната интерпретация: Нека събитието A е резултат от някакъв експеримент. Вероятностите $\mathbf{P}(H_k)$ обикновено се наричат *априорни вероятности на хипотезите* H_k, т.е. такива, които се дефинират от условията на реализацията на опита. Условните вероятности $\mathbf{P}(H_k|A)$ се наричат *апостериорни вероятности* и се дефинират в зависимост от изхода на опита A, който може би носи допълнителна информация за условията на опита. Формулата на Бейс позволява по априорните вероятности на хипотезите и по условните вероятности на събитието A при хипотезите $\{H_k\}$ да се пресмятат апостериорните вероятности $\mathbf{P}(H_k|A)$.

Пример 3. В урна има n топки, от които известен брой са бели. Възможни са $n+1$ хипотези H_i, $i = 0, 1, \ldots, n$, където H_i е хипотезата в урната да има i на брой бели топки. При липса на друга информация имаме всички основания да предположим, че априорните вероятности са равни помежду си, т.е.

$$\mathbf{P}(H_0) = \mathbf{P}(H_1) = \cdots = \mathbf{P}(H_n) = \frac{1}{n+1}.$$

От урната случайно е извадена една топка, която се е оказала бяла. Да означим с B събитието „случайно извадена топка от урната е бяла“ и да пресметнем $\mathbf{P}(H_k|B)$, $k = 0, 1, \ldots, n$.

Тъй като $\mathbf{P}(B|H_i) = \dfrac{i}{n}$, по формулата на Бейс (3) получаваме

$$\mathbf{P}(H_k|B) = \frac{\frac{1}{n+1} \cdot \frac{k}{n}}{\sum\limits_{i=0}^{n} \frac{1}{n+1} \cdot \frac{i}{n}} = \frac{2k}{n(n+1)}.$$

Оттук се вижда, че най-вероятна е хипотезата H_n и за нейната юстериорна вероятност имаме $\mathbf{P}(H_n|B) = 2/(n+1) = 2\mathbf{P}(H_n)$.

§ 8. Теорема за продължение на вероятностите

Преди да преминем към доказателството на теорема 3.2, ще жажем някои помощни резултати.

Разглеждаме вероятностното пространство $(\Omega, \mathfrak{F}, \mathbf{P})$, където $\mathfrak{F}$ булова алгебра от подмножества на Ω.

Лема 1. *Нека* $A_n \uparrow$, $B_m \uparrow$, *където* A_n, $B_m \in \mathfrak{F}$ *и* $\bigcup\limits_{n=1}^{\infty} A_n \subset \bigcup\limits_{m=1}^{\infty} B_m$. *огава* $\lim\limits_{n\to\infty} \uparrow \mathbf{P}(A_n) \leqq \lim\limits_{m\to\infty} \uparrow \mathbf{P}(B_m)$.

Доказателство. За всяко фиксирано n редицата A_nB_m е расща по $m \geqq 1$ и $\bigcup\limits_{m=1}^{\infty} A_nB_m = A_n$. Тогава от непрекъснатостта на ›роятността (вж. свойство 3.8.) получаваме

$$\lim_{m\to\infty} \uparrow \mathbf{P}(B_m) \geqq \lim_{n\to\infty} \uparrow \mathbf{P}(A_nB_m) = \mathbf{P}(A_n).$$

ттук при $n \to \infty$ следва твърдението на лемата.

Следствие. *При условията на лема* 1, *ако* $\bigcup\limits_{n=1}^{\infty} A_n = \bigcup\limits_{m=1}^{\infty} B_m$, *то* $\lim\limits_{n\to\infty} \uparrow \mathbf{P}(A_n) = \lim\limits_{m\to\infty} \uparrow \mathbf{P}(B_m)$.

Да означим с $\mathcal{U}$ класа на всевъзможните изброими обединения ι елементи от $\mathfrak{F}$ или, което е еквивалентно, класа на границите ι монотонно растящите редици от елементи на $\mathfrak{F}$, т.е.

$$) \quad \mathcal{U} = \{G : G = \lim \uparrow A_n,\ A_n \in \mathfrak{F}\} = \left\{G : G = \bigcup_{m=1}^{\infty} B_m,\ B_m \in \mathfrak{F}\right\}.$$

Тогава съгласно следствието от лема 1 в $\mathcal{U}$ е еднозначно де›нирана следната функция:

$$) \qquad Q(G) = \lim_{n\to\infty} \uparrow \mathbf{P}(A_n), \quad A_n \uparrow G, \quad A_n \in \mathfrak{F}.$$

Лема 2. *Класът* $\mathcal{U}$ *е минималният клас от подмножества на , породен от* $\mathfrak{F}$, *и имащ следните свойства:*

а) $0 \leqq Q(G) \leqq 1$ *за всяко* $G \in \mathcal{U}$;

б) $\varnothing$, $\Omega \in \mathcal{U}$, *като* $Q(\varnothing) = 0$ *и* $Q(\Omega) = 1$;

в) *ако* $G_1 \subset G_2$; G_1, $G_2 \in \mathcal{U}$, *то* $Q(G_1) \leqq Q(G_2)$;

г) *ако* G_1, $G_2 \in \mathcal{U}$, *то*

$$G_1 \cup G_2 \in \mathcal{U}, G_1 \cap G_2 \in \mathcal{U} \quad \text{и} \quad Q(G_1) + Q(G_2) = Q(G_1 \cup G_2) + Q(G_1 \cap G_2);$$

д) *ако* $G_n \in \mathcal{U}$ *и* $G_n \uparrow G$, *то* $G \in \mathcal{U}$ *и* $Q(G) = \lim_{n\to\infty} Q(G_n)$;

е) *за всяко* $A \in \mathfrak{F}$ *е в сила* $Q(A) = \mathbf{P}(A)$.

Доказателство. Свойствата а) и б) са очевидни, а свойство в) следва от лема 1. Нека A_n, $B_n \in \mathfrak{F}$ и $A_n \uparrow G_1$, $B_n \uparrow G_2$. Тъй като $A_n \cup B_n \uparrow G_1 \cup G_2$, $A_n B_n \uparrow G_1 G_2$ и освен това по свойство 3.5 $\mathbf{P}(A_n) + \mathbf{P}(B_n) = \mathbf{P}(A_n \cup B_n) + \mathbf{P}(A_n B_n)$, то при $n \to \infty$ получаваме свойство г).

Нека за всяко $n \geqq 1$ имаме $A_{m,n} \uparrow G_n$ при $m \to \infty$, където $A_{m,n} \in \mathfrak{F}$ и $G_n \uparrow G$. Нека $C_m = \sup_{n \leqq m} A_{m,n}$. Очевидно $C_m \in \mathfrak{F}$ и $C_m \uparrow$. От друга страна, $A_{m,n} \subset C_m \subset G_m$ и следователно $\mathbf{P}(A_{m,n}) \leqq \mathbf{P}(C_m) \leqq Q(G_m)$ при $n \leqq m$. Оттук при $m \to \infty$ получаваме $G_n \subset \lim \uparrow C_m \subset G$ и $Q(G_n) \leqq \lim_{m\to\infty} \uparrow \mathbf{P}(C_m) \leqq \lim_{m\to\infty} Q(G_m)$, откъдето при $n \to \infty$ окончателно намираме

$$G_n \uparrow G \in \mathcal{U}, \quad Q(G) = \lim_{m\to\infty} \mathbf{P}(C_m) = \lim_{n\to\infty} Q(G_n),$$

което доказва свойството д).

Твърдение е) на лемата е очевидно.

Върху $\mathfrak{B}(\Omega)$ — множеството от всички подмножества на Ω, дефинираме следната функция:

$$Q^*(B) = \inf_G \{Q(G) : B \subset G \in \mathcal{U}\}, \tag{3}$$

която се нарича *външна мярка*.

Лема 3. *Външната мярка* $Q^*(B)$, *дефинирана в* (3), *има свойствата*:

а) $0 \leqq Q^*(B) \leqq 1$ *и* $Q^*(G) = Q(G)$ *при* $G \in \mathcal{U}$;

б) *ако* $B_1 \subset B_2$, *то* $Q^*(B_1) \leqq Q^*(B_2)$;

в) $Q^*(B_1 \cup B_2) + Q^*(B_1 \cap B_2) \leqq Q^*(B_1) + Q^*(B_2)$, *като в частност* $Q^*(B) + Q^*(\overline{B}) \geqq 1$;

г) *ако* $B_n \uparrow B$ *при* $n \to \infty$, *то* $Q^*(B_n) \uparrow Q^*(B)$.

Доказателство. Свойство а) е очевидно, а свойство б) следва от монотонността на Q.

От (3) следва, че за всяко $\varepsilon > 0$ можем да изберем $G_i \in U$ така,
: $B_i \subset G_i$ и $Q^*(B_i) + \frac{\varepsilon}{2} \geqq Q(G_i)$, $i = 1, 2$. Тогава

$$Q^*(B_1) + Q^*(B_2) + \varepsilon \geqq Q(G_1) + Q(G_2)$$

$$= Q(G_1 \cup G_2) + Q(G_1 \cap G_2) \geqq Q^*(B_1 \cup B_2) + Q^*(B_1 \cap B_2),$$

:където при $\varepsilon \to 0$ получаваме в).

За да докажем г), ще отбележим, че за всяко $\delta > 0$ можем да :берем редица от числа $\delta_n > 0$ и съответни подмножества $G_n \in \mathcal{U}$ ка, че $\sum_{n=1}^{\infty} \delta_n = \delta$, $B_n \subset G_n$ и $Q^*(B_n) + \delta_n \geqq Q(G_n)$. Да означим $_n = \sup_{m \leqq n} G_m$. Очевидно $D_n \in \mathcal{U}$, $D_n \uparrow$ и $B_n \subset D_n$. Не е трудно да покаже по индукция, че

$$) \qquad Q^*(B_n) + \sum_{k \leqq n} \delta_k \geqq Q(D_n), \qquad n \geqq 1.$$

:йствително, да допуснем, че (4) е вярно за някакво $n \geqq 1$ (при $= 1$ то очевидно е вярно). Тъй като $B_n \subset D_n \cap G_{n+1} \in \mathcal{U}$, от лема свойство в) получаваме

$$\begin{aligned}
Q(D_{n+1}) &= Q(D_n \cup G_{n+1}) \\
&= Q(D_n) + Q(G_{n+1}) - Q(D_n G_{n+1}) \\
&\leqq Q^*(B_n) + \sum_{k \leqq n} \delta_k + Q^*(B_{n+1}) + \delta_{n+1} - Q^*(B_n) \\
&= Q^*(B_{n+1}) + \sum_{k \leqq n+1} \delta_k.
\end{aligned}$$

й като $B \subset \lim_{n\to\infty} \uparrow D_n \in \mathcal{U}$, при $n \to \infty$ от (4) получаваме

$$\lim_{n\to\infty} \uparrow Q^*(B_n) + \delta \geqq \lim_{n\to\infty} \uparrow Q(D_n) = Q(\lim_{n\to\infty} \uparrow D_n) \geqq Q^*(B).$$

тук, като вземем предвид, че съгласно б) $Q^*(B_n) \leqq Q^*(B)$, по- чаваме г), с което лемата е доказана.

Да дефинираме класа от подмножества на Ω

$$\mathfrak{A} \rightleftharpoons \{A : Q^*(A) + Q^*(\overline{A}) = 1, \quad A \in \mathfrak{B}(\Omega)\}.$$

Лема 4. *Класът* $\mathfrak{A}$ *е* σ*-булова алгебра и* Q^* *е вероятностна мяр ка в измеримото пространство* $(\Omega, \mathfrak{A})$.

Доказателство. От дефиниция (5) следва, че класът $\mathfrak{A}$ е устой чив относно операцията допълнение, т.е. от $A \in \mathfrak{A}$ следва $\overline{A} \in \mathfrak{A}$.

От леми 3 а) и 2 б) следва, че $Q^*(\varnothing)+Q^*(\Omega) = 1$, т.е. $\varnothing$, $\Omega \in \mathfrak{A}$

За A_1, $A_2 \in \mathfrak{A}$ по лема 3 в) получаваме

$$(6) \qquad \begin{cases} Q^*(A_1 \cup A_2) + Q^*(A_1 \cap A_2) \leqq Q^*(A_1) + Q^*(A_2), \\ Q^*(\overline{A_1 \cup A_2}) + Q^*(\overline{A_1 \cap A_2}) \leqq Q^*(\overline{A_1}) + Q^*(\overline{A_2}), \end{cases}$$

където съгласно (5) сумата от десните части на (6) е равна на 2

От лема 3 в) следва, че

$$(7) \qquad \begin{cases} Q^*(A_1 \cup A_2) + Q^*(\overline{A_1 \cup A_2}) \geqq 1, \\ Q^*(A_1 \cap A_2) + Q^*(\overline{A_1 \cap A_2}) \geqq 1. \end{cases}$$

Не е трудно да се съобрази, че (6) и (7) са изпълнени едновременно само в случай на равенство. Следователно $A_1 \cup A_2 \in \mathfrak{A}$ и $A_1 \cap A_2 \in \mathfrak{A}$, т.е. класът $\mathfrak{A}$ е устойчив относно крайните обединения и сечения, като освен това Q^* е силно адитивна в $\mathfrak{A}$ (вж. свойство 5) от § 3).

Досега показахме, че $\mathfrak{A}$ е булова алгебра и Q^* е неотрицателна, нормирана и адитивна мярка в нея. От друга страна, ако $A_n \in \mathfrak{A}$ и $A_n \uparrow$, съгласно лема 3 б) и г) при $m \geqq 1$ получаваме

$$Q^*\left(\overline{\bigcup_{n=1}^{\infty} A_n}\right) \leqq Q^*(\overline{A_m}), \qquad Q^*\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \uparrow Q^*(A_n).$$

Като съберем двете съотношения и положим $m \to \infty$, получаваме

$$Q^*\left(\bigcup_{n=1}^{\infty} A_n\right) + Q^*\left(\overline{\bigcup_{n=1}^{\infty} A_n}\right) \leqq 1.$$

От лема 3 в) следва, че в горното неравенство трябва да имаме равенство, т.е. $\bigcup_{n=1}^{\infty} A_n \in \mathfrak{A}$. Това показва, че $\mathfrak{A}$ е σ-булова алгебра и Q^* е вероятностна мярка в нея.

Доказателство на теорема 3.2 (за продължение на вероятностите). Нека $\mathbf{P}$ е вероятност в измеримото пространство $(\Omega, \mathfrak{F})$, където $\mathfrak{F}$ е булова алгебра от подмножества на Ω.

Съгласно (1) и (2) построяваме класа $\mathcal{U}$ и дефинираната в не-
о функция Q, за която по (3) определяме външната мярка Q^*, ткъдето в съответствие с (5) се дефинира σ-алгебрата $\mathfrak{A}$.

Поради леми 2 и 3 за всяко $A \in \mathfrak{F}$ имаме $\mathbf{P}(A) = Q(A) = Q^*(A)$, ткъдето следва, че $\mathfrak{F} \subset \mathfrak{A}$. Тогава за $\mathfrak{F}' = \sigma(\mathfrak{F})$ (т.е. σ-алгебрата, ородена от $\mathfrak{F}$) имаме $\mathfrak{F} \subset \mathfrak{A}$ и ограничението $\mathbf{P}'$ на вероятностната ярка Q^* върху $\mathfrak{F}'$ е търсеното продължение на $\mathbf{P}$ като вероятност тази σ-алгебра.

Остава да докажем единствеността на това продължение.

Нека $\mathbf{P}_1$ е вероятност в $\mathfrak{F}'$, ограничението на която върху $\mathfrak{F}$ ъвпада с $\mathbf{P}$, т.е. $\mathbf{P}_1(A) = \mathbf{P}(A)$ за всяко $A \in \mathfrak{F}$. Тогава очевидно $_1(G)=Q(G)$ за всяко $G \in \mathcal{U}$ и следователно $\mathbf{P}_1(B) \leqq \mathbf{P}'(B)$ за всяко $\in \mathfrak{F}$. Да допуснем, че за някакво $B_0 \in \mathfrak{F}'$ имаме $\mathbf{P}_1(B) < \mathbf{P}'(B_0)$. огава

$$\mathbf{P}_1(\Omega) = \mathbf{P}_1(B_0) + \mathbf{P}_1(\overline{B_0}) < \mathbf{P}'(B_0) + \mathbf{P}'(\overline{B_0}) = 1.$$

Полученото противоречие доказва единствеността на продъл-
ението.

Задачи

1. Докажете, че за всяка тройка събития A, B, C е изпълнено:

а) $A(BC) = (AB)C$, $A \cup (B \cup C) = (A \cup B) \cup C$ (асоциативен закон);

б) $A \cap (B \cup C) = AB \cup AC$, $A \cup (BC) = (A \cup B)(A \cup C)$ (дистрибутивен закон).

2. Докажете, че следните събития са еквивалентни:

а) $\overline{A\overline{B} \cup A} \cap B$ и $B \setminus A$;

б) AB и $A \setminus (A \setminus B)$.

3. Докажете, че $\mathbf{P}(A) \leqq \mathbf{P}(\overline{B})$ при $A \cap B = \varnothing$.

4. Докажете, че за всеки три събития A, B и C е в сила формулата

$$\mathbf{P}(A \cup B \cup C) = \mathbf{P}(A) + \mathbf{P}(B) + \mathbf{P}(C) - \mathbf{P}(AB) - \mathbf{P}(AC) - \mathbf{P}(BC) - \mathbf{P}(ABC).$$

5. Числата $1, 2, \ldots, n$ се нареждат в случаен ред. Покажете, че вероят-
стта числата 1 и 2 да са съседни в такъв ред е $1/n$.

6. На фиш от ТОТО 2 са зачертани шест числа. Покажете, че вероят-
стта точно три от тях да бъдат изтеглени в поредния тираж е 0,017650...

7. Покажете, че вероятността сред k случайни цифри ($k \geqq 3$) цифрата 0
се появи точно три пъти е $k(k-1)(k-2)(0,9)^k/4374$.

8. Върху квадрат случайно е хвърлена точка. Докажете, че вероятнос та точката да бъде отдалечена от върховете на квадрата на разстояние, н по-малко от дължината на половината страна, е $1 - \pi/4$.

9. В квадрат $\Omega = [0, 1]^2$ случайно е хвърлена точка. Нека A е събитиет „абсцисата на точката е не по-малка от p“, $0 < p < 1$, а B — събитието „о динатата на точката е не по-малка от q“, $0 < q < 1$. Покажете, че събитият A и B са независими.

10. Урна съдържа N билета, от които $M \leqq N$ печелят. Нека A_k е събі тието да се изтегли печеливш билет, след като от урната вече са изтеглен k билета. Покажете, че $\mathbf{P}(A_k) = M/N$, т.е. шансовете за печалба не завися от момента на теглене.

11. Три машини произвеждат 25, 35 и 40% от даден вид изделия пр брак съответно 5, 4 и 2%. Покажете, че вероятността случайно избран изделие да бъде дефектно е 0,0345 и ако случайно избраното изделие се оказало дефектно, вероятностите то да е произведено от първата, втора или третата машина са съответно 25/69, 28/69 и 16/69.

12. Нека $A_1, \ldots, A_n$ са случайни събития. Докажете по индукция фо мулата

$$\mathbf{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbf{P}(A_i) - \sum_{1 \leqq i<j \leqq n} \mathbf{P}(A_i A_j)$$

$$+ \sum_{1 \leqq i<j<k \leqq n} \mathbf{P}(A_i A_j A_k) - \cdots + (-1)^{n-1} \mathbf{P}\left(\bigcap_{i=1}^{n} A_i\right).$$

13. Секретарка е написала n писма, които сложила в пликове и запечата ла, а след това написала адресите в случаен ред. Покажете, че вероятностт поне едно писмо да попадне на верния адрес е

$$p_n = 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n+1} \frac{1}{n!}$$

и при $n \to \infty$ клони към $1 - e^{-1} \approx 0,63$.

Упътване. Използвайте зад. 12 с $A_i = \{i$-тият адресант получава пред назначеното за него писмо$\}$.

Втора глава

Случайни величини и функции на разпределение

§ 1. Емпирични представи

Случайните величини са едни от основните обекти на изслед- ние в теорията на вероятностите. Преди да дадем формална финиция, ще се спрем на някои примери.

Броят на точките при хвърляне на зар се мени от опит на опит, то взема различни стойности от множеството {1, 2, 3, 4, 5, 6}. умата от точките при хвърляне на два зара също се колебае от пит на опит със стойности от 2 до 12. Броят на обажданията от онатите на телефонна станция за даден интервал от време при- ма различни стойности в зависимост от случайни обстоятелства. ри стрелба отклонението на попадението от центъра на мише- та има случаен характер и очевидно може да взема различни тойности в даден интервал $[0, R]$. Въобще всяко измерване (както че бе отбелязано в първа глава) има случаен характер. Същото ожем да кажем за температурата на въздуха, броя на падащи- космически частици, радиоактивното разпадане, скоростта на олекулите в газ и т. н.

Независимо от разнообразния характер на приведените приме- и от математическа гледна точка те представляват една и съща ртина. Във всеки пример имаме величина, която характеризира следваното явление и която под влиянието на случайни фактори иема различни стойности. При това предварително не можем кажем точно каква стойност ще вземе дадената величина, тъй то тя се мени случайно от опит на опит, т. е. случайната ве- чина е функция на елементарните събития. Следователно пър- то нещо, което трябва да знаем за една случайна величина, е вкупността от стойностите, които тя може да приема.

Това, разбира се, е недостатъчно, за да направим някакви съ- ествени изводи. При задаването на случайна величина бихме искали да знаем не само стойностите, които тя взема, но и колко често, т. е. с каква вероятност, приема тя тези стойности. Сега ще се постараем да формализираме тези идеи в рамките на математическия модел.

§ 2. Пространство на случайните величини

Нека $(\Omega, \mathfrak{F}, \mathbf{P})$ е вероятностно пространство, а $(\mathbf{R}^1, \mathfrak{B}_1)$ — бореловото измеримо пространство (вж. дефиниция 1.3.1).

Дефиниция 1. *Числова функция $\xi = \xi(\omega)$, изобразяваща Ω в $\mathbf{R}^1$, се нарича $\mathfrak{F}$-измерима (или накратко измерима), ако*

$$\xi^{-1}(B) = \{\omega : \xi(\omega) \in B\} \in \mathfrak{F} \quad \text{за всяко } B \in \mathfrak{B}_1. \tag{1}$$

В теорията на вероятностите е прието измеримите функции да се наричат *случайни величини.* Специално, ако $\Omega = \mathbf{R}^1$ и $\mathfrak{F} = \mathfrak{B}_1$, функцията $\xi(\omega)$ се нарича *борелова.* Същността на тази дефиниция е следната. Тъй като не всяко подмножество на Ω е събитие, естествено трябва да се разглеждат функции $\xi(\omega)$, чиито прообрази на борелови множества $\xi^{-1}(B)$ съгласно (1) са събития и следователно има смисъл да се говори за тяхната вероятност $\mathbf{P}\{\omega : \xi(\omega) \in B\}$.

Теорема 1. *Условието* (1) *е еквивалентно на следното условие:*

$$\{\omega : \xi(\omega) < x\} \in \mathfrak{F} \quad \text{за всяко } x \in \mathbf{R}^1. \tag{2}$$

Доказателство. В едната посока твърдението е очевидно, тъй като $I_x = (-\infty, x) \in \mathfrak{B}_1$ и следователно $\xi^{-1}(I_x) \in \mathfrak{F}$. Нека сега е в сила (2). Да отбележим, че операцията „вземане на прообраз" (т. е. преход от $B \subset \mathbf{R}^1$ към $\xi^{-1}(B) \subset \Omega$) запазва теоретико-множествените операции, т. е.

$$\begin{aligned} \xi^{-1}(\cup B_k) &= \cup \xi^{-1}(B_k), \\ \xi^{-1}(\cap B_k) &= \cap \xi^{-1}(B_k), \\ \xi^{-1}(\overline{B}) &= \overline{\xi^{-1}(B)}. \end{aligned} \tag{3}$$

Въвеждаме клас $\mathfrak{A} \subset \mathfrak{B}_1$, такъв че

$$\mathfrak{A} = \{B : B \in \mathfrak{B}_1,\ \xi^{-1}(B) \in \mathfrak{F}\}.$$

Съгласно (3) $\mathfrak{A}$ е σ-алгебра. Ще покажем, че $\mathfrak{A}$ съдържа всички интервали. Действително, съотношенията

$$\{\omega : \xi(\omega) \geqq x\} = \Omega \setminus \{\omega : \xi(\omega) < x\},$$

$$\{\omega : \xi(\omega) \leqq x\} = \bigcap_{n=1}^{\infty} \{\omega : \xi(\omega) < x + 1/n\},$$

$$\{\omega : \xi(\omega) > x\} = \Omega \setminus \{\omega : \xi(\omega) \leqq x\},$$

$$\{\omega : a \leqq \xi(\omega) < b\} = \{\omega : \xi(\omega) < b\} \setminus \{\omega : \xi(\omega) < a\},$$

$$\{\omega : a < \xi(\omega) < b\} = \{\omega : \xi(\omega) < b\} \setminus \{\omega : \xi(\omega) \leqq a\},$$

$$\{\omega : \xi(\omega) = x\} = \{\omega : \xi(\omega) \leqq x\} \setminus \{\omega : \xi(\omega) < x\},$$

$$\{\omega : a \leqq \xi(\omega) \leqq b\} = \{\omega : \xi(\omega) \leqq b\} \setminus \{\omega : \xi(\omega) < a\}$$

показват, че горните множества са събития, т. е. принадлежат на $\mathfrak{F}$. Следователно $\mathfrak{A}$ съдържа и минималната σ-алгебра, породена от интервалите, т. е. $\mathfrak{B}_1 \subset \mathfrak{A}$, което доказва теоремата.

Теорема 1 ни дава възможност на практика да използваме по-простото условие (2).

От (3) следва, че съвкупността $\mathfrak{F}_\xi = \{A : A = \xi^{-1}(B), B \in \mathfrak{B}_1\}$ е σ-алгебра, която се нарича *σ-алгебра, породена от ξ*, и се бележи $\mathfrak{F}_\xi = \sigma(\xi)$.

В теорията на вероятностите е прието случайните величини да се отбелязват с една буква (най-често гръцка), без да се пише аргументът (но той трябва винаги да се подразбира!).

Да означим с $\mathfrak{S} = \mathfrak{S}(\Omega, \mathfrak{F}, \mathbf{P})$ *пространството на случайните величини*, дефинирани в дадено вероятностно пространство $(\Omega, \mathfrak{F}, \mathbf{P})$, и да изследваме структурата му.

1) $\mathfrak{S}$ ***съдържа константите***, т. е. за всяко $a \in \mathbf{R}^1$ имаме $a \in \mathfrak{S}$. Действително,

$$\{a < x\} = \begin{cases} \varnothing & \text{за } x \leqq a, \\ \Omega & \text{за } x > a. \end{cases}$$

2) $\mathfrak{S}$ ***съдържа индикаторите на всички събития***, т. е. ако $A \in \mathfrak{F}$, то $I_A \in \mathfrak{S}$. Да напомним, че

$$I_A(\omega) = \begin{cases} 0 & \text{за } \omega \notin A \text{ (т.е. } \omega \in \overline{A}), \\ 1 & \text{за } \omega \in A \end{cases}$$

и следователно

$$\{I_A < x\} = \begin{cases} \varnothing & \text{за } x \leqq 0, \\ \overline{A} & \text{за } 0 < x \leqq 1, \\ \Omega & \text{за } x > 1. \end{cases}$$

3) $\mathfrak{S}$ *е линейно пространство*, т. е. ако $\xi_1, \xi_2 \in \mathfrak{S}$, то $\xi_1 + \xi_2 \in \mathfrak{S}$ и $a\xi_i \in \mathfrak{S}$ за всяко $a \in \mathbf{R}^1$.

Второто твърдение е очевидно, тъй като за всички $x \in \mathbf{R}^1$ имаме

$$\{a\xi_i < x\} = \left\{\xi_i < \frac{x}{a}\right\} \in \mathfrak{F} \quad \text{при } a > 0,$$

$$\{a\xi_i > x\} = \left\{\xi_i > \frac{x}{a}\right\} \in \mathfrak{F} \quad \text{при } a < 0.$$

Не е трудно да се провери, че

$$\{\xi_1 + \xi_2 < x\} = \bigcup_k [\{\xi_1 < x - r_k\} \cap \{\xi_2 < r_k\}], \tag{4}$$

където $\{r_k\}$ е множеството (изброимо) на рационалните числа.

Действително, нека $\omega \in \{\xi_1 + \xi_2 < x\}$. Тогава $\xi_2(\omega) < -\xi_1(\omega) + x$. Следователно съществува рационално число r_k, такова че $\xi_2(\omega) < r_k < -\xi_1(\omega) + x$, т. е. $\xi_2(\omega) < r_k$ и $\xi_1(\omega) < x - r_k$, което показва, че $\omega \in \{\xi_2 < r_k\} \cap \{\xi_1 < x - r_k\}$. Обратно, нека ω принадлежи на дясната страна на (4). Тогава съществува такова k, че $\omega \in \{\xi_1 < x - r_k\} \cap \{\xi_2 < r_k\}$, т. е. $\xi_1(\omega) < x - r_k$ и $\xi_2(\omega) < r_k$, откъдето $\xi_1(\omega) + \xi_2(\omega) < x$.

4) Аналогично се показва, че $\mathfrak{S}$ *е алгебра*, т. е. от $\xi_1, \xi_2 \in \mathfrak{S}$ следва, че $\xi_1\xi_2 \in \mathfrak{S}$ и $\xi_1/\xi_2 \in \mathfrak{S}$. Естествено, неопределеностите от вида $\infty - \infty$, $0.\infty$, $\frac{0}{0}$ и $\frac{\infty}{\infty}$ трябва да се изключват.

5) В $\mathfrak{S}$ приемаме естествената наредба $\xi_1 \leqq \xi_2$, ако $\xi_1(\omega) \leqq \xi_2(\omega)$ за всички $\omega \in \Omega$. Тогава от $\xi_1, \xi_2 \in \mathfrak{S}$ следва, че $\max(\xi_1, \xi_2) \in \mathfrak{S}$ и $\min(\xi_1, \xi_2) \in \mathfrak{S}$. Действително, за всяко $x \in \mathbf{R}^1$ имаме

$$\{\max(\xi_1, \xi_2) < x\} = \{\xi_1 < x\} \cap \{\xi_2 < x\} \in \mathfrak{F},$$

$$\{\min(\xi_1, \xi_2) > x\} = \{\xi_1 > x\} \cap \{\xi_2 > x\} \in \mathfrak{F}.$$

6) $\mathfrak{S}$ *е затворено относно граничен преход.* Действително, нека $\{\xi_n\}$ е редица от случайни величини. Тогава $\sup \xi_n \in \mathfrak{S}$ и $\inf \xi_n \in \mathfrak{S}$,

което за всяко $x \in \mathbf{R}^1$ имаме

$$\left\{\sup_{n \geqq 1} \xi_n < x\right\} = \bigcap_{n=1}^{\infty} \{\xi_n < x\} \in \mathfrak{F},$$

$$\left\{\inf_{n \geqq 1} \xi_n < x\right\} = \bigcup_{n=1}^{\infty} \{\xi_n < x\} \in \mathfrak{F}.$$

Оттук следва, че $\limsup \xi_n \in \mathfrak{S}$ и $\liminf \xi_n \in \mathfrak{S}$, понеже

$$\limsup \xi_n = \inf_{n \geqq 1} \left(\sup_{m \geqq n} \xi_m\right), \liminf \xi_n = \sup_{n \geqq 1} \left(\inf_{m \geqq n} \xi_m\right).$$

Да означим $\overline{\xi} = \limsup \xi_n$, $\underline{\xi} = \liminf \xi_n$ и да разгледаме множеството

$$\{\overline{\xi} = \underline{\xi}\} = \overline{\{\overline{\xi} \neq \underline{\xi}\}} = \overline{\{\overline{\xi} < \underline{\xi}\} \cup \{\overline{\xi} > \underline{\xi}\}}.$$

Следователно $\{\overline{\xi} = \underline{\xi}\} \in \mathfrak{F}$, понеже за всяка двойка случайни величини ξ и η са в сила съотношенията

$$\{\xi < \eta\} \in \mathfrak{F}, \quad \{\xi > \eta\} \in \mathfrak{F},$$

$$\{\xi \leqq \eta\} \in \mathfrak{F}, \quad \{\xi = \eta\} \in \mathfrak{F}.$$

Действително,

$$\{\xi < \eta\} = \bigcup_{r_k} \{\xi < r_k < \eta\} \in \mathfrak{F},$$

където $\{r_k\}$ е множеството на рационалните числа, а

$$\{\xi < r_k < \eta\} = \{\xi < r_k\} \cap \{\eta > r_k\} \in \mathfrak{F}.$$

Оттук лесно се вижда, че

$$\{\xi \leqq \eta\} = \Omega \setminus \{\xi > \eta\} \in \mathfrak{F}, \ \{\xi = \eta\} = \{\xi \leqq \eta\} \cap \{\xi \geqq \eta\} \in \mathfrak{F}.$$

Дефиниция 2. *Ако за всяко $\omega \in \Omega$ имаме*

$$\limsup \xi_n(\omega) = \liminf \xi_n(\omega) = \xi(\omega),$$

ще казваме, че редицата $\{\xi_n\}$ клони към случайната величина $\xi \in \mathfrak{S}$.

Понякога тази сходимост се нарича *поточкова.*

7) За всяка борелова функция $\varphi(x)$ и за всяка случайна величина $\xi \in \mathfrak{S}$ е в сила $\eta = \varphi(\xi) \in \mathfrak{S}$ (*т. е. борелова функция от случайна величина е случайна величина*).

Доказателството следва веднага от факта, че за всяко $B \in \mathfrak{B}_1$

$$\{\omega : \eta(\omega) \in B\} = \{\omega : \varphi(\xi(\omega)) \in B\} = \{\omega : \xi(\omega) \in \varphi^{-1}(B)\} \in \mathfrak{F},$$

тъй като $\varphi^{-1}(B) = \{x : \varphi(x) \in B\} \in \mathfrak{B}_1$. Оттук веднага се вижда, че ако $\xi \in \mathfrak{S}$, то например ξ^n, $\xi^+ = \max(\xi, 0)$, $\xi^- = \max(-\xi, 0)$, $|\xi| = \xi^+ - \xi^-$, $\lg \xi$ (при $\xi > 0$) и т. н. са също случайни величини.

Дефиниция 3. *Нека $\{A_k\}$ е разлагане на Ω, т. е. $\sum_{k=1}^{\infty} A_k = \Omega$, $A_k \in \mathfrak{F}$. Случайните величини, представими във вида*

$$\xi(\omega) = \sum_{k=1}^{\infty} x_k I_{A_k}(\omega), \tag{5}$$

се наричат дискретни (стъпаловидни) случайни величини. Ако в (5) събираемите са краен брой, то съответната случайна величина се нарича проста (елементарна). Класа на простите случайни величини ще означаваме с $\mathfrak{S}_0$.

Простите случайни величини приемат краен брой стойности съответно върху непресичащите се събития.

Теорема 2. *Класът $\mathfrak{S}$ на случайните величини, дефинирани в дадено вероятностно пространство $(\Omega, \mathfrak{F}, \mathbf{P})$, съвпада с класа S на измеримите функции, съдържащ елементарните случайни величини и затворен относно поточковата сходимост.*

Доказателство. Нека $\xi = \xi(\omega) \in \mathfrak{S}$. Полагаме

$$\xi_n(\omega) = \sum_{k=1-n^2}^{n^2} \frac{k-1}{n} I_{A_k}(\omega) + n I_{B_n}(\omega) - n I_{C_n}(\omega), \tag{6}$$

където

$$A_k = \{\omega : \frac{k-1}{n} \leqq \xi(\omega) < \frac{k}{n}\}, \quad B_n = \{\omega : \xi(\omega) \geqq n\}, \quad C_n = \{\omega : \xi(\omega) < -n\}.$$

Очевидно $\xi_n(\omega)$ е проста случайна величина, т. е. $\xi \in \mathfrak{S}_0$, тъй като приема краен брой стойности. Ще трябва да покажем, че за всяко $\omega \in \Omega$ $\xi_n(\omega) \to \xi(\omega)$ при $n \to \infty$, т. е. за всяко $\omega \in \Omega$ и всяко $\varepsilon > 0$ съществува $N < \infty$, така че за всички $n \geqq N$ имаме $|\xi_n(\omega) - \xi(\omega)| < \varepsilon$.

Нека $\omega \in \Omega$ и ε е произволно положително число. Избираме N така, че $\frac{1}{N} < \varepsilon$. Тогава съществува такова k_0, че

$$-N^2 + 1 \leqq k_0 \leqq N^2 \quad \text{и} \quad \frac{k_0 - 1}{N} \leqq \xi(\omega) < \frac{k_0}{N}.$$

Съгласно (6) това означава, че $\xi_N(\omega) = \dfrac{k_0 - 1}{N}$ и следователно $\xi(\omega) - \xi_N(\omega) < \dfrac{1}{N} < \varepsilon$. Сега за всяко $n \geqq N$ съществува k, такова че $1 - n^2 \leqq k \leqq n^2$ и $\dfrac{k-1}{n} \leqq \xi(\omega) \leqq \dfrac{k}{n}$. Съгласно (6) $\xi_n(\omega) = \dfrac{k-1}{n}$. Тогава $\xi(\omega) - \xi_n(\omega) < \dfrac{1}{n} < \varepsilon$.

И така, показахме, че всяка случайна величина може да се представи като граница на елементарни случайни величини, което доказва теоремата. От (6) непосредствено се вижда, че ако $\xi \geqq 0$, то $C_n = \varnothing$, $n = 1, 2, \ldots$ Следователно $\xi_n \uparrow \xi$, т. е. *всяка неотрицателна случайна величина може да се представи като граница на монотонно растяща редица от неотрицателни прости случайни величини.*

Изобщо от (6) следва $|\xi_n| \leqq |\xi|$.

§ 3. Функция на разпределение

Както видяхме в предишните два параграфа, разнообразието от случайни величини е голямо, като стойностите им могат да са краен брой, изброимо или неизброимо много, могат да са разположени дискретно или да запълват цели интервали, дори цялата реална права. За единно вероятностно характеризиране на толкова различни по своя характер случайни величини в теорията на вероятностите се въвежда важното понятие *функция на разпределение.*

Дефиниция 1. *За всички $x \in \mathbf{R}^1$ вероятността случайната величина $\xi \in \mathfrak{S}(\Omega, \mathfrak{F}, \mathbf{P})$ да вземе стойности, по-малки от x, разглеждана като функция на x, се нарича функция на разпределение на ξ:*

$$F_\xi(x) = \mathbf{P}(\{\omega : \xi(\omega) < x\}) = \mathbf{P}\{\xi < x\}. \tag{1}$$

По-нататък (когато няма опасност от недоразумение) ще пишем само $F(x)$.

С помощта на (1) можем да пресмятаме вероятностите и на по-сложни събития. Например

$$\mathbf{P}\{a \leqq \xi < b\} = F(b) - F(a), \tag{2}$$

което следва от представянето

$$\{\xi < b\} = \{\xi < a\} + \{a \leqq \xi < b\}$$

и от аксиома **3°** за адитивността.

Теорема 1. *Всяка функция на разпределение има следните свойства:*

1^0) *за всяко* $x \in \mathbf{R}^1$ *е изпълнено* $0 \leqq F(x) \leqq 1$;

2^0) *за* $x_1 < x_2$ $(x_1, x_2 \in \mathbf{R}^1)$ *е в сила* $F(x_1) \leqq F(x_2)$ (***монотонност***);

3^0) *за всяка редица* $x_n \in \mathbf{R}^1$, $x_n \uparrow x_0 \in \mathbf{R}^1$, *следва* $\lim\limits_{x_n \to x_0} F(x_n) = F(x_0)$ (***непрекъснатост отляво***);

4^0) $\lim\limits_{x \to -\infty} F(x) = 0$, $\lim\limits_{x \to +\infty} F(x) = 1$.

Доказателство. Свойство 1^0) следва от факта, че $F(x)$ е вероятност, а свойство 2^0) — от монотонността на вероятността $\mathbf{P}$, тъй като $\{\xi < x_1\} \subset \{\xi < x_2\}$.

От $x_1 < x_2 < \cdots < x_n < \cdots \to x_0$ следва, че $B_n = \{x_n \leqq \xi < x_0\} \downarrow \varnothing$ при $n \to \infty$. Тогава от аксиома **4°** за непрекъснатостта и (2) получаваме

$$\begin{aligned} 0 &= \lim_{n\to\infty} \mathbf{P}(B_n) = \lim_{n\to\infty} (F(x_0) - F(x_n)) \\ &= F(x_0) - \lim_{n\to\infty} F(x_n) = F(x_0) - F(x_0 - 0), \end{aligned}$$

т. е. $F(x_0) = F(x_0 - 0)$ (непрекъснатост отляво).

Тъй като $\{\omega : \xi(\omega) < \infty\} = \Omega$, то $F(+\infty) = \mathbf{P}\{\xi < +\infty\} = 1$. От друга страна, $\Omega = \sum\limits_{k=-\infty}^{+\infty} A_k$, където $A_k\{k-1 \leqq \xi < k\}$. Следователно

$$\begin{aligned} 1 = \mathbf{P}(\Omega) &= \sum_{k=-\infty}^{+\infty} \mathbf{P}(A_k) = \lim_{n\to\infty} \sum_{k=1-n}^{n} \mathbf{P}(A_k) \\ &= \lim_{n\to\infty} \sum_{k=1-n}^{n} [F(k) - F(k-1)] = \lim_{n\to\infty} (F(n) - F(-n)) \\ &= F(+\infty) - F(-\infty), \end{aligned}$$

откъдето получаваме $F(-\infty) = 0$.

З а б е л е ж к а. Понякога вместо (1) се използва дефиницията

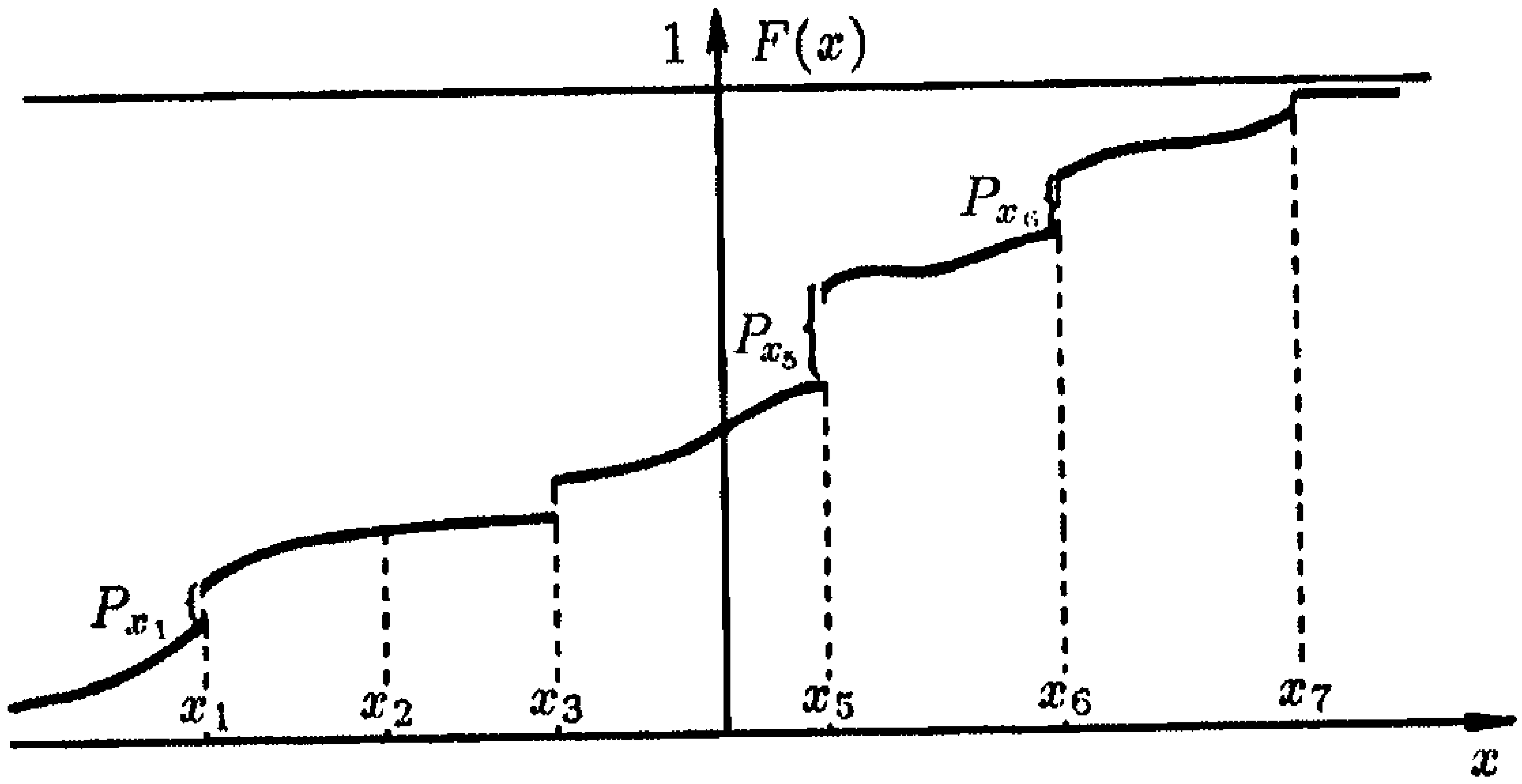
(1*) $$F_\xi(x) = \mathbf{P}\{\xi \leqq x\}.$$

Тогава теорема 1 е в сила с единствената разлика, че вместо условието 3^0) (непрекъснатост отляво) е изпълнено условието

3^*) *за всяка редица* $x_n \in \mathbf{R}^1$, $x_n \downarrow x_0 \in \mathbf{R}^1$ *е в сила* $\lim\limits_{n\to\infty} F(x_n) = F(x_0)$ (*непрекъснатост отдясно*).

В условията на дефиниция (1) доказахме, че $F(x) = F(x-0)$. Нека сега $x_n \downarrow x_0$. Тъй като редицата $F(x_n)$ е монотонна и ограничена, тя е сходяща, т. е. $\lim\limits_{n\to\infty} F(x_n) = F(x+0)$. В общия случай $p_x = F(x+0) - F(x) \geqq 0$. Ако $F(x)$ е непрекъсната в точката x, то $p_x = 0$.

Точките, в които $p_x > 0$, ще наричаме *точки на скок* (*точки на прекъсване*). В такава точка графиката на функцията $F(x)$ има скок с големина p_x (фиг. 5).



Фиг. 5. Функция на разпределение със скокове

Функцията на разпределение $F(x)$ може да има не повече от изброимо много скокове (респ. точки на скок).

Наистина функцията на разпределение $F(x)$ може да има не повече от 1 скок, по-голям от $\frac{1}{2}$; скоковете, при които $\frac{1}{4} < p_x \leqq \frac{1}{2}$, са не повече от 3; при $\frac{1}{8} < p_x \leqq \frac{1}{4}$ те са не повече от 7. В общия случай функцията на разпределение $F(x)$ може да има не повече от $2^n - 1$ такива скока, че $\frac{1}{2^n} < p_x \leqq \frac{1}{2^{n-1}}$, $n = 1, 2, \ldots$

Ясно е, че всички скокове могат да бъдат номерирани, т.е. те са не повече от изброимо много.

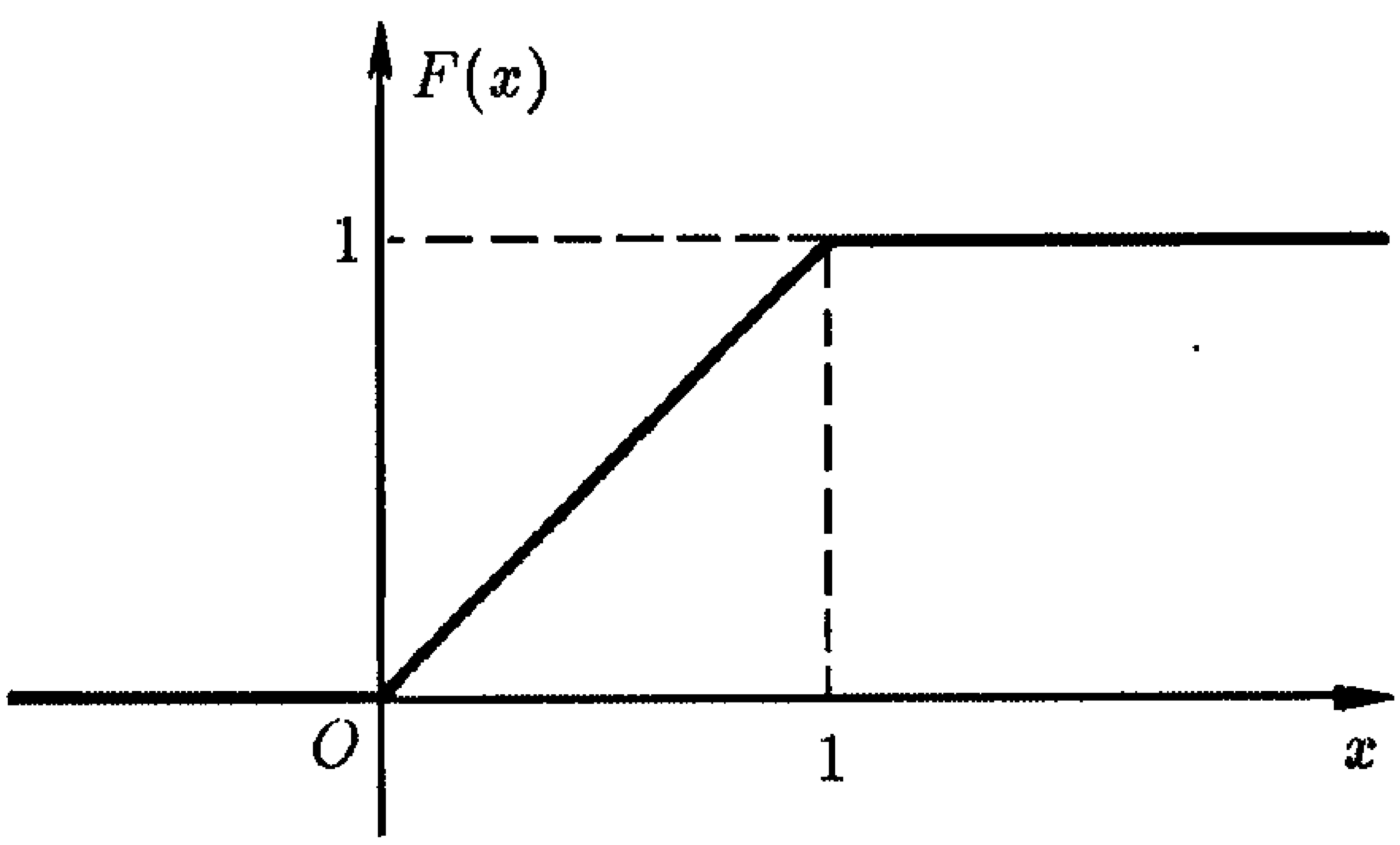
И така, на всяка случайна величина по формула (1) съпоставяме функция на разпределение. Обратното не е вярно — различни случайни величини от едно и също вероятностно пространство могат да имат една и съща функция на разпределение.

Пример 1. Нека $\Omega = [0,1]$ и $\mathfrak{F} = \mathfrak{B}_1 \cap [0,1]$, т. е. $\mathfrak{F}$ се състои от борелови подмножества в интервала $[0,1]$. За всяко $A \in \mathfrak{F}$ имаме $\mathbf{P}(A) = \lambda(A)$, където λ е лебеговата мярка (дължината) върху реалната права (вж. 1.5). За всяко $\omega \in [0,1]$ разглеждаме случайните величини $\xi(\omega) = \omega$ и $\eta(\omega) = 1-\omega$. Очевидно това са различни случайни величини. При $0 \leqq x \leqq 1$ директно пресмятаме

$$\begin{aligned}
\mathbf{P}\{\xi < x\} &= \mathbf{P}\{\omega < x\} = \lambda([0,x]) = x,\\
\mathbf{P}\{\eta < x\} &= \mathbf{P}\{1-\omega < x\} = \mathbf{P}\{\omega > 1-x\}\\
&= 1 - \mathbf{P}\{\omega \leqq 1-x\} = 1 - \lambda([0,1-x])\\
&= 1-(1-x) = x.
\end{aligned}$$

Следователно случайните величини ξ и η имат една и съща функция на разпределение (фиг. 6).

$$F(x) = \begin{cases} 0 & \text{за } x \leqq 0,\\ x & \text{за } 0 \leqq x \leqq 1,\\ 1 & \text{за } x \geqq 1. \end{cases}$$



Фиг. 6. Равномерно разпределение

За случайна величина, имаща тази функция на разпределение, се казва, че е ***равномерно разпределена в интервала*** $(0,1)$, като се означава $\xi \in \mathcal{U}(0,1)$ (фиг. 6).

В теорията на вероятностите в повечето случаи се работи не директно със случайните величини, а с техните функции на разпределение. Ако е зададена една функция на разпределение, тя не определя еднозначно случайната величина, а цял клас случайни величини, имащи дадената функция на разпределение.

В много случаи обаче ще видим, че от теоретико-вероятностна гледна точка случайни величини, имащи една и съща функция на разпределение, е уместно да считаме за неразлични.

Нека ξ е дискретна случайна величина (вж. дефиниция 2.3), т. е. $\xi = \sum_{k=1}^{\infty} x_k I_{A_k}$, $\sum_{k=1}^{\infty} A_k = \Omega$, където стойностите ѝ са подредени по големина $x_1 < x_2 < \cdots < x_n < \cdots$

Тогава, като означим

$$p_n = \mathbf{P}(A_n) = \mathbf{P}\{\omega : \xi(\omega) = x_n\}, \tag{3}$$

за функцията на разпределение получаваме

$$F(x) = \mathbf{P}\{\xi < x\} = \mathbf{P}\left(\sum_{x_k < x} A_k\right) = \sum_{x_k < x} \mathbf{P}\{\xi = x_k\} = \sum_{x_k < x} p_k. \tag{4}$$

Сега от (4) непосредствено се проверява, че при $x \uparrow x_m$

$$\lim F(x) = F(x_m) = p_1 + p_2 + \cdots + p_{m-1},$$

а при $x \downarrow x_m$

$$\lim F(x) = F(x_m + 0) = p_1 + p_2 + \cdots + p_m,$$

т. е. големината на скока на $F(x)$ в точката x_m е $F(x_m+0) - F(x_m) = p_m$.

И така функцията на разпределение на дискретна случайна величина е напълно зададена, ако са дадени нейните стойности, както и съответните вероятности, с които тя приема тези стойности. Често това се представя във вид на таблица и се нарича *дискретно разпределение*:

Стойности	x_1	x_2	$\dots$	x_n	$\dots$
Вероятности	p_1	p_2	$\dots$	p_n	$\dots$

където освен това са в сила условията

$$p_n \geqq 0, \quad \sum_{n=1}^{\infty} p_n = 1. \tag{5}$$

Пример 2. Нека ξ е сумата от точките при две последователни хвърляния на зар.

В този случай вероятностите (3) непосредствено се пресмятат по формулата за класическата вероятност и разпределението има следния вид:

k	2	3	4	5	6	7	8	9	10	11	12
p_k	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Друг важен клас случайни величини са тези, за които съществува неотрицателна функция $f_\xi(t)$, такава че за всяко $x \in \mathbf{R}^1$

$$F_\xi(x) = \int_{-\infty}^{x} f_\xi(t)dt. \tag{6}$$

Такива случайни величини се наричат *непрекъснати*, а функцията $f_\xi(t)$ — *плътност на разпределението* (или накратко *плътност*). От (6) следва, че функцията на разпределение $F(x)$ е абсолютно непрекъсната и $F'_\xi(x) = f_\xi(x)$.

Очевидно

$$\mathbf{P}\{x_1 \leqq \xi < x_2\} = F_\xi(x_2)\} - F_\xi(x_1) = \int_{x_1}^{x_2} f_\xi(t)dt,$$

където

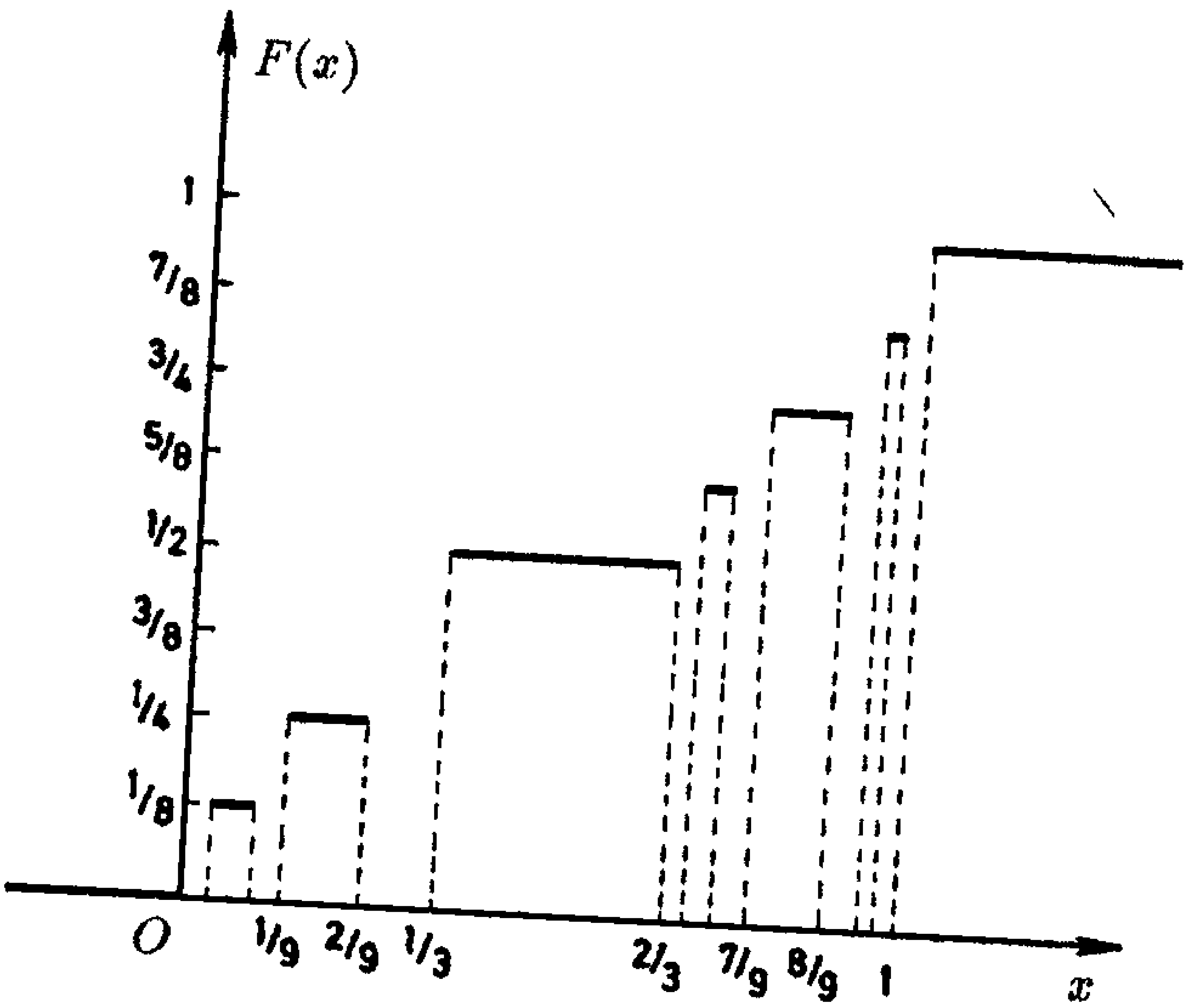
$$f_\xi(t) \geqq 0, \quad \int_{-\infty}^{+\infty} f_\xi(t)dt = 1. \tag{7}$$

Както ще видим по-късно, свойствата (6) и (7) (съответно (4) и (5)) в известен смисъл са характеризиращи, т. е. ако съществуват функции с такива свойства, могат да се намерят случайни величини, на които те са съответно функция на разпределение и плътност. (Понякога (5) се нарича *дискретна плътност*.)

Освен дискретни и абсолютно непрекъснати разпределения (както и смес от тях) съществуват и т. нар. *сингулярни разпределения*. За тях функцията на разпределение $F(x)$ е непрекъсната, но не абсолютно непрекъсната (т.е. няма плътност). Типичен пример е известната канторова функция на разпределение, която може да се дефинира с равенствата

$$F(x) = \begin{cases} 0 & \text{за } x \leqq 0, \\ \dfrac{F(3x)}{2} & \text{за } 0 \leqq x \leqq \dfrac{1}{3}, \\ \dfrac{1}{2} & \text{за } \dfrac{1}{3} \leqq x \leqq \dfrac{2}{3}, \\ \dfrac{1}{2} + \dfrac{F(3x-2)}{2} & \text{за } \dfrac{2}{3} \leqq x \leqq 1, \\ 1 & \text{за } x \geqq 1. \end{cases}$$

Графиката на тази функция е представена на фиг. 7.



Фиг. 7. Разпределение на Кантор

Очевидно $F(x)$ е непрекъсната, но $F'(x) = 0$ почти навсякъде, ъй като сумарната дължина на интервала, в който функцията

$F(x)$ не нараства, е

$$\sum_{k=0}^{\infty} \frac{1}{3}\left(\frac{2}{3}\right)^k = 1,$$

т. е. $F(x)$ расте върху множество с мярка нула, но без скокове.

§ 4. Функции на разпределение и вероятностни мерки в борелово пространство

В предишния параграф видяхме, че всяка функция на разпределение, дефинирана с (3.1), притежава свойствата $1^0) - 4^0)$ от теорема 3.1. Оказва се, че тези свойства са характеризиращи, т.е. определят еднозначно функцията на разпределение в смисъл на следната

Теорема 1. *Всяка реална функция $F(x)$, $x \in \mathbf{R}^1$, със свойствата $1^0) - 4^0)$ е функция на разпределение на някаква случайна величина, дефинирана в подходящо вероятностно пространство.*

Тази теорема ще докажем с помощта на следващата теорема, която установява връзка между вероятностите в бореловото измеримо пространство $(\mathbf{R}^1, \mathfrak{B}_1)$ и функциите $F(x)$ със свойствата $1^0) - 4^0)$.

Теорема 2 (теорема за съответствие). *Съотношението*

$$Q(I_x) = F(x), \tag{1}$$

където $I_x = (-\infty, x)$, установява взаимно еднозначно съответствие между функциите $F(x)$, дефинирани със свойствата $1^0) - 4^0)$, и вероятностните мерки Q в бореловото измеримо пространство $(\mathbf{R}^1, \mathfrak{B}_1)$.

Доказателство. В едната посока твърдението е частен случай на теорема 3.1, в смисъл че във вероятностното пространство $(\mathbf{R}^1, \mathfrak{B}_1, Q)$ от (1), което е равносилно на (3.1), следва, че $F(x)$ има свойствата $1^0) - 4^0)$.

Обратно, нека $F(x)$ има свойствата $1^0) - 4^0)$ от теорема 3.1. Ще покажем, че съотношението (1) определя вероятност в $(\mathbf{R}^1, \mathfrak{B}_1)$.

Нека J е класът на интервалите в $\mathbf{R}^1$ (крайни, безкрайни, отворени, затворени, полуотворени). В J дефинираме функция Q

чрез равенствата

$$
(2) \quad \begin{aligned} Q([a,b)) &= F(b) - F(a), \qquad & Q([a,b]) &= F(b+0) - F(a), \\ Q((a,b)) &= F(b) - F(a+0), \qquad & Q((a,b]) &= F(b+0) - F(a+0). \end{aligned}
$$

Нека $\mathfrak{B}_0$ е минималната булова алгебра, породена от J. Не е трудно да се съобрази, че $\mathfrak{B}_0$ е съвкупността на всички крайни обединения на непресичащи се елементи от J, т.е. за всяко $A \in \mathfrak{B}_0$ съществуват $C_k \in J$, $k = 1, 2, \ldots, n$, така че

$$
(3) \qquad A = \sum_{k=1}^{n} C_k, \quad C_i \cap C_j = \varnothing \quad \text{за } i \neq j.
$$

Продължаваме еднозначно Q върху $\mathfrak{B}_0$ чрез полагането

$$
(4) \qquad Q(A) = \sum_{k=1}^{n} Q(C_k).
$$

Сега ще проверим, че (4) дефинира вероятност в $(\mathbf{R}^1, \mathfrak{B}_0)$, т. е. че Q удовлетворява аксиомите 1° – 4° от 1.3.

Непосредствено от (4) се вижда, че за Q са в сила аксиомите 1° – 3°, т. е. Q е неотрицателна, нормирана и адитивна мярка. Остава да покажем, че тя е непрекъсната, т. е. аксиома 4°.

Първо ще покажем, че за всяко $A \in \mathfrak{B}_0$ функцията $Q(A)$ може да се апроксимира отдолу с компактни множества, т. е.

$$
(5) \qquad Q(A) = \sup Q(S), \quad S \subset A, \quad S = \sum_{k=1}^{n} I_k, \quad I_k \in G,
$$

където G е съвкупността на затворените интервали $[a,b] \in \mathbf{R}^1$, $G \subset \mathfrak{B}_0$.

Действително, от непрекъснатостта отляво на $F(x)$ следва, че за всеки интервал $C_k \in J$ съществува $I_k \in G$, $I_k \subset C_k$, така че

$$
Q(C_k) \leqq Q(I_k) + \frac{\varepsilon}{n}, \quad k = 1, \ldots, n.
$$

Следователно

$$
\sum_{k=1}^{n} Q(I_k) \leqq \sum_{k=1}^{n} Q(C_k) \leqq \sum_{k=1}^{n} Q(I_k) + \varepsilon,
$$

т. е. $Q(S) \leqq Q(A) \leqq Q(S) + \varepsilon$, което доказва (5).

Нека $A_n \in \mathfrak{B}_0$ и $A_n \downarrow \varnothing$. Да означим с G_0 съвкупността на компактните множества в $\mathbf{R}^1$. Тогава от (5) следва, че за всяко $n \geqq 1$ и за всяко $\varepsilon > 0$ съществува $S_n \in G_0$, $S_n \subset A_n$, така че

$$Q(A_n) \leqq Q(S_n) + \frac{\varepsilon}{2^n}. \tag{6}$$

Тъй като $\bigcap_{n=1}^{\infty} S_n \subset \bigcap_{n=1}^{\infty} A_n = \varnothing$, от компактността на множествата S_n следва, че съществува крайно N, така че $\bigcap_{n=1}^{N} S_n = \varnothing$. Тогава

$$\Omega = \overline{\bigcap_{n=1}^{N} S_n} = \bigcup_{n=1}^{N} \overline{S}_n,$$

$$A_N = A_N \cap \Omega = A_N \cap \left(\bigcup_{n=1}^{N} \overline{S_n} \right) = \bigcup_{n=1}^{N} A_N \overline{S_n} \subset \bigcup_{n=1}^{N} A_n \overline{S}_n,$$

тъй като $A_N \subset A_n$, $n = 1, 2, \ldots, N$. Следователно (вж. свойства 2) и 7) от § 3 на първа глава)

$$Q(A_N) \leqq Q\left(\bigcup_{n=1}^{N} (A_n \setminus S_n) \right) \leqq \sum_{n=1}^{N} Q(A_n \setminus S_n)$$

$$= \sum_{n=1}^{N} [Q(A_n) - Q(S_n)] \leqq \sum_{n=1}^{\infty} [Q(A_n) - Q(S_n)] \leqq \varepsilon,$$

където сме използвали и неравенствата (6).

Показахме, че за всяко $\varepsilon > 0$ съществува крайно N, така че при $n \geqq N$ имаме $Q(A_n) \leqq Q(A_N) < \varepsilon$, т. е. $Q(A_n) \to 0$ при $n \to \infty$. Тогава по теоремата за продължение на вероятностите (вж. теорема 1.3.2) вероятността Q може да бъде еднозначно продължена върху σ-алгебрата $\mathfrak{B}_1$.

Доказателство на теорема 1. Нека Q е вероятността в $(\mathbf{R}^1, \mathfrak{B}_1)$ съгласно теорема 2. Във вероятностното пространство $(\mathbf{R}^1, \mathfrak{B}_1, Q)$ разглеждаме случайната величина $\xi(\omega) = \omega$, $\omega \in \mathbf{R}^1$. Тогава $F(x)$ е функция на разпределение на ξ, както се вижда от следните равенства:

$$Q\{\omega : \xi(\omega) < x\} = Q\{\omega < x\} = Q(I_x) = F(x)$$

съгласно теорема 2.

§ 5. Многомерни функции на разпределение

Нека $\xi_1(\omega)$, ..., $\xi_n(\omega)$ са случайни величини в дадено вероятностно пространство $(\Omega, \mathfrak{F}, \mathbf{P})$. Тъй като за всяко $x_k \in \mathbf{R}^1$ имаме $\{\omega : \xi_k(\omega) < x_k\} \in \mathfrak{F}$, $k = 1, \ldots, n$, то сечението им също е събитие, т.е. $\bigcap_{k=1}^{n} \{\omega : \xi_k(\omega) < x_k\} \in \mathfrak{F}$. Следователно може да бъде дефинирана вероятността на това събитие, разглеждана като функция на $(x_1, x_2, \ldots, x_n)$, която е прието да се нарича *многомерна функция на разпределение*

$$F_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_n) = \mathbf{P}\left(\bigcap_{k=1}^{n} \{\omega : \xi_k(\omega) < x_k\}\right) \tag{1}$$

$$\equiv \mathbf{P}\{\xi_1 < x_1, \ldots, \xi_n < x_n\}.$$

Понякога $\xi(\omega) = (\xi_1(\omega), \ldots, \xi_n(\omega))$ се нарича *случаен вектор* и се дефинира като такова изображение на Ω в $\mathbf{R}^n$, че за всяко $B \in \mathfrak{B}_n$ да имаме

$$\xi^{-1}(B) = \{\omega : \xi(\omega) \in B\} \in \mathfrak{F}.$$

Функцията, дефинирана с (1), се нарича *съвместно разпределение* на случайните величини $\xi_1, \ldots, \xi_n$. Понякога индексите се изпускат и се пише само $F(x_1, \ldots, x_n)$.

Теорема 1. *Всяка съвместна функция на разпределение, дефинирана с (1), има следните свойства:*

1) *за всяко* $x = (x_1, \ldots, x_n) \in \mathbf{R}^n$ *имаме* $0 \leqq F(x_1, \ldots, x_n) \leqq 1$;

2) $F(x_1, \ldots, x_n)$ *е ненамаляваща функция по всеки от аргументите си;*

3) $F(x_1, \ldots, x_n)$ *е непрекъсната отляво по всеки от аргументите си;*

4) $F(x_1, \ldots, x_n) \to 0$, *когато поне един от аргументите* $x_k \to -\infty$ *и* $F(x_1, \ldots, x_n) \to 1$, *когато всички аргументи* $x_k \to \infty$, $k = 1, 2, \ldots, n$, *т. е.* $F(x_1, \ldots, x_{k-1}, -\infty, x_{k+1}, \ldots, x_n) = 0$ *и* $F(+\infty, \ldots, +\infty) = 1$;

5) *нека* $a_k < b_k$, $k = 1, \ldots, n$, *и* $q_{ij\ldots s} = F(\alpha_1, \ldots, \alpha_n)$, *където* $\alpha_i = a_i$, $\alpha_j = a_j$, ..., $\alpha_s = a_s$, *а всички останали аргументи са*

$a_t = b_t$. *Тогава*

$$\Delta = F(b_1, b_2, \ldots, b_n) - \sum_{i=1}^{n} q_i + \sum_{i<j} q_{ij} - \sum_{i<j<k} q_{ij\ldots k} + \cdots \tag{2}$$
$$+ (-1)^n F(a_1, a_2, \ldots, a_n) \geqq 0,$$

където $\sum\limits_{i<j} = \sum\limits_{i=1}^{n-1} \sum\limits_{j=i+1}^{n}$, $\sum\limits_{i<j<k} = \sum\limits_{i=1}^{n-2} \sum\limits_{j=i+1}^{n-1} \sum\limits_{k=j+1}^{n}$ *и т. н.*

Доказателство. Свойствата 1) – 4) са аналогични на тези в едномерния случай (вж. теорема 3.1) и следват непосредствено от дефиницията (1).

Да отбележим, че

$$F(x_1, \ldots, x_{k-1}, -\infty, x_{k+1}, \ldots, x_n)$$
$$= \mathbf{P}\ \{\xi_1 < x_1, \ldots, \xi_{k-1} < x_{k-1}, \varnothing, \xi_{k+1} < x_{k+1}, \ldots, \xi_n < x_n\} = 0,$$

$$F_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_{k-1}, +\infty, x_{k+1}, \ldots, x_n)$$
$$= \mathbf{P}\{\xi_1 < x_1, \ldots, \xi_{k-1} < x_{k-1}, \Omega, \xi_{k+1} < x_{k+1}, \ldots, \xi_n < x_n\}$$
$$= F_{\xi_1, \ldots, \xi_{k-1}, \xi_{k+1}, \ldots, \xi_n}(x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_n),$$

т. е. получихме функция на разпределение на $(n-1)$-мерен вектор.

Специално

$$F_{\xi_1, \ldots, \xi_n}(+\infty, \ldots, +\infty, x, +\infty, \ldots, +\infty) = \mathbf{P}\{\xi_k < x\} = F_{\xi_k}(x)$$

се нарича *маргинално разпределение* на случайната величина ξ_k.

За да докажем (2), ще покажем, че

$$\Delta = \mathbf{P}\{a_1 \leqq \xi_1 < b_1, a_2 \leqq \xi_2 < b_2, \ldots, a_n \leqq \xi_n < b_n\}, \tag{3}$$

с което се доказва свойство 5).

Да означим $A_k = \{\xi_k < a_k\}$, $B_k = \{\xi_k < b_k\}$, $A = \bigcup\limits_{k=1}^{n} A_k$, $B = \bigcap\limits_{k=1}^{n} B_k$. Очевидно $A_k \subset B_k$ и

$$\mathbf{P}\{a_k \leqq \xi_k < b_k,\ k = 1, 2, \ldots, n\} = \mathbf{P}\left(\bigcap_{k=1}^{n} \overline{A_k} B_k\right)$$
$$= \mathbf{P}(\overline{A}B) = \mathbf{P}(B \setminus AB) = \mathbf{P}(B) - \mathbf{P}(AB) \tag{4}$$
$$= F(b_1, \ldots, b_n) - \mathbf{P}\left(\bigcup_{k=1}^{n} C_k\right),$$

където

$C_k = A_k B = \{\xi_1 < b_1, \ldots, \xi_{k-1} < b_{k-1}, \xi_k < a_k, \xi_{k+1} < b_{k+1}, \ldots, \xi_n < b_n\}.$

Не е трудно да се докаже по индукция, че за всяка редица от събития $\{D_k\}$ е в сила представянето (вж. зад. 1.12):

$$\mathbf{P}\left(\bigcup_{k=1}^{n} D_k\right) = \sum_{k=1}^{n} \mathbf{P}(D_k) - \sum_{i<j} \mathbf{P}(D_i D_j) \tag{5}$$
$$+ \sum_{i<j<k} \mathbf{P}(D_i D_j D_k) + \cdots + (-1)^{n-1} \mathbf{P}\left(\bigcap_{k=1}^{n} D_k\right).$$

Действително при $n = 2$ формула (5) е доказаното свойство силна адитивност в 1.3, т.е.

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(AB).$$

От тази формула по индукционното предположение (5) последователно получаваме

$$\mathbf{P}\left(\bigcup_{k=1}^{n+1} D_k\right) = \mathbf{P}\left(\bigcup_{k=1}^{n} D_k\right) + \mathbf{P}(D_{n+1}) - \mathbf{P}\left(\bigcup_{k=1}^{n} D_k D_{n+1}\right)$$
$$= \sum_{k=1}^{n} \mathbf{P}(D_k) + \mathbf{P}(D_{n+1}) - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbf{P}(D_i D_j) + \cdots + (-1)^{n-1} \mathbf{P}\left(\bigcap_{k=1}^{n} D_k\right)$$
$$- \left[\sum_{i=1}^{n} \mathbf{P}(D_i D_{n+1}) + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbf{P}(D_i D_j D_{n+1}) + \cdots + (-1)^{n-1} \mathbf{P}\left(\bigcap_{k=1}^{n+1} D_k\right)\right]$$
$$= \sum_{k=1}^{n+1} \mathbf{P}(D_k) - \sum_{i=1}^{n} \sum_{j=i+1}^{n+1} \mathbf{P}(D_i D_j) + \cdots + (-1)^{n} \mathbf{P}\left(\bigcap_{k=1}^{n+1} D_k\right),$$

т. е. доказахме (5) за $n + 1$.

Тъй като $\mathbf{P}(C_i) = q_i$, $\mathbf{P}(C_i C_j) = q_{ij}$, $\mathbf{P}(C_i C_j C_k) = q_{ijk}$ и т. н., $\mathbf{P}\left(\bigcap_{k=1}^{n} C_k\right) = F(a_1, \ldots, a_n)$, то от (4), прилагайки формула (5), получаваме (3).

Свойствата 1) – 4) са характеризиращи, т. е. ако някаква реална функция $F(x_1, \ldots, x_n)$ има свойствата 1) – 4), то съществува n-мерен вектор $\xi = (\xi_1, \ldots, \xi_n)$, дефиниран в подходящо вероятностно пространство $(\Omega, \mathfrak{F}, \mathbf{P})$, така че $F(x_1, \ldots, x_n)$ е негова функция на разпределение (в смисъл на (1)). Нещо повече, в сила е следният важен резултат:

Теорема 2 (фундаментална теорема на Колмогоров за съгласуваните вероятности). *Нека функциите $F_n(x_1, \ldots, x_n)$, където $n = 1, 2, \ldots$, са определени от условията 1) – 4) и освен това удовлетворяват следното условие за съгласуваност:*

$$F_{n+m}(x_1, \ldots, x_n, +\infty, \ldots, +\infty) = F_n(x_1, \ldots, x_n)$$

за всички n, $m = 1, 2, \ldots$

Тогава съществува вероятностно пространство $(\Omega, \mathfrak{F}, \mathbf{P})$, в което може да се дефинира безкрайна редица от случайни величини $\xi_1, \xi_2, \ldots, \xi_n, \ldots$, така че за всяко $n = 1, 2, \ldots$ съвместното разпределение на $\xi_1, \ldots, \xi_n$ е

$$F_n(x_1, \ldots, x_n) = \mathbf{P}\{\xi_1 < x_1, \ldots, \xi_n < x_n\}.$$

Очевидно доказаната теорема 4.1 се получава като частен случай от тази теорема при $n = 1$. Ние няма да привеждаме доказателството на теорема 2, тъй като то излиза извън рамките на един първоначален курс по теория на вероятностите. Любознателният читател може да намери нейното доказателство например в [12, 17, 18, 20].

Аналогично на едномерния случай *съвместната дискретна плътност* на дискретни случайни величини $\xi_1, \ldots, \xi_n$ се дефинира с равенството

$$(6) \qquad f_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n) = \mathbf{P}(\xi_1 = x_1, \ldots, \xi_n = x_n).$$

Ще казваме, че случайните величини $\xi_1, \ldots, \xi_n$ имат *абсолютно непрекъснато разпределение*, ако съществува такава функция $f_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n)$, че

$$(7) \quad F_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} f_{\xi_1, \ldots, \xi_n}(t_1, \ldots, t_n) dt_1 \ldots dt_n$$

за всяка n-орка от реални числа $(x_1, \ldots, x_n)$.

Функцията $f_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n)$ се нарича *съвместна плътност* на случайните величини $\xi_1, \ldots, \xi_n$. Очевидно

$$(8) \qquad f(x_1, \ldots, x_n) \geqq 0, \quad \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} f(t_1, \ldots, t_n) dt_1 \ldots dt_n = 1.$$

От (7) следва, че за всяка точка $(x_1, \ldots, x_n)$ в n-мерното евклидово пространство, в която $f(x_1, \ldots, x_n)$ е непрекъсната, имаме

(9) $$\frac{\partial^n}{\partial x_1 \ldots \partial x_n} F(x_1, \ldots, x_n) = f(x_1, \ldots, x_n).$$

В (8) и (9) за по-кратко сме изпуснали индексите $\xi_1, \ldots, \xi_n$, което ще правим и по-нататък, когато това не води до недоразумение.

С конкретни примери на многомерни разпределения ще се запознаем в трета и пета глава.

§ 6. Независимост на случайни величини

Понятието независимост на събития, въведено в 1.6, се пренася естествено и за случайни величини.

Дефиниция 1. *Случайните величини ξ и η, дефинирани във вероятностното пространство $(\Omega, \mathfrak{F}, \mathbf{P})$, се наричат независими, ако*

(1) $$\mathbf{P}\{\xi < x,\, \eta < y\} = \mathbf{P}\{\xi < x\} \cdot \mathbf{P}\{\eta < y\}$$

за всички $x, y \in \mathbf{R}^1$.

Равенството (1) можем да запишем във вида

(2) $$F_{\xi,\,\eta}(x, y) = F_\xi(x) \cdot F_\eta(y).$$

Дефиниция 2. *Случайните величини $\xi_1, \ldots, \xi_n$, дефинирани в едно и също пространство $(\Omega, \mathfrak{F}, \mathbf{P})$, се наричат независими две по две, ако (1) (или равносилното му (2)) е в сила за всеки две от тях.*

Дефиниция 3. *Случайните величини $\xi_1, \ldots, \xi_n$, дефинирани в $(\Omega, \mathfrak{F}, \mathbf{P})$, се наричат независими в съвкупност (или накратко независими), ако*

(3) $$\mathbf{P}\{\xi_1 < x_1, \ldots, \xi_n < x_n\} = \prod_{m=1}^{n} \mathbf{P}\{\xi_m < x_m\}$$

за всички $x_m \in \mathbf{R}^1$, $m = 1, \ldots, n$.

З а б е л е ж к а. Тъй като в (3) можем да полагаме $x_i = +\infty$ за някои i, от (1) следва

$$\mathbf{P}\{\xi_{i_1} < x_{i_1}, \ldots, \xi_{i_m} < x_{i_m}\} = \prod_{k=1}^{m} \{\mathbf{P}\{\xi_{i_k} < x_{i_k}\}$$

за всяка редица $1 \leqq i_1 \leqq \ldots \leqq i_m \leqq n$.

Очевидно от независимостта на $\xi_1, \ldots, \xi_n$ следва независимостта им две по две. Обратното не е вярно.

Теорема 1. *Ако случайните величини $\xi_1, \ldots, \xi_n$, дефинирани в дадено пространство $(\Omega, \mathfrak{F}, \mathbf{P})$, са независими, то за всички борелови множества $B_k \in \mathfrak{B}_1$, $k = 1, 2, \ldots, n$, имаме*

$$\mathbf{P}\{\xi_1 \in B_1, \ldots, \xi_n \in B_n\} = \prod_{k=1}^{n} \mathbf{P}\{\xi_k \in B_k\}. \tag{4}$$

Доказателство. От (3) следва, че (4) е вярно за интервалите от вида $I_k = (-\infty, x_k)$, $k = 1, \ldots, n$. Нека $\mathfrak{A}_1$ е съвкупността на множествата B_1, за които (4) е в сила. Не е трудно да се покаже, че $\mathfrak{A}_1$ е σ-булова алгебра. Тъй като $\mathfrak{B}_1$ се поражда от интервалите $I_1 = (-\infty, x_1)$, очевидно $\mathfrak{B}_1 \subset \mathfrak{A}_1$, т. е. (4) е вярно за всяко борелово множество B_1. Аналогични разсъждения са валидни за всяко от множествата B_k, $k = 2, \ldots, n$.

Следствие. *Ако $\xi_1, \ldots, \xi_n$ са независими случайни величини, а $g_1(x), g_2(x), \ldots, g_n(x)$ — борелови функции, случайните величини $g_1(\xi_1), g_2(\xi_2), \ldots, g_n(\xi_n)$ са също независими.*

Доказателство. Нека $A_k = \{x : g_k(x) < y_k\}$. Тъй като $g_k(x)$ е борелова функция, то $A_k \in \mathfrak{B}_1$. Тогава от (4) получаваме

$$\begin{aligned}\mathbf{P}\left(\bigcap_{k=1}^{n} \{g_k(\xi_k) < y_k\}\right) &= \mathbf{P}\left(\bigcap_{k=1}^{n} \{\xi_k \in A_k\}\right) \\ &= \prod_{k=1}^{n} \mathbf{P}\{\xi_k \in A_k\} = \prod_{k=1}^{n} \mathbf{P}\{g_k(\xi_k) < y_k\}.\end{aligned}$$

Теорема 2. *Ако случайните величини $\xi_1, \ldots, \xi_n$ имат съвместно абсолютно непрекъснато разпределение, то необходимо и достатъчно условие за независимост е съвместната им плътност да бъде*

$$f_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{\xi_i}(x_i). \tag{5}$$

Доказателството следва непосредствено от дефинициите (5.7) и (3).

Задачи

1. Нека една партида съдържа N изделия, от които M са дефектни ($M < N$). Прави се случайна извадка без връщане с обем $n\,(n \leq N)$ от партидата. Да означим с ξ броя на дефектните изделия в партидата. Покажете, че

$$p_k = \mathbf{P}\{\xi = k\} = \binom{M}{k}\frac{\binom{N-M}{n-k}}{\binom{N}{n}}, \quad \max(0, n - N + M) \leqq k \leqq \min(n, M)$$

(**хипергеометрично разпределение**).

Упътване. Вж. пример 1 от 1.4.

2. В зад. 1 намерете вероятността извадката да съдържа не повече от s бракувани изделия ($s \leq \min(n, M)$).

3. Върху единична окръжност с център в началото на координатната система се избира случайна точка, през която се прекарва допирателна към окръжността. Нека ξ е дължината на допирателната от точката до пресичането ѝ с оста Ox. Покажете, че $F_\xi(x) = \dfrac{2}{\pi}\operatorname{arc\,tg}(x)$, $x > 0$ и $F_\xi(x) = 0$ при $x \leq 0$ (**разпределение на Коши**). Определете плътността на разпределението.

4. Нека случайната величина ξ има функция на разпределение $F(x)$ и $F'(x) = f(x)$. Намерете функцията на разпределение на случайната величина $\eta = \xi^2$ и покажете, че

$$f_\eta(x) = \begin{cases} \dfrac{f(\sqrt{x}) + f(-\sqrt{x})}{2\sqrt{x}} & \text{при } x > 0\,, \\ 0 & \text{при } x < 0\,. \end{cases}$$

5. Нека $F_\eta(x)$ е непрекъсната функция на разпределение и $\eta = F_\xi(\xi)$. Покажете, че $\eta \in \mathcal{U}(0, 1)$, т. е. η е равномерно разпределена в интервала $[0, 1]$.

6. Нека $\xi \in \mathcal{U}(0, 1)$ и $\eta = \log\dfrac{1}{\xi}$. Покажете, че $F_\eta(x) = 0$ при $x < 0$ и $F_\eta(x) = 1 - e^{-x}$ при $x \geqq 0$ (**експоненциално разпределение с параметър 1**).

7. Покажете, че при $\lambda > 0$ съотношенията

$$p_n = \frac{\lambda^n e^{-n}}{n!}, \quad n = 0, 1, 2, \ldots,$$

определят дискретно разпределение (нарича се **разпределение на Поасон**).

8. Нека $p + q = 1$ и $0 < p < 1$. Докажете, че за всяко n равенствата

$$p_k = \binom{n}{k}p^k q^{n-k}, \quad k = 0, 1, \ldots, n,$$

задават дискретно разпределение (нарича се **биномно разпределение**).

9. Покажете, че при $0 < p < 1$ съотношенията

$$p_k = (1-p)^k p, \quad k = 0, 1, 2, \ldots,$$

дефинират дискретно разпределение (известно като **геометрично разпределение**).

10. Покажете, че

$$f_\xi(x) = \begin{cases} 0 & \text{за} \quad x < 0, \\ \dfrac{1}{\lambda} e^{-x/\lambda} & \text{за} \quad x \geqq 0 \end{cases}$$

е плътност на разпределение. Намерете съответната функция на разпределение (**експоненциално разпределение с параметър** λ; означава се $\mathrm{Ex}(\lambda)$).

11. Покажете, че функцията

$$f_\xi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\} \quad \text{за} \quad x \in \mathbf{R}^1$$

дефинира плътност на разпределение, където $-\infty < \mu < \infty$, $\sigma^2 > 0$. Намерете съответната функция на разпределение (**нормално разпределение с параметри** μ **и** σ^2; означава се $\xi \in \mathcal{N}(\mu, \sigma^2)$).

Трета глава

Независими опити. Схема на Бернули

§ 1. Произведение на вероятностни пространства

В предишните две глави въведохме понятията независимост на събития (§ 1.6) и независимост на случайни величини (§ 2.6). Всъщност те могат да бъдат обединени по следния начин:

Дефиниция 1. *Буловите алгебри (или σ-алгебри) от събития $\mathfrak{F}_1$, $\mathfrak{F}_2, \dots, \mathfrak{F}_n$ се наричат независими, ако за всички $A_k \in \mathfrak{F}_k$, $k = 1, \dots, n$, имаме*

$$\mathbf{P}\left(\bigcap_{k=1}^{n} A_k\right) = \prod_{k=1}^{n} \mathbf{P}(A_k). \tag{1}$$

Естествено тук се подразбира, че е зададено вероятностно пространство $(\Omega, \mathfrak{F}, \mathbf{P})$ и $\mathfrak{F}_k \subset \mathfrak{F}$, $k = 1, \dots, n$.

От дефиниция 1 и теорема 2.6.1 следва, че понятието независимост на случайни величини е еквивалентно на независимост на породените от тях σ-алгебри (вж. § 2.6).

В приетата от нас аксиоматика на всеки опит съпоставихме вероятностно пространство. Нека сега разгледаме n опита, т. е. да считаме, че са зададени n вероятностни пространства

$$(\Omega_1, \mathfrak{F}_1, \mathbf{P}_1), \dots, (\Omega_n, \mathfrak{F}_n, \mathbf{P}_n). \tag{2}$$

Ако тези вероятностни пространства са модели на причинно независими опити, то σ-алгебрите $\mathfrak{F}_1, \dots, \mathfrak{F}_n$ трябва да са независими. Но за да можем да говорим изобщо за теоретико–вероятностна независимост, трябва да разглеждаме $\mathfrak{F}_k$ като σ-подалгебри на σ-алгебра $\mathfrak{F}$ от едно общо вероятностно пространство $(\Omega, \mathfrak{F}, \mathbf{P})$. Оказва се, че такова вероятностно пространство винаги може да бъде построено. Сега ще демонстрираме този факт в случая, когато вероятностните пространства (2) са крайни.

Нека $(\Omega_k, \mathfrak{F}_k, \mathbf{P}_k)$, $k = 1, \dots, n$ е крайно вероятностно пространство, т. е. $\Omega_k = \{\omega_k\}$, $\#(\Omega_k) < \infty$, $\mathfrak{F}_k = \mathfrak{B}(\Omega_k)$ — множеството от всички подмножества на Ω_k, а вероятността $\mathbf{P}_k(A) = \sum_{\omega_k \in A} p_k(\omega_k)$

се задава с вероятностите на елементарните събития $p_k(\omega_k) \geqq 0$, $\sum\limits_{\omega_k \in \Omega_k} p_k(\omega_k) = 1$.

Ще построим *декартовото произведение* $(\Omega, \mathfrak{F}, \mathbf{P})$ *на вероятностните пространства* (2). Да положим

$$\Omega = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_n.$$

Точките (елементарните събития) $\omega \in \Omega$ са вектори $\omega = (\omega_1, \ldots, \omega_n)$ с компоненти $\omega_i \in \Omega_i$, $i = 1, \ldots, n$, $\mathfrak{F} = \mathfrak{B}(\Omega)$ е множеството от всички подмножества на Ω, а

$$\text{(3)} \qquad \mathbf{P}(A) = \sum_{\omega \in A} p(\omega), \quad p(\omega) = \prod_{k=1}^{n} p_k(\omega_k), \quad A \in \mathfrak{F}.$$

Така дефинираната вероятност се нарича *декартово произведение на вероятностите* $\mathbf{P}_k$ и се означава

$$\mathbf{P} = \mathbf{P}_1 \times \mathbf{P}_2 \times \ldots \times \mathbf{P}_n.$$

Аналогично

$$\mathfrak{F} = \mathfrak{F}_1 \times \mathfrak{F}_2 \times \ldots \times \mathfrak{F}_n$$

се нарича *декартово произведение на алгебри.*

Събитията от вида

$$\text{(4)} \qquad A = A_1 \times A_2 \times \ldots \times A_n,$$

състоящи се от $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$, за които $\omega_k \in A_k$, $k = 1, 2, \ldots, n$, ще наричаме *правоъгълници.* От (3) следва

$$\text{(5)} \quad \mathbf{P}(A) = \sum_{\omega \in A} p(\omega) = \sum_{\omega_1 \in A_1} p_1(\omega_1) \ldots \sum_{\omega_n \in A_n} p_n(\omega_n) = \prod_{k=1}^{n} \mathbf{P}_k(A_k).$$

Правоъгълниците от вида

$$\text{(6)} \qquad A_i' = \Omega_1 \times \ldots \times \Omega_{i-1} \times A_i \times \Omega_{i+1} \times \ldots \times \Omega_n$$

понякога се наричат *цилиндри.* Да означим с $\mathfrak{F}_i'$ съответната им подалгебра на $\mathfrak{F}$. Очевидно между $A_i' \in \mathfrak{F}_i'$ и $A_i \in \mathfrak{F}_i$ съществува естествен изоморфизъм $A_i' \sim A_i$, така че вместо събитията A_i от вероятностното пространство $(\Omega_i, \mathfrak{F}_i, \mathbf{P}_i)$ можем да разглеждаме изоморфните събития A_i' от вероятностното пространство $(\Omega, \mathfrak{F}, \mathbf{P})$.

От (5) следва, че $\mathbf{P}(A_i') = \mathbf{P}_i(A_i)$. Тъй като

$$A = A_1 \times \ldots \times A_n = \bigcap_{k=1}^{n} A_k',$$

то от (5) за всички $A'_k \in \mathfrak{F}'_k$ получаваме

$$\mathbf{P}\left(\bigcap_{k=1}^{n} A'_k\right) = \prod_{k=1}^{n} \mathbf{P}(A'_k),$$

което показва (съгласно (1)), че алгебрите $\mathfrak{F}'_1, \ldots, \mathfrak{F}'_n$ са независими.

Това дава възможност да интерпретираме $(\Omega, \mathfrak{F}, \mathbf{P})$ като вероятностно пространство, съответстващо на опит, състоящ се от n независими опита.

§ 2. Схема на Бернули. Биномно разпределение

Частният случай на независими опити с два изхода при всеки опит може да се дефинира по следния начин:

Нека вероятностните пространства в (1.2) имат вида

$$\Omega_k = \{0,1\}, \quad \mathfrak{F}_k = \{\varnothing, \{0\}, \{1\}, \Omega_k\}, \quad \mathbf{P}_k(1) = p, \quad \mathbf{P}_k(0) = q = 1 - p.$$

Тогава съгласно (1.3) за декартовото произведение $(\Omega, \mathfrak{F}, \mathbf{P})$ имаме $\Omega = \{\omega\}$, $\omega = (\omega_1, \ldots, \omega_n)$, където $\omega_i = 0$ или 1,

$$p(\omega) = \prod_{i=1}^{n} p_i(\omega_i) = \prod_{i=1}^{n} p^{\omega_i} q^{1-\omega_i}. \tag{1}$$

Тази схема на независими опити се нарича *схема на Бернули* или *бернулиеви опити* и допуска следната интерпретация: Нека един от изходите A (който е прието да се нарича „успех“) да настъпва при всеки опит с една и съща вероятност p. Тогава противоположният му изход $\overline{A}$ (наричан „неуспех“) настъпва при всеки опит с вероятност $q = 1 - p$. В елементарното събитие $\omega = (\omega_1, \ldots, \omega_n)$ имаме $\omega_i = 1$, ако в резултат на i-тия опит е настъпил успех, и $\omega_i = 0$ при неуспех.

Да означим $\nu_n = \omega_1 + \cdots + \omega_n$. Очевидно ν_n е случайна величина и може да се интерпретира като брой на успехите в редица от n бернулиеви опита.

Да означим с $B_k = \{\omega : \nu_n(\omega) = k\}$ събитието да се сбъднат точно k успеха. Тъй като от (1) следва, че при $\omega \in B_k$ имаме $p(\omega) = p^k q^{n-k}$, то $\mathbf{P}(B_k) = p^k q^{n-k} \#(B_k)$. Не е трудно да се съобрази, че броят на елементарните събития $\omega \in B_k$ е $\#(B_k) = C_n^k$, (т. е. броят на комбинациите на n елемента от k-ти клас). Следователно

(2) $$p_k = \mathrm{P}\{\nu_n = k\} = C_n^k p^k q^{n-k}, \quad k = 0,\ 1,\ 2,\ \ldots,\ n.$$

Вероятностите (2) определят *биномно разпределение*, а за случайната величина ν_n се казва, че е *биномно разпределена* и се означава $\nu_n \in Bi(n,p)$ (чете се: „биномно разпределение с параметри n и p"). Понякога в (2) се използва означението $p_k = b(k; n, p)$.

За да проверим, че (2) е дискретно разпределение (вж. условия (1.3.5)), достатъчно е да отбележим, че

$$\sum_{k=0}^{n} p_k = \sum_{k=0}^{n} C_n^k p^k q^{n-k} = (p+q)^n = 1.$$

Оттук веднага се вижда, че p_k е коефициентът пред x^k в развитието на бинома $(px+q)^n$.

При фиксирани n и p от (2) получаваме

$$\frac{p_k}{p_{k-1}} = 1 + \frac{(n+1)p - k}{kq}.$$

Следователно $p_k \geqq p_{k-1}$ тогава и само тогава, когато $k \leqq (n+1)p$.

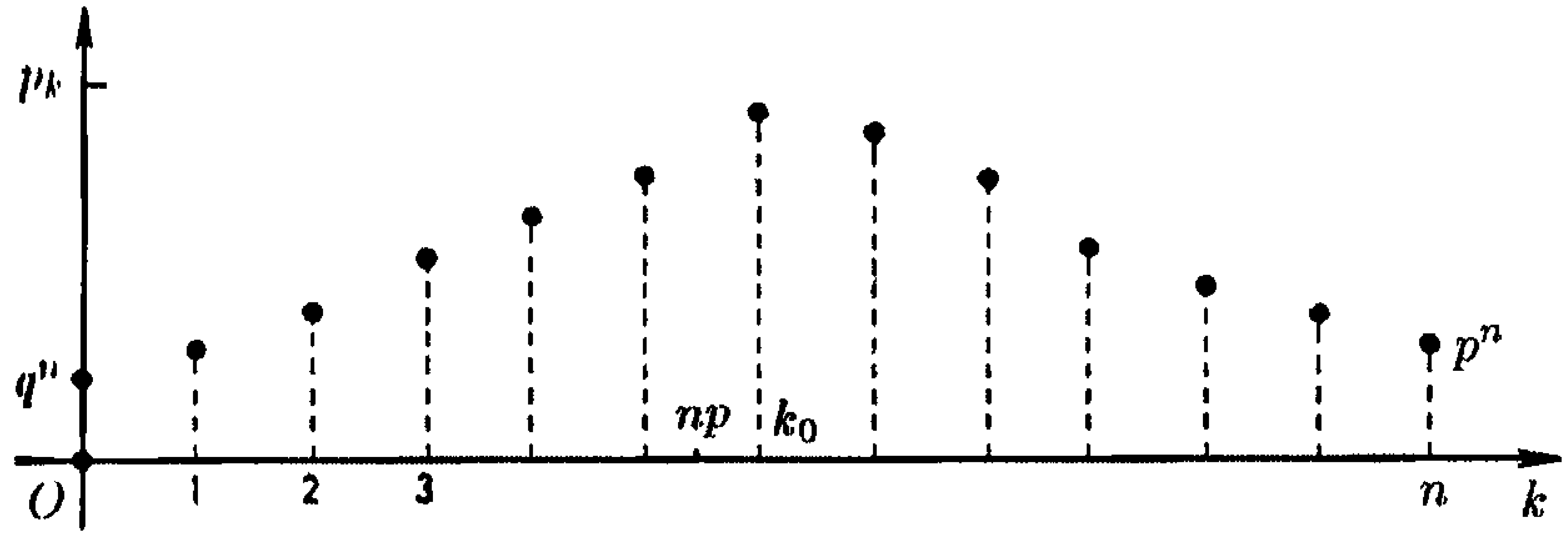
Аналогично от

$$\frac{p_{k+1}}{p_k} = 1 + \frac{(n+1)p - 1 - k}{(k+1)q}$$

следва, че $p_{k+1} \leqq p_k$ тогава и само тогава, когато $k \geqq (n+1)p - 1$.

По този начин окончателно получаваме $p_{k-1} \leqq p_k \geqq p_{k+1}$ тогава и само тогава, когато $(n+1)p - 1 \leqq k \leqq (n+1)p$.

Ако $(n+1)p$ не е цяло число, то редицата (2) има единствен максимален член с номер k_0, равен на цялата част на $(n+1)p$, т. е. $k_0 = [(n+1)p]$. Ако $k_0 = [(n+1)p] = (n+1)p$, то редицата (2) има два максимални члена с номера $k_0 = (n+1)p$ и $k_0 - 1 = (n+1)p - 1 = np - q$. Очевидно $-q \leqq k_0 - np \leqq p$ и $|k_0 - np| < 1$, т. е. можем да считаме, че най-вероятните стойности на биномните вероятности са групирани около np. Както ще видим в следващата глава, np е математическото очакване (важна интегрална характеристика) на случайната величина ν_n.



Фиг. 8. Биномно разпределение

Примерна графика на биномните вероятности (2) е дадена на фиг. 8.

При $0 \leqq x \leqq n$ за функцията на разпределение на случайната величина ν_n имаме

$$F(x) = \mathbf{P}\{\nu_n \leqq x\} = \sum_{k=0}^{m} \binom{n}{k} (1-q)^k q^{n-k} = f(q), \tag{3}$$

където $m = [x]$. Като диференцираме (3) по q, получаваме

$$f'(q) = \binom{n}{m} (n-m)(1-q)^m q^{n-m-1}.$$

Оттук, като вземем предвид, че $f(0) = 0$, и интегрираме, окончателно намираме

$$F(x) = \binom{n}{m} (n-m) \int_0^q (1-t)^m t^{n-m-1} dt, \quad m = [x]. \tag{4}$$

Нека сега да разгледаме една безкрайна редица от бернулиеви опити. Тук основното пространство $\Omega = \{\omega\}$ се състои от всички $\omega = (\omega_1, \dots, \omega_n, \dots)$, където ω_i е 0 или 1. Нека ξ е броят на опитите, предшестващи настъпването на първия успех, т. е. ξ приписва на всяко елементарно събитие ω броя на нулите, намиращи се отляво на първата единица. Тогава

$$p_k = \mathbf{P}\{\xi = k\} = q^k p, \quad k = 0, 1, 2, \dots \tag{5}$$

От (5) веднага се вижда, че

$$\sum_{k=0}^{\infty} p_k = p \sum_{k=0}^{\infty} q^k = \frac{p}{1-q} = 1,$$

т. е. (5) е дискретно разпределение. То е прието да се нарича *геометрично разпределение.* От (5) лесно следва, че $p_k \downarrow 0$ при $k \to \infty$.

§ 3. Приближение на Поасон

В много практически важни случаи се налага да се разглежда схема на Бернули, когато n е голямо, а p — малко, т. е. $n \to \infty$, а $p \to 0$, така че $np \to \lambda > 0$. Оказва се, че в този случай възникват нови явления.

Теорема 1 (приближение на Поасон). *Разглеждаме серия от бернулиеви опити с дължина n и вероятност за успех във всеки опит p_n, т. е.*

$$p_k(n) = \mathbf{P}\{\nu_n = k\} = \binom{n}{k} p_n^k (1-p_n)^{n-k}, \quad n = 1, 2, \ldots$$

Ако $np_n \to \lambda > 0$ при $n \to \infty$, то за всяко фиксирано $k = 0, 1, 2, \ldots$

$$\lim_{n\to\infty} p_k(n) = \frac{\lambda^k}{k!} e^{-\lambda} = r_k. \tag{1}$$

Доказателство. По условие

$$p_n = \frac{\lambda}{n} + o\left(\frac{1}{n}\right) = \frac{\lambda}{n}(1 + o(1)) \quad \text{при } n \to \infty.$$

Следователно при $n \to \infty$

$$\begin{aligned} p_k(n) &= \binom{n}{k} p_n^k (1-p_n)^{n-k} \\ &= \frac{n(n-1)\cdots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right)^{n-k} (1+o(1))^k \\ &= \frac{(1-\frac{1}{n})(1-\frac{2}{n})\cdots(1-\frac{k-1}{n})}{k!} \lambda^k e^{-\lambda}(1+o(1)), \end{aligned}$$

откъдето се получава (1).

Тук сме използвали факта, че

$$\left(1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right)^n = e^{-\lambda}(1+o(1)),$$

което веднага следва от представянето

$$n\log\left(1-\frac{\lambda}{n}+o\left(\frac{1}{n}\right)\right)\sim n\left(-\frac{\lambda}{n}+o\left(\frac{1}{n}\right)\right)=-\lambda+o(1).$$

З а б е л е ж к а. Казваме, че $f(x)=o(g(x))$ при $x\to x_0$ (четем „o малко"), ако за всяко $\varepsilon>0$ съществува околност $U(x_0,\varepsilon)$, такава че $|f(x)|\leqq\varepsilon|g(x)|$ при $x\in U(x_0,\varepsilon)$.

Казваме, че $f(x)=O(g(x))$ при $x\to x_0$ (четем „O голямо"), ако съществуват околност $U(x_0)$ и константа $C>0$, за които $|f(x)|\leqq C|g(x)|$ за всички $x\in U(x_0)$.

Казваме, че $f(x)\sim g(x)$ (четем „$f(x)$ и $g(x)$ са еквивалентни") при $x\to x_0$, ако $\dfrac{f(x)}{g(x)}\to 1$ при $x\to x_0$.

От горните дефиниции следва, че при $x\to x_0$

$$\frac{o(g(x))}{g(x)}\to 0,\qquad \left|\frac{O(g(x))}{g(x)}\right|\leqq C.$$

Естествено, ако $f(x)=o(g(x))$, ще имаме и $f(x)=O(g(x))$.

От (1) лесно се проверява, че

$$\sum_{k=0}^{\infty}r_k=\sum_{k=0}^{\infty}\frac{\lambda^k e^{-\lambda}}{k!}=e^{-\lambda}e^{\lambda}=1,$$

т. е. (1) дефинира дискретно разпределение.

Казваме, че случайната величина η има ***поасоново разпределение с параметър*** $\lambda>0$ (което се записва $\eta\in Po(\lambda)$), ако

$$(2)\qquad r_k=\mathbf{P}\{\eta=k\}=\frac{e^{-\lambda}\lambda^k}{k!},\qquad k=0,\,1,\,2,\,\ldots$$

От (2) имаме $\dfrac{r_{k-1}}{r_k}=\dfrac{k}{\lambda}$ и следователно $r_{k-1}\leqq r_k$ тогава и само тогава, когато $k\leqq\lambda$. Аналогично $r_k\geqq r_{k+1}$ тогава и само тогава, когато $k\geqq\lambda-1$. Следователно, ако λ не е цяло число, то $k_0=[\lambda]$ е най-вероятната стойност на поасоновото разпределение, т. е. $r_{k_0}=\max r_k$; ако λ е цяло число, то най-вероятните стойности са λ и $\lambda-1$, т. е. $r_\lambda=r_{\lambda-1}=\max r_k$.

Граничното съотношение (1) дава основание поасоновото разпределение често да се интерпретира като „разпределение на редките събития". Много реални явления се описват чрез поасоново разпределение. Едно от тях е радиоактивното разпадане, т. е. ако $\xi(t)$ е броят на α-частиците, които изпуска радиоактивен радий за време t, оказва се, че $\xi(t) \in Po(\lambda t)$. Поасоново разпределени също са броят на клетките с изменени хромозоми при рентгеново облъчване, броят на заявките в телефонни централи и т. н.

§ 4. Теореми на Моавър – Лаплас и Бернули

Тъй като схемата на Бернули служи като стохастичен модел на много реални явления, значителен интерес представлява задачата за намиране на асимптотични формули за пресмятане на биномните вероятности (2.2) при големи n и k, а също на вероятностите

$$\mathbf{P}\{l_1 \leqq \nu_n \leqq l_2\} = \sum_{i=l_1}^{l_2} C_n^i p^i q^{n-i}, \tag{1}$$

които зависят вече от четири параметъра n, p, l_1, l_2 .

Теорема 1 (локална теорема на Моавър – Лаплас). *Ако в схемата на Бернули $\sigma = \sqrt{npq} \to \infty$, то за всяко $C > 0$ равномерно по всички $|x| \leqq C$ от вида $x = \dfrac{k - np}{\sigma}$, където k е цяло число, е в сила*

$$\mathbf{P}\left\{\frac{\nu_n - np}{\sqrt{npq}} = x\right\} = \frac{1}{2\pi\sigma^2} e^{-x^2/2}(1 + o(1)). \tag{2}$$

Доказателство. Нека $k = np + \sigma x$. Тогава

$$\mathbf{P}\{\nu_n = k\} = \mathbf{P}\left\{\frac{\nu_n - np}{\sigma} = x\right\} = \frac{n!}{k!(n-k)!} p^k q^{n-k}, \tag{3}$$

$$\log \mathbf{P}\{\nu_n = k\} = \log n! - \log k! - \log(n-k)! + k \log p + (n-k) \log q.$$

При $n \to \infty$ ще използваме следната формула на Стирлинг:

$$\log n! = n \log n + \log \sqrt{2\pi n} - n + \alpha_n, \tag{4}$$

където $\alpha_n = O\left(\dfrac{1}{n}\right)$.

Да означим $m = n - k = nq - \sigma x$. Тъй като

$$k = np\left(1 + \frac{qx}{\sigma}\right) \to \infty \quad \text{и} \quad m = nq\left(1 - \frac{px}{\sigma}\right) \to \infty,$$

от условията на теоремата и от равенствата (3) и (4) получаваме

$$\log \mathbf{P}\{\nu_n = k\} \tag{5}$$

$$= n\log n - k\log k - m\log m + k\log p + m\log q + \frac{1}{2}\log\frac{n}{2\pi k m} + \beta_n,$$

където $\beta_n = \alpha_n - \alpha_k - \alpha_m = O\left(\frac{1}{\sigma^2}\right)$.

Тъй като $\log(1+z) = O(z)$ при $z \to 0$, то $\log\frac{n}{km} = \log\frac{1}{\sigma^2} + O\left(\frac{1}{\sigma}\right)$ и от (5) намираме

$$\log \mathbf{P}\{\nu_n = k\} = -k\log\frac{k}{np} - m\log\frac{m}{nq} + \log\frac{1}{\sqrt{2\pi\sigma^2}} + O\left(\frac{1}{\sigma}\right) \tag{6}$$

$$= -(np+\sigma x)\log\left(1+\frac{xq}{\sigma}\right) - (nq-\sigma x)\log\left(1-\frac{xp}{\sigma}\right) + \log\frac{1}{\sqrt{2\pi\sigma^2}} + O\left(\frac{1}{\sigma}\right).$$

Като имаме предвид, че $\log(1\pm z) = \pm z - \frac{z^2}{2} + O(z^3)$ при $z \to 0$, от (6) получаваме

$$\log \mathbf{P}\{\nu_n = k\} = -(np + x\sigma)\left(\frac{xq}{\sigma} - \frac{x^2q^2}{2\sigma^2} + O\left(\frac{1}{\sigma^3}\right)\right)$$

$$- (nq - x\sigma)\left(-\frac{xp}{\sigma} + \frac{x^2p^2}{2\sigma^2} + O\left(\frac{1}{\sigma^3}\right)\right) + \log\frac{1}{\sqrt{2\pi\sigma^2}} + O\left(\frac{1}{\sigma}\right)$$

$$= -\frac{x^2}{2} + \log\frac{1}{\sqrt{2\pi\sigma^2}} + O\left(\frac{1}{\sigma}\right),$$

което доказва (2).

За приближено пресмятане на формула (1) се използва следната

Теорема 2 (интегрална теорема на Моавър – Лаплас). *При $\sigma = \sqrt{npq} \to \infty$ равномерно по $-\infty < a < b < +\infty$ е изпълнено*

$$\lim \mathbf{P}\left\{a \leqq \frac{\nu_n - np}{\sqrt{npq}} \leqq b\right\} = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-x^2/2}dx. \tag{7}$$

Доказателство. Нека $[x]$ е най-голямото цяло число, за което $[x] \leqq x$, а $]x[$ е най-малкото цяло число, за което $x \leqq]x[$. Да означим

$$l_1 =]np + a\sqrt{npq}[\,, \quad l_2 = [np + b\sqrt{npq}].$$

Тогава

$$\mathbf{P}\left\{a \leqq \frac{\nu_n - np}{\sqrt{npq}} \leqq b\right\} = \sum_{k=l_1}^{l_2} \mathbf{P}\{\nu_n = k\} = \sum_{a \leqq x_k \leqq b} \mathbf{P}\left\{\frac{\nu_n - np}{\sqrt{npq}} = x_k\right\}, \tag{8}$$

където $x_k = \dfrac{k - np}{\sqrt{npq}}$. Тогава $\Delta x_k = x_{k+1} - x_k = \dfrac{1}{\sigma}$ и от (8), прилагайки локалната гранична теорема, получаваме

$$\mathbf{P}\left\{a \leqq \frac{\nu_n - np}{\sqrt{npq}} \leqq b\right\} = \sum_{a \leqq x_k \leqq b} \frac{1}{\sqrt{2\pi}} e^{-x_k^2/2} \Delta x_k \left(1 + O\left(\frac{1}{\sigma}\right)\right).$$

В дясната страна на горното равенство стои интегрална сума, която при $\sigma \to \infty$ клони равномерно по a и b към интеграла $\dfrac{1}{\sqrt{2\pi}} \int\limits_a^b \exp\left(-\dfrac{x^2}{2}\right) dx$, което доказва (7).

З а б е л е ж к а. Теорема 2 е изпълнена и в по-общия случай $-\infty \leqq a < b \leqq \infty$. Сега няма да доказваме този факт, тъй като в девета глава ще докажем много по-силен резултат (т. нар. централна гранична теорема), откъдето ще следва теоремата на Моавър – Лаплас.

Както вече бе отбелязано, в схемата на Бернули случайната величина ν_n се интерпретира като брой на успехите (т. е. сбъдвания на събитие A) при n независими опита с вероятност за успех $p = \mathbf{P}(A)$ при всеки опит. Тогава $\dfrac{\nu_n}{n}$ е честотата на събитието A.

В началото посочихме, че в аксиоматичните основи на теорията на вероятностите са положени някои свойства на стабилните (устойчивите) честоти. Естествено е сега да си зададем въпроса, дали в рамките на тази теория при големи n честотата $\dfrac{\nu_n}{n}$ е близка до вероятността p. Отговор на този въпрос дава следната

Теорема 3 (теорема на Бернули). *За всяко $\varepsilon > 0$ имаме*

$$\lim_{n \to \infty} \mathbf{P}\left\{\left|\frac{\nu_n}{n} - p\right| > \varepsilon\right\} = 0. \tag{9}$$

Доказателство. Нека $C > 0$ е произволна константа. Тогава за всички достатъчно големи n имаме

$$\mathbf{P}\left\{\left|\frac{\nu_n}{n} - p\right| \leqq \varepsilon\right\} = \mathbf{P}\left\{-\frac{\varepsilon n}{\sqrt{npq}} \leqq \frac{\nu_n - np}{\sqrt{npq}} \leqq \frac{\varepsilon n}{\sqrt{npq}}\right\}$$

$$= \mathbf{P}\left\{-\frac{\varepsilon\sqrt{n}}{\sqrt{pq}} \leqq \frac{\nu_n - np}{\sqrt{npq}} \leqq \frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\right\} \geqq \mathbf{P}\left\{-C \leqq \frac{\nu_n - np}{\sqrt{npq}} \leqq C\right\}.$$

Към последното съотношение можем да приложим теорема 2, откъдето следва, че

$$\lim_{n\to\infty} \mathbf{P}\left\{\left|\frac{\nu_n}{n} - p\right| \leqq \varepsilon\right\} \geqq \frac{1}{\sqrt{2\pi}} \int_{-C}^{C} e^{-x^2/2} dx.$$

При $C \to \infty$, получаваме

$$\lim_{n\to\infty} \mathbf{P}\left\{\left|\frac{\nu_n}{n} - p\right| \leqq \varepsilon\right\} \geqq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx = 1, \tag{10}$$

което доказва (9), тъй като

$$\mathbf{P}\left\{\left|\frac{\nu_n}{n} - p\right| > \varepsilon\right\} = 1 - \mathbf{P}\left\{\left|\frac{\nu_n}{n} - p\right| \leqq \varepsilon\right\}$$

Да въведем означенията

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt. \tag{11}$$

Очевидно функциите (11) са дефинирани за $-\infty < x < +\infty$ и $\varphi(x) > 0$. Лесно се показва, че

$$\Phi(+\infty) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2/2} dt = 1.$$

Действително

$$\Phi^2(+\infty) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varphi(x)\varphi(y) dx dy = \frac{1}{2\pi} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}(x^2 + y^2)\right\} dx dy,$$

откъдето чрез смяна в полярни координати $x = r\cos\theta$, $y = r\sin\theta$ лесно се пресмята, че $\Phi^2(+\infty) = 1$.

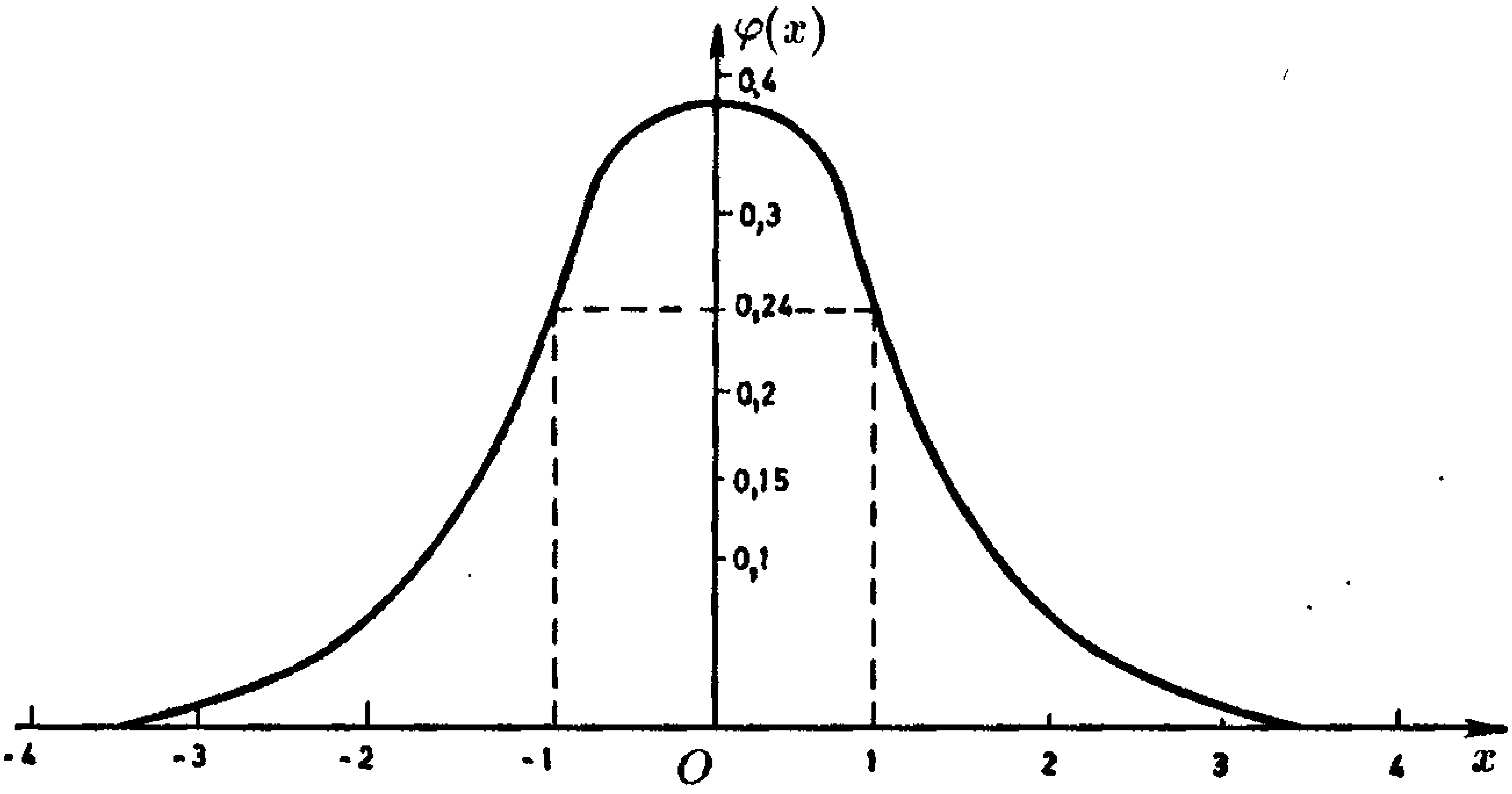
Съгласно (2.3.6) и (2.3.7) функциите $\varphi(x)$ и $\Phi(x)$ са съответно плътността и функцията на разпределение на непрекъснатата случайна величина ξ, която обикновено се нарича *гаусова*, а разпределението ѝ — *гаусово или стандартно нормално.*

По-общо ще казваме, че случайната величина η е разпределена *нормално* или *гаусово* с параметри $-\infty < a < +\infty$ и $\sigma^2 > 0$, ако има плътност (гаусова, нормална)

$$\varphi_\eta(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-a)^2/2\sigma^2} \quad \text{при } -\infty < x < \infty. \tag{12}$$

Очевидно $\int\limits_{-\infty}^{\infty} \varphi_\eta(x)dx = \int\limits_{-\infty}^{\infty} \varphi(x)dx = 1$, т. е. (12) е наистина плътност. Използва се означението $\eta \in \mathcal{N}(a, \sigma^2)$. Тогава $\xi \in \mathcal{N}(0,1)$ и не е трудно да се провери, че $\dfrac{\eta - a}{\sigma} \in \mathcal{N}(0,1)$.

Графиката на $\varphi(x)$ е симетрична камбановидна крива, показана на фиг. 9 при различен мащаб по двете оси.



Фиг. 9. Стандартно нормално разпределение

Не е трудно да се съобрази, че графиката на общата гаусова плътност (12) ще бъде отместена спрямо тази на фиг. 9, т. е. $\max\limits_{-\infty<x<\infty} \varphi_\eta(x) = \varphi_\eta(a) = \dfrac{1}{\sqrt{2\pi\sigma^2}}$. Очевидно параметрите a и σ^2

определят съответно положението на максимума и неговата големина, както и инфлексните точки $a - \sigma$ и $a + \sigma$. При това колкото σ^2 е по-малко, толкова максимумът е по-голям и кривата е „посъбрана" около точката a. В следващата глава ще видим, че тези „характеризационни" параметри имат и друга интерпретация. Същото се отнася за параметрите np и $\sigma^2 = npq$ при биномното разпределение.

Теоремите на Моавър – Лаплас и Бернули се оказват валидни (както ще видим по-късно в осма и девета глава) в много по-обща ситуация. За да поясним това, на получените резултати ще придадем малко по-друг смисъл.

При означенията от § 2 дефинираме случайните величини ξ_k по следния начин:

$$\xi_k(\omega) = \xi_k(\omega_1, \ldots, \omega_k, \ldots, \omega_n) = \begin{cases} 1 & \text{при } \omega_k = 1, \\ 0 & \text{при } \omega_k = 0, \end{cases} \tag{13}$$

т. е. $\xi_k = 1$, ако при k-тия опит е настъпил „успех", и $\xi_k = 0$ в противоположния случай.

Очевидно случайните величини ξ_k, $k = 1, \ldots, n$, са независими и еднакво разпределени, т. е. $\mathbf{P}\{\xi_k = 1\} = p$ и $\mathbf{P}\{\xi_k = 0\} = q$, $k = 1, \ldots, n$.

От друга страна,

$$\nu_n = \xi_1 + \xi_2 + \cdots + \xi_n. \tag{14}$$

Тогава (9) и (7) могат да бъдат записани в следния вид:

$$\frac{\nu_n}{n} = \frac{1}{n} \sum_{k=1}^{n} \xi_k \xrightarrow[n \to \infty]{\mathbf{P}} p \tag{15}$$

(клони по вероятност),

$$\frac{\nu_n - np}{\sqrt{npq}} = \frac{\frac{1}{n} \sum_{k=1}^{n} \xi_k - p}{\sqrt{\frac{pq}{n}}} \xrightarrow[n \to \infty]{d} \zeta \in \mathcal{N}(0,1) \tag{16}$$

(клони по разпределение към гаусова случайна величина).

Граничното съотношение (15) е известно като *слаб закон за големите числа*, при това ограничението (13) е несъществено. С други думи, законът за големите числа гласи, че средното аритметично на независими и еднакво разпределени случайни величини (с крайно математическо очакване, за което ще стане въпрос в следващата глава) клони по вероятност към константа (естествено, свързана с разпределението на ξ_k).

Граничното съотношение (16), което дава скоростта на сходимост на средното аритметично към тази константа (която ще се окаже математическото очакване на случайната величина ξ_k), се нарича *централна гранична теорема.*

Това са два фундаментални резултата в теорията на вероятностите с многобройни приложения в науката и практиката, в което ще имаме възможност да се убедим нееднократно.

§ 5. Полиномно разпределение

По-сложна схема на n независими опита се получава, когато при всеки опит е възможно да настъпи 1 от m несъвместими събития. Сега вероятностните пространства в (1.2) имат следния вид: $(\Omega_k, \mathfrak{F}_k, \mathbf{P}_k)$, $k = 1, \ldots, n$, където $\Omega_k = \{1, 2, \ldots, m\}$, $\mathfrak{F}_k = \mathfrak{B}(\Omega_k)$, $\mathbf{P}_k(i) = p_i$, $\sum\limits_{i=1}^{m} p_i = 1$. Съгласно (1.3) за декартовото произведение $(\Omega, \mathfrak{F}, \mathbf{P})$ имаме $\Omega = \{\omega\}$, $\omega = (\omega_1, \ldots, \omega_n)$, където $\omega_k = i$, $i = 1, \ldots, m$ и

$$p(\omega) = \prod_{k=1}^{n} p_k(\omega_k) = \prod_{k=1}^{n} p_{\omega_k}. \tag{1}$$

Нека $\nu_i = \nu_i(\omega) = \#\{\omega_k : \omega_k = i,\ k = 1, \ldots, n\}$, $i = 1, \ldots, m$. Очевидно ν_i са случайни величини и $\nu_1 + \nu_2 + \cdots + \nu_m = n$. Следователно те са зависими. Ще намерим съвместното многомерно разпределение на случайния вектор $\nu = (\nu_1, \ldots, \nu_n)$. Да означим

$$B_{n_1, \ldots, n_m} = \left\{ \omega : \nu_1(\omega) = n_1, \ldots, \nu_m(\omega) = n_m,\ \sum_{i=1}^{m} n_i = n \right\}.$$

Тъй като при $\omega \in B_{n_1, \ldots, n_m}$ съгласно (1) е изпълнено $p(\omega) = \prod\limits_{k=1}^{n} p_k^{n_k}$, то от (1.3) намираме

$$\mathrm{P}(B_{n_1, \ldots, n_m}) = \#(B_{n_1, \ldots, n_m}) \prod_{k=1}^{m} p_k^{n_k}.$$

Като се имат предвид комбинаторните разсъждения от § 1.4, не е трудно да се съобрази, че броят на елементарните събития $\omega \in B_{n_1,\dots,n_m}$ е

$$\#(B_{n_1,\dots,n_m}) = P^n_{n_1,\dots,n_m} = \frac{n!}{n_1!\dots n_m!},$$

т. е. броят на пермутациите с повторение. Следователно

$$\mathbf{P}\{\nu_1 = n_1, \dots, \nu_m = n_m\} = \frac{n!}{n_1!\dots n_m!} p_1^{n_1} \dots p_m^{n_m}, \tag{2}$$

където $n_1 + n_2 + \dots + n_m = n$. Тъй като

$$\sum_{n_1+\dots+n_m=n} \frac{n!}{n_1!\dots n_m!} p_1^{n_1} \dots p_m^{n_m} = (p_1 + \dots + p_m)^n = 1,$$

то (2) дефинира многомерно дискретно разпределение, което се нарича *полиномно разпределение*. Очевидно при $m = 2$ се получава биномното разпределение на Бернули.

Задачи

1. Като допуснете, че вероятността за раждане на момиче и на момче е $\frac{1}{2}$, покажете, че в семейство с 6 деца вероятността да има:

а) 3 момичета е 0,3125;

б) не по-малко от 1 и не повече от 5 момчета е 0,96875.

2. Покажете, че минималният брой числа, които трябва да бъдат избрани от таблица за случайни числа така, че вероятността точно 3 от тях да завършват на 5 да бъде максимална, е 29.

Упътване. Всяка цифра на едно случайно число е равновероятна.

3. При игра на бридж покажете, че вероятността даден играч 3 пъти поред да не получи асо е 0,028.

4. При n независими опита с вероятност за успех p е реализиран точно един успех. Покажете, че вероятността този успех да се случи при предпоследния опит е $\frac{1}{n}$.

5. При n бернулиеви опита с вероятност за успех p във всеки от тях са настъпили точно два успеха. Покажете, че вероятността успехите да са реализирани при последователни опити е $\frac{2}{n}$.

6. От урна, съдържаща бяла, зелена и червена топка, се вадят с връщане 5 топки. Покажете, че вероятността зелената и червената топка да бъдат извадени поне по два пъти е $\frac{50}{243}$.

7. Каква е вероятността в група от 30 студенти никой да не се е родил през даден месец? Покажете, че тази вероятност, пресметната по точна формула, е 0,07351, а чрез приближението на Поасон е 0,08208.

8. Покажете, че правилна монета трябва да се хвърли поне 16 587 пъти така, че с вероятност, не по-малка от 0,99, честотата на появяване на герб да се отличава от $\frac{1}{2}$ с не повече от $\frac{1}{100}$.

9. Нека случайната величина ξ е геометрично разпределена (т. е. съгласно (2.5)). Докажете, че при $n = 1, 2, \dots$; $k = 0, 1, 2, \dots$

$$\mathbf{P}(\{\xi = k + n\} | \{\xi \geqq k\}) = \mathbf{P}\{\xi = n\}$$

(отсъствие на последействие).

10. Докажете обратното твърдение на зад. 9, т. е. нека ξ е случайна величина, множеството от възможните стойности на която са целите неотрицателни числа, със свойствата:

а) $0 < \mathbf{P}\{\xi = 0\} < 1$;

б) $\mathbf{P}(\{\xi = k + n\} | \{\xi \geqq k\}) = \mathbf{P}\{\xi = n\}$ за всички цели неотрицателни n и k.

Тогава ξ е геометрично разпределена с параметър $p = \mathbf{P}\{\xi = 0\}$.

З а б е л е ж к а. Задачи 9 и 10 показват, че свойството „без последействие" дава една характеризация на геометричното разпределение, т.е. определя го еднозначно сред дискретните разпределения.

Четвърта глава

Математическо очакване на случайни величини

§ 1. Математическо очакване на елементарни случайни величини

В предишните глави бе отбелязано, че считаме една случайна величина за зададена, ако знаем стойностите, които тя взема, както и вероятностите, с които тя приема тези стойности, т.е. ако знаем нейното разпределение. В много случаи се оказва достатъчно да знаем много по-малко, за да можем да характеризираме случайната величина. Една такава важна (интегрална) характеристика е математическото очакване или средната стойност на случайната величина. Както ще видим, това понятие не е ново, то всъщност представлява един функционал (интеграл) от случайните величини (т. е. измеримите функции), но за разлика от анализа той има определен „вероятностен" смисъл. За този смисъл ще поговорим по-късно, а сега да изградим строго формално това понятие.

Нека $\mathfrak{S}_0$ е съвкупността на елементарните случайните величини, дефинирани в дадено вероятностно пространство $(\Omega, \mathfrak{F}, \mathbf{P})$, т. е. $\xi \in \mathfrak{S}_0$ тогава и само тогава, когато $\xi = \sum\limits_{k=1}^{n} x_k I_{A_k}$, което означава, че ξ приема краен брой стойности $\{x_k\}$, $k = 1, 2, \ldots, n$, с вероятности

$$p_k = \mathbf{P}\{\xi = x_k\} = \mathbf{P}(A_k), \quad \sum_{k=1}^{n} p_k = 1.$$

Следните свойства следват непосредствено от дефиницията:

1) $\mathfrak{S}_0$ е линейно пространство, т. е. ако ξ, $\eta \in \mathfrak{S}_0$, то $\xi + \eta \in \mathfrak{S}_0$ и $c\xi \in \mathfrak{S}_0$ за всяко $c \in \mathbf{R}^1$;

2) $\mathfrak{S}_0$ е алгебра, т. е. освен свойство 1), за всеки ξ, $\eta \in \mathfrak{S}_0$ имаме $\xi\eta \in \mathfrak{S}_0$ и $\dfrac{\xi}{\eta} \in \mathfrak{S}_0$;

3) От ξ, $\eta \in \mathfrak{S}_0$ следва, че $\max(\xi,\eta) \in \mathfrak{S}_0$ и $\min(\xi,\eta) \in \mathfrak{S}_0$, а също така $|\xi| \in \mathfrak{S}_0$.

Дефиниция 1. *Нека $\xi \in \mathfrak{S}_0$, т. е. $\xi = \sum\limits_{k=1}^{n} x_k I_{A_k}$. Математическо очакване (средна стойност, математическа надежда) на случайната величина ξ наричаме функционала*

$$\mathbf{E}\xi = \sum_{k=1}^{n} x_k \mathbf{P}(A_k) = \sum_{k=1}^{n} x_k p_k. \tag{1}$$

Да разгледаме някои примери. Като частен случай от (1) за произволно събитие $A \in \mathfrak{F}$ получаваме $\mathbf{E}I_A = \mathbf{P}(A)$, т. е. вероятността на едно събитие може да се разглежда като математическо очакване на неговия индикатор.

Нека $p_k = \mathbf{P}(A_k) = \dfrac{1}{n}$. Тогава от (1) следва, че

$$\mathbf{E}\xi = \frac{(x_1 + \cdots + x_n)}{n},$$

т. е. математическото очакване е средното аритметично от стойностите на случайната величина ξ.

Дефиниция 1 допуска следната механична интерпретация: Да си представим, че в точките $x_1, \ldots, x_n$ са разположени съответни точкови маси $p_1, \ldots, p_n$; $\sum\limits_{k=1}^{n} p_k = 1$. Тогава $\mathbf{E}\xi = \sum\limits_{k=1}^{n} x_k p_k$ е център на тежестта на тази механична система.

Теорема 1. *Математическото очакване $\mathbf{E}$ е единственият нормиран, линеен, позитивен и непрекъснат функционал в $\mathfrak{S}_0(\Omega, \mathfrak{F}, \mathbf{P})$, за който $\mathbf{E}I_A = \mathbf{P}(A)$.*

Доказателство. Непосредствено от (1) следват свойствата:

1) $\mathbf{E}I_\Omega = 1$ (**нормираност**).

2) Ако ξ, $\eta \in \mathfrak{S}_0$, за всички $a, b \in \mathbf{R}^1$ следва, че е изпълнено $\mathbf{E}(a\xi + b\eta) = a\mathbf{E}\xi + b\mathbf{E}\eta$ (**линейност**).

Действително, ако

$$\xi = \sum_{i=1}^{n} x_i I_{A_i}, \quad \eta = \sum_{j=1}^{m} y_j I_{B_j}, \quad \sum_{i=1}^{n} A_i = \Omega = \sum_{j=1}^{m} B_j,$$

то

$$\xi + \eta = \sum_k z_k I_{C_k} = \sum_i \sum_j (x_i + y_j) I_{A_i B_j}.$$

Следователно (използвайки адитивността на мярката $\mathbf{P}$)

$$\begin{aligned}
\mathbf{E}(\xi+\eta) &= \sum_k z_k \mathbf{P}(C_k) = \sum_i \sum_j (x_i + y_j)\mathbf{P}(A_i B_j) \\
&= \sum_i \sum_j x_i \mathbf{P}(A_i B_j) + \sum_i \sum_j y_j \mathbf{P}(A_i B_j) \\
&= \sum_i x_i \mathbf{P}\left(A_i \sum_j B_j\right) + \sum_j y_j \mathbf{P}\left(B_j \sum_i A_i\right) \\
&= \sum_i x_i \mathbf{P}(A_i) + \sum_j y_j \mathbf{P}(B_j) = \mathbf{E}\xi + \mathbf{E}\eta.
\end{aligned}$$

3) От $\xi \in \mathfrak{S}_0$, $\xi \geqq 0$ (т. е. $\xi \in \mathfrak{S}^+$), следва $\mathbf{E}\xi \geqq 0$ (**позитивност**).

От свойства 2) и 3) следва равносилност на свойствата позитивност и **монотонност** ($\mathbf{E}\xi \leqq \mathbf{E}\eta$ при $\xi \leqq \eta$). Действително, от $\xi \leqq \eta$ следва, че $\eta - \xi \geqq 0$ и $\mathbf{E}(\eta - \xi) \geqq 0$, откъдето

$$\mathbf{E}\eta = \mathbf{E}(\xi + \eta - \xi) = \mathbf{E}\xi + \mathbf{E}(\eta - \xi) \geqq \mathbf{E}\xi.$$

4) От $\xi_n \in \mathfrak{S}_0$ и $\xi_n \downarrow 0$ следва $\mathbf{E}\xi_n \downarrow 0$ (**непрекъснатост**).

Нека $M = \max\limits_{\omega \in \Omega} \xi_1(\omega)$. Тогава $M < \infty$, тъй като $\xi_1 \in \mathfrak{S}_0$. Следователно за всяко $\varepsilon > 0$ имаме

$$0 \leqq \xi_n \leqq \xi_n I_{\{\xi_n > \varepsilon\}} + \xi_n I_{\{\xi_n \leqq \varepsilon\}} \leqq M I_{\{\xi_n > \varepsilon\}} + \varepsilon.$$

Тогава за всяко $\varepsilon > 0$ и за всички $n \geqq 1$

$$0 \leqq \mathbf{E}\xi_n \leqq M \mathbf{E} I_{\{\xi_n > \varepsilon\}} + \varepsilon = M\mathbf{P}\{\xi_n > \varepsilon\} + \varepsilon.$$

Тъй като $\xi_n \downarrow 0$, то $\{\xi_n > \varepsilon\} \downarrow \varnothing$, откъдето по аксиома 4° (за непрекъснатост на мярката $\mathbf{P}$) следва $\mathbf{P}\{\xi_n > \varepsilon\} \xrightarrow[n\to\infty]{} 0$, т. е. $\mathbf{E}\xi_n \downarrow 0$.

Свойство 4) допуска следното обобщение:

Ако $\xi_n \in \mathfrak{S}_0$ и $\xi_n \uparrow \xi$ ($\xi_n \downarrow \xi$), $\xi \in \mathfrak{S}_0$, то $\mathbf{E}\xi_n \uparrow \mathbf{E}\xi$ ($\mathbf{E}\xi_n \downarrow \mathbf{E}\xi$).

Действително, от $\xi_n \uparrow \xi$ следва $(\xi - \xi_n) \downarrow 0$, откъдето $\mathbf{E}(\xi - \xi_n) \downarrow 0$. Но $\mathbf{E}(\xi - \xi_n) = \mathbf{E}\xi - \mathbf{E}\xi_n$ и следователно $\mathbf{E}\xi_n \uparrow \mathbf{E}\xi$.

Това свойство често се нарича **монотонна непрекъснатост.**

5) **Единственост**. Да разгледаме функцията от събития $A \in \mathfrak{F}$, дефинирана чрез полагането $\mathbf{P}'(A) = \mathbf{J}I_A$, където $\mathbf{J}$ е нормиран, линеен, позитивен и непрекъснат функционал в $\mathfrak{S}_0$, т. е. удовлетворява свойствата 1) – 4). Лесно се проверява, че $\mathbf{P}'(A)$, $A \in \mathfrak{F}$, е вероятност. Действително:

а) $\mathbf{P}'(A) \geqq 0$ — следва от свойство 3);

б) $\mathbf{P}'(\Omega) = 1$ — следва от свойство 1);

в) ако $A \cap B = \varnothing$, то $I_{A+B} = I_A + I_B$ и следователно

$$\mathbf{P}'(A+B) = \mathbf{J}I_{A+B} = \mathbf{J}(I_A + I_B) = \mathbf{J}I_A + \mathbf{J}I_B = \mathbf{P}'(A) + \mathbf{P}'(B),$$

т. е. адитивност;

г) ако $A_n \downarrow \varnothing$, то $I_{A_n} \downarrow 0$ и $\mathbf{J}I_{A_n} = \mathbf{P}'(A_n) \downarrow 0$ (по свойство 4)), т.е. непрекъснатост.

Тогава за всяко $\xi \in \mathfrak{S}_0$ имаме

$$\mathbf{J}\xi = \mathbf{J}\sum_{k=1}^{n} x_k I_{A_k} = \sum_{k=1}^{n} x_k \mathbf{J}I_{A_k} = \sum_{k=1}^{n} x_k \mathbf{P}'(A_k).$$

И така, доказахме, че всеки функционал $\mathbf{E}$ в $\mathfrak{S}_0$ със свойствата 1) – 4) представлява математическо очакване относно вероятност $\mathbf{P}$, дефинирана чрез $\mathbf{P}(A) = \mathbf{E}I_A$ за всяко $A \in \mathfrak{F}$.

Следващият пример илюстрира „стохастичния" смисъл на математическото очакване.

Пример 1. Нека в лотария (спорт-тото или друга игра) при всеки тираж печалбата е случайна и взема една от стойностите $x_1, x_2, \ldots, x_k$. Ако лотарията се е провела N пъти, като N_i пъти се е паднала печалба x_i, при което $N = \sum_{i=1}^{k} N_i$, то $\frac{N_i}{N}$ е честотата на печалбата x_i, а $\overline{x} = \frac{1}{N}\sum_{i=1}^{k} x_i N_i$ — средната печалба в лотарията. Нека ξ е случайна величина, равна на печалбата в лотарията. От стохастичната устойчивост на честотите следва, че $\dfrac{N_i}{N} \approx \mathbf{P}\{\xi = x_i\}$, поради което

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{N} x_i N_i \approx \sum_{i=1}^{k} x_i \mathbf{P}\{\xi = x_i\} = \mathbf{E}\xi,$$

т. е. средната печалба $\overline{x}$ се колебае около математическото очакване $\mathbf{E}\xi$.

Нека $\xi \in Bi(n,p)$, т. е.

$$p_k = \mathbf{P}\{\xi = k\} = \binom{n}{k} p^k q^{n-k}, \quad k = 0,\, 1,\, \ldots,\, n; \quad q = 1 - p.$$

Тогава от (1) намираме

$$\begin{aligned}
\mathbf{E}\xi &= \sum_{k=0}^{n} k p_k = \sum_{k=1}^{n} k \frac{n!}{k!(n-k)!} p^k q^{n-k} \\
&= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} q^{n-k} \\
&= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j q^{n-1-j} = np(p+q)^{n-1} = np.
\end{aligned}$$

§ 2. Интегруеми и квазиинтегруеми случайни величини. Теорема на Фату – Лебег

Да означим с $\mathfrak{S}^+$ съвкупността на неотрицателните случайни величини в $(\Omega, \mathfrak{F}, \mathbf{P})$. В 2.2 показахме, че за всяка случайна величина $\xi \in \mathfrak{S}^+$ имаме $\xi = \lim \uparrow \xi_n$, където $\xi_n \in \mathfrak{S}_0^+$ (т. е. $\xi_n \in \mathfrak{S}_0$ и $\xi_n \geqq 0$) за всяко n. Това ни дава основание да дадем следната

Дефиниция 1. *За всяка случайна величина $\xi \in \mathfrak{S}^+$ съотношението*

(1) $$\mathbf{E}\xi = \lim \uparrow \mathbf{E}\xi_n,$$

където $\xi_n \in \mathfrak{S}_0^+$ и $\xi_n \uparrow \xi$, се нарича математическо очакване $\mathbf{E}\xi$.

Теорема 1. *Формула* (1) *дефинира еднозначно продължение на функционала* $\mathbf{E}$ *от* $\mathfrak{S}_0$ *в* $\mathfrak{S}^+$ *със свойствата:*

1) *ако* $\xi \in \mathfrak{S}^+$, *то* $0 \leqq \mathbf{E}\xi \leqq \infty$;

2) *ако* a *е положителна константа и* $\xi \in \mathfrak{S}^+$, *то* $\mathbf{E}a\xi = a\mathbf{E}\xi$;

3) *ако* $\xi, \eta \in \mathfrak{S}^+$, *то* $\mathbf{E}(\xi + \eta) = \mathbf{E}\xi + \mathbf{E}\eta$;

4) *ако* $\xi \leqq \eta$ *и* $\xi, \eta \in \mathfrak{S}^+$, *то* $\mathbf{E}\xi \leqq \mathbf{E}\eta$ (***монотонност***);

5) *ако* $\xi_n \in \mathfrak{S}^+$ *и* $\xi_n \uparrow$, *то съществуват* $\lim \uparrow \xi_n = \xi \in \mathfrak{S}^+$ *и* $\mathbf{E}\xi = \lim \uparrow \mathbf{E}\xi_n \leqq \infty$ (***монотонна непрекъснатост***).

Доказателство. Първо ще докажем еднозначността, т. е. че стойността на функционала $\mathbf{E}\xi$ в (1) не зависи от избора на редицата.

Нека $\xi = \lim \uparrow \xi_n \leqq \eta = \lim \uparrow \eta_n$, където $\xi_n, \eta_n \in \mathfrak{S}_0^+$ за всяко n. От $\xi_m, \eta_n \in \mathfrak{S}_0^+$ следва $\min(\xi_m, \eta_n) \in \mathfrak{S}_0^+$. Тъй като $\min(\xi_m, \eta_n) \leqq \eta_n$, то

$$\xi_m = \lim_{n\to\infty} \uparrow \min(\xi_m, \eta_n) \leqq \lim_{n\to\infty} \uparrow \eta_n = \eta.$$

Тогава от (1) получаваме

$$\mathbf{E}\xi_m = \lim_{n\to\infty} \uparrow \mathbf{E}\min(\xi_m, \eta_n) \leqq \lim_{n\to\infty} \uparrow \mathbf{E}\eta_n.$$

Оттук при $m \to \infty$ следва

$$\lim_{m\to\infty} \uparrow \mathbf{E}\xi_m = \mathbf{E}\xi \leqq \mathbf{E}\eta = \lim_{n\to\infty} \uparrow \mathbf{E}\eta_n.$$

Ако $\xi = \lim \uparrow \xi_n = \lim \uparrow \eta_n$, то $\mathbf{E}\xi = \lim \uparrow \mathbf{E}\xi_n = \lim \uparrow \mathbf{E}\eta_n$, т. е. стойността на функционала в (1) не зависи от избора на редицата.

Свойствата 1) – 4) следват от (1) и току-що доказаното.

Ще докажем 5). Нека $\xi_n \uparrow$ и $\xi_n \in \mathfrak{S}^+$. Тогава $\xi_n = \lim\limits_{m\to\infty} \uparrow X_{nm}$, където $X_{nm} \in \mathfrak{S}_0^+$. Означаваме $Y_m = \max\limits_{n \leqq m} X_{nm}$. Очевидно $Y_m \in \mathfrak{S}_0^+$ и $Y_m \uparrow$. При $n \leqq m$ имаме

$$X_{nm} \leqq Y_m \leqq \xi_m, \quad \mathbf{E}X_{nm} \leqq \mathbf{E}Y_m \leqq \mathbf{E}\xi_m. \tag{2}$$

Оттук при $m \to \infty$ получаваме

$$\lim_{m\to\infty} \uparrow X_{nm} = \xi_n \leqq \lim_{m\to\infty} \uparrow Y_m \leqq \lim_{m\to\infty} \uparrow \xi_m.$$

От горните неравенства при $n \to \infty$ следва, че съществува

$$\lim_{n\to\infty} \uparrow \xi_n = \lim \uparrow Y_n = \xi \in \mathfrak{S}^+.$$

От (2) при $m \to \infty$ намираме

$$\lim_{m\to\infty} \mathbf{E}X_{nm} = \mathbf{E}\xi_n \leqq \lim_{m\to\infty} \uparrow \mathbf{E}Y_m = \mathbf{E}\xi \leqq \lim_{m\to\infty} \uparrow \mathbf{E}\xi_m.$$

Оттук при $n \to \infty$ следва

$$\lim_{n\to\infty} \uparrow \mathbf{E}\xi_n \leqq \mathbf{E}\xi \leqq \lim_{m\to\infty} \uparrow \mathbf{E}\xi_m,$$

т. е. $\mathbf{E}\xi = \mathbf{E} \lim\limits_{n\to\infty}\uparrow \xi_n = \lim\limits_{m\to\infty} \uparrow \mathbf{E}\xi_m$, което доказва монотонната непрекъснатост.

За всяка случайна величина ξ можем да образуваме

$$\xi^+ = \max(\xi, 0) \quad \text{и} \quad \xi^- = \max(-\xi, 0).$$

Очевидно $\xi^+, \xi^- \in \mathfrak{S}^+$, $\xi = \xi^+ - \xi^-$, $|\xi| = \xi^+ + \xi^-$.

Дефиниция 2. *Ако $\mathbf{E}\xi^+ < \infty$ и $\mathbf{E}\xi^- < \infty$, случайната величина ξ се нарича интегруема. Тогава математическото очакване (или средната стойност) се определя чрез съотношението*

$$\mathbf{E}\xi = \mathbf{E}\xi^+ - \mathbf{E}\xi^-. \tag{3}$$

Множеството на интегруемите случайни величини се означава с L_1 (или $L_1(\Omega, \mathfrak{F}, \mathbf{P})$). Очевидно

$$0 \leqq \mathbf{E}|\xi| = \mathbf{E}\xi^+ + \mathbf{E}\xi^- < \infty.$$

Дефиниция 3. *Ако поне едно от числата $\mathbf{E}\xi^+$ или $\mathbf{E}\xi^-$ е крайно, случайната величина ξ се нарича квазиинтегруема и*

$$\mathbf{E}\xi = \mathbf{E}\xi^+ - \mathbf{E}\xi^- \tag{4}$$

определя математическото очакване.

Множеството от квазиинтегруемите случайни величини ще означаваме с $\widetilde{L_1}$. Ако $\xi \in \widetilde{L_1}$, то $0 \leqq \mathbf{E}|\xi| = \mathbf{E}\xi^+ + \mathbf{E}\xi^- \leqq \infty$. Очевидно $L_1 \subset \widetilde{L_1}$. Често се използват означенията

$$\mathbf{E}\xi = \mathbf{E}_{\mathbf{P}}\xi = \int_\Omega \xi d\mathbf{P} = \int_\Omega \xi(\omega) d\mathbf{P}(\omega) = \int_\Omega \xi(\omega)\mathbf{P}(d\omega),$$

откъдето естествено идва названието интеграл, по-точно за (3) и (4) се казва, че дефинират *интеграл на Лебег.*

Теорема 2. *Дефинираното в (4) математическо очакване (т. е. функционал) запазва свойствата нормираност, позитивност, линейност и монотонна непрекъснатост върху квазиинтегруемите случайни величини (следователно и върху интегруемите).*

Доказателство. Първите две свойства са очевидни и затова ще докажем линейността.

Ще покажем, че $\mathbf{E}c\xi = c\mathbf{E}\xi$ за всяко $c \in \mathbf{R}^1$ и $\xi \in \widetilde{L_1}$. Нека $c > 0$. Тогава $(c\xi)^+ = c\xi^+$ и $(c\xi)^- = c\xi^-$, откъдето

$$\mathbf{E}c\xi = \mathbf{E}(c\xi)^+ - \mathbf{E}(c\xi)^- = c(\mathbf{E}\xi^+ - \mathbf{E}\xi^-) = c\mathbf{E}\xi.$$

Ако $c < 0$, то $-c > 0$ и

$$(c\xi)^+ = ((-c)(-\xi))^+ = \max(0, -(-c)\xi) = (-c\xi)^- = (-c)\xi^-,$$
$$(c\xi)^- = ((-c)(-\xi))^- = \max(0, (-c)\xi) = (-c\xi)^+ = (-c)\xi^+.$$

Следователно

$$\mathbf{E}c\xi = \mathbf{E}(c\xi)^+ - \mathbf{E}(c\xi)^- = -c\mathbf{E}\xi^- - (-c)\mathbf{E}\xi^+ = c\mathbf{E}\xi^+ - c\mathbf{E}\xi^- = c\mathbf{E}\xi.$$

Сега ще докажем, че $\mathbf{E}(\xi + \eta) = \mathbf{E}\xi + \mathbf{E}\eta$, ако $\xi + \eta$ има смисъл (т. е. изключваме възможността $\infty - \infty$), като или $\xi^-, \eta^- < \infty$, или $\xi^+, \eta^+ < \infty$ (т. е. $\xi, \eta \in \widetilde{L_1}$). Тогава от $(\xi + \eta)^+ \leqq \xi^+ + \eta^+$ и $(\xi + \eta)^- \leqq \xi^- + \eta^-$ следва, че $(\xi + \eta) \in \widetilde{L_1}$. Тъй като

$$\xi + \eta = \xi^+ - \xi^- + \eta^+ - \eta^- = \xi^+ + \eta^+ - (\xi^- + \eta^-) = (\xi + \eta)^+ - (\xi + \eta)^-,$$

то $(\xi + \eta)^+ + \xi^- + \eta^- = \xi^+ + \eta^+ + (\xi + \eta)^-$. Оттук по теорема 1 получаваме

$$\mathbf{E}(\xi + \eta)^+ + \mathbf{E}\xi^- + \mathbf{E}\eta^- = \mathbf{E}\xi^+ + \mathbf{E}\eta^+ + \mathbf{E}(\xi + \eta)^-,$$

откъдето

$$\mathbf{E}(\xi + \eta) = \mathbf{E}(\xi + \eta)^+ - \mathbf{E}(\xi + \eta)^- = \mathbf{E}\xi^+ - \mathbf{E}\xi^- + \mathbf{E}\eta^+ - \mathbf{E}\eta^- = \mathbf{E}\xi + \mathbf{E}\eta.$$

Монотонната непрекъснатост следва от следната

Лема 1. *Нека е дадена редица от случайни величини* $\{\xi_n\}$. *Тогава:*

а) *ако* $\xi_n \uparrow \xi$ *и за някое* n_0 *имаме* $\mathbf{E}\xi_{n_0}^- < \infty$, *то* $\xi \in \widetilde{L_1}$ *и* $\mathbf{E}\xi_n \uparrow \mathbf{E}\xi$;

б) *ако* $\xi_n \downarrow \xi$ *и за някое* n_0 *е изпълнено* $\mathbf{E}\xi_{n_0}^+ < \infty$, *то* $\xi \in \widetilde{L_1}$ *и* $\mathbf{E}\xi_n \downarrow \mathbf{E}\xi$.

Доказателство. а) От $\xi_n = \xi_n^+ - \xi_n^- \uparrow$ следва, че $\xi_n^- \downarrow$, т. е. $\xi_n^- \leqq \xi_{n_0}^-$ за $n \geqq n_0$. Следователно $\xi^- \leqq \xi_{n_0}^-$ и $\mathbf{E}\xi^- \leqq \mathbf{E}\xi_{n_0}^- < \infty$, т. е. ξ е квазиинтегруема. От друга страна, при $n \geqq n_0$ имаме

$$0 \leqq \xi_n^+ = \xi_n + \xi_n^- \leqq \xi_n + \xi_{n_0}^-,$$

т. е. $(\xi_n + \xi_{n_0}^-) \in \mathfrak{S}^+$. Тъй като $\xi_n + \xi_{n_0}^- \uparrow \xi + \xi_{n_0}^-$, то по теорема 1

$$\mathbf{E}(\xi_n + \xi_{n_0}^-) = \mathbf{E}\xi_n + \mathbf{E}\xi_{n_0}^- \uparrow \mathbf{E}(\xi + \xi_{n_0}^-) = \mathbf{E}\xi + \mathbf{E}\xi_{n_0}^-.$$

Следователно
$$\lim_{n\to\infty} \uparrow \mathbf{E}\xi_n = \mathbf{E}\xi.$$

б) Доказва се аналогично. От $\xi_n = \xi_n^+ - \xi_n^- \downarrow$ следва, че $\xi_n^+ \downarrow$, т. е. при $n \geqq n_0$ имаме $\xi_n^+ \leqq \xi_{n_0}^+$. Следователно $\xi^+ \leqq \xi_{n_0}^+$, откъдето $\mathbf{E}\xi^+ \leqq \mathbf{E}\xi_{n_0}^+ < \infty$ и тогава $\xi \in \widetilde{L_1}$. От друга страна, при $n \geqq n_0$ имаме
$$0 \leqq \xi_n^- = \xi_n^+ - \xi_n \leqq \xi_{n_0}^+ - \xi_n \uparrow \xi_{n_0}^+ - \xi.$$
Тогава по теорема 1
$$\mathbf{E}(\xi_{n_0}^+ - \xi_n) = \mathbf{E}\xi_{n_0}^+ - \mathbf{E}\xi_n \uparrow \mathbf{E}(\xi_{n_0}^+ - \xi) = \mathbf{E}\xi_{n_0}^+ - \mathbf{E}\xi,$$
т. е. $\mathbf{E}\xi_n \downarrow \mathbf{E}\xi$.

Лема 1 показва, че при монотонна сходимост може да извършваме граничен преход под знака на интеграла (което наричаме *монотонна непрекъснатост*).

Следващата теорема снема ограничението за монотонност.

Теорема 3 (теорема на Фату – Лебег). *Нека $\{\xi_n\}$ е редица от случайни величини. Тогава при $n \to \infty$ са изпълнени съотношенията:*

а) *ако $\xi_n \leqq X \in L_1$, то $\lim\limits_{n\to\infty} \sup \mathbf{E}\xi_n \leqq \mathbf{E}(\lim\limits_{n\to\infty} \sup \xi_n)$;*

б) *ако $\xi_n \geqq Y \in L_1$, то $\mathbf{E}(\lim\limits_{n\to\infty} \inf \xi_n) \leqq \lim\limits_{n\to\infty} \mathbf{E} \inf \xi_n$;*

в) *ако $\xi_n \to \xi$ и $|\xi| \leqq Z \in L_1$, то $\xi \in L_1$ и $\mathbf{E}\xi = \lim\limits_{n\to\infty} \mathbf{E}\xi_n$.*

Доказателство. а) От $\xi_n \leqq X$ следва неравенството $\xi_n^+ \leqq X^+$, откъдето $\mathbf{E}\xi_n^+ \leqq \mathbf{E}X^+ < \infty$, т. е. $\xi_n \in \widetilde{L_1}$. От друга страна, $\sup \xi_n \leqq X$ и аналогично $\sup \xi_n \in \widetilde{L_1}$. Тъй като $\lim\limits_n \sup \xi_n = \lim\limits_{n\to\infty} \downarrow \sup\limits_{m \geqq n} \xi_m$, то от лема 1б) следва $\mathbf{E} \lim\limits_{n\to\infty} \sup \xi_n = \lim\limits_{n\to\infty} \downarrow \mathbf{E} \sup\limits_{m \geqq n} \xi_m$. Но при $m \geqq n$ имаме $\sup\limits_{l \geqq n} \xi_l \geqq \xi_m$, откъдето $\mathbf{E} \sup\limits_{l \geqq n} \xi_l \geqq \mathbf{E}\xi_m$ и тогава $\mathbf{E} \sup\limits_{l \geqq n} \xi_l \geqq \sup\limits_{m \geqq n} \mathbf{E}\xi_m$. Оттук окончателно получаваме
$$\mathbf{E} \lim_n \sup \xi_n = \lim_{n\to\infty} \downarrow \mathbf{E} \sup_{l \geqq n} \xi_l \geqq \lim_{n\to\infty} \downarrow \sup_{m \geqq n} \mathbf{E}\xi_m = \lim_n \sup \mathbf{E}\xi_n.$$

б) От $\xi_n \geqq Y$ следва, че $\xi_n^- \leqq Y^-$ и $\mathbf{E}\xi_n^- \leqq \mathbf{E}Y^- < \infty$, т. е. $\xi_n \in \widetilde{L_1}$. От $\xi_n \geqq Y$ се вижда, че $\inf \xi_n \geqq Y$, откъдето аналогично $\inf \xi_n \in \widetilde{L_1}$, но $\lim\limits_n \inf \xi_n = \lim\limits_{n\to\infty} \uparrow \inf\limits_{m \geqq n} \xi_m$ и по лема 1а) получаваме
$$\mathbf{E} \lim_{n\to\infty} \inf \xi_n = \lim_{n\to\infty} \uparrow \mathbf{E} \inf_{m \geqq n} \xi_m.$$

От друга страна, при $m \geqq n$ имаме $\inf_{l \geqq n} \xi_l \leqq \xi_m$, откъдето следва $\mathbf{E} \inf_{l \geqq n} \xi_l \leqq \mathbf{E}\xi_m$ и тогава $\mathbf{E} \inf_{l \geqq n} \xi_l \leqq \inf_{m \geqq n} \mathbf{E}\xi_m$. При $n \to \infty$ окончателно намираме

$$\mathbf{E} \liminf_n \xi_n = \lim_{n\to\infty} \uparrow \mathbf{E} \inf_{l \geqq n} \xi_l \leqq \lim_{n\to\infty} \uparrow \inf_{m \geqq n} \mathbf{E}\xi_m = \liminf_n \mathbf{E}\xi_n.$$

в) От $Z \in L_1$ и $-Z \leqq \xi_n \leqq Z$ следва, че са изпълнени условията а) и б), откъдето

$$\mathbf{E} \lim_{n\to\infty} \inf \xi_n \leqq \lim_{n\to\infty} \inf \mathbf{E}\xi_n \leqq \lim_{n\to\infty} \sup \mathbf{E}\xi_n \leqq \mathbf{E} \lim_{n\to\infty} \sup \xi_n. \tag{5}$$

Тъй като $\lim_n \xi_n = \lim_{n\to\infty} \inf \xi_n = \lim_{n\to\infty} \sup \xi_n$, от (5) следва

$$\mathbf{E} \lim_{n\to\infty} \xi_n = \lim_{n\to\infty} \mathbf{E}\xi_n.$$

§ 3. Математическото очакване като стилтесов интеграл. Моменти

Нека ξ е случайна величина във вероятностно пространство $(\Omega, \mathfrak{F}, \mathbf{P})$, т. е. ξ е такова изображение на $(\Omega, \mathfrak{F})$ в бореловото измеримо пространство $(\mathbf{R}^1, \mathfrak{B}_1)$, че за всяко $B \in \mathfrak{B}_1$ е изпълнено $\xi^{-1}(B) = \{\omega : \xi(\omega) \in B\} \in \mathfrak{F}$.

Лема 1. *Съотношението*

$$Q_\xi(B) = \mathbf{P}(\xi^{-1}(B)) \tag{1}$$

дефинира вероятност в $(\mathbf{R}^1, \mathfrak{B}_1)$.

Доказателство. Очевидно за всяко $B \in \mathfrak{B}_1$ имаме $Q_\xi(B) \geqq 0$ и $Q_\xi(\mathbf{R}^1) = \mathbf{P}(\Omega) = 1$, т. е. в сила са аксиоми **1°** и **2°** от § 1.3.

Нека $\{B_n\}$ е редица от несъвместими борелови множества, т. е. $B_i \cap B_j = \varnothing$, $i \neq j$. Тогава $\xi^{-1}(B_i) \cap \xi^{-1}(B_j) = \varnothing$ и

$$\xi^{-1}\left(\sum_n B_n\right) = \left\{\omega : \xi(\omega) \in \sum_n B_n\right\} = \sum_n \{\omega : \xi(\omega) \in B_n\} = \sum_n \xi^{-1}(B_n).$$

Следователно

$$Q_\xi\left(\sum_n B_n\right) = \mathbf{P}\left(\xi^{-1}\left(\sum_n B_n\right)\right) = \mathbf{P}\left(\sum_n \xi^{-1}(B_n)\right)$$

$$= \sum_n \mathbf{P}(\xi^{-1}(B_n)) = \sum_n Q_\xi(B_n),$$

т. е. в сила е аксиома **3*** от § 1.3 за σ-адитивност, което доказва лемата.

Нека $B_x = (-\infty, x)$. Тогава от теоремата за съответствие (теорема 2.4.2) следва

$$Q_\xi(B_x) = F_\xi(x). \tag{2}$$

Равенство (2) дефинира еднозначно обратимо съответствие между вероятностната мярка Q_ξ и функцията на разпределение $F_\xi(x)$.

Теорема 1. *Нека ξ е случайна величина в $(\Omega, \mathfrak{F}, \mathbf{P})$, а $g(x)$ е борелова функция. Тогава*

$$\mathbf{E}_\mathbf{P} g(\xi) = \int_\Omega g(\xi(\omega)) d\mathbf{P}(\omega) = \int_{\mathbf{R}^1} g(x) dQ_\xi(x) = \mathbf{E}_Q g. \tag{3}$$

Доказателство. Да отбележим първо, че $\zeta(\omega) = g(\xi(\omega))$ е случайна величина в $(\Omega, \mathfrak{F}, \mathbf{P})$, съгласно свойство 7) от 2.2.

1) Ще докажем (3) в случая, когато g е индикатор, т. е. за всяко $B \in \mathfrak{B}_1$ е изпълнено

$$g(x) = I_B(x) = \begin{cases} 1 & \text{за } x \in B, \\ 0 & \text{за } x \notin B. \end{cases}$$

Тогава

$$g(\xi(\omega)) = \begin{cases} 1, & \text{когато } \xi(\omega) \in B, \text{ т. е. } \omega \in \xi^{-1}(B), \\ 0, & \text{когато } \xi(\omega) \notin B, \text{ т. е. } \omega \notin \xi^{-1}(B). \end{cases}$$

Имаме

$$g(\xi(\omega)) = I_{\xi^{-1}(B)}(\omega),$$

$$\mathbf{E}_\mathbf{P} g(\xi) = \int_\Omega g(\xi(\omega)) d\mathbf{P}(\omega) = \mathbf{E}_\mathbf{P} I_{\xi^{-1}(B)} = \mathbf{P}(\xi^{-1}(B))$$

$$= Q_\xi(B) = \mathbf{E}_{Q_\xi} I_B = \mathbf{E}_{Q_\xi} g = \int_{\mathbf{R}^1} g(x) dQ_\xi(x).$$

2) От 1) веднага се вижда, че (3) е вярно, когато g е елементарна случайна величина, т. е. $g = \sum_{k=1}^{n} x_k I_{B_k}$.

3) Нека g е ограничена, т. е. $|g(x)| \leqq C$. Следователно $\mathbf{E}_Q|g| \leqq C$. Съгласно теорема 2.2.2 съществува редица от елементарни случайни величини (т. е. елементарни борелови функции) g_n, такива че $g_n \to g$ и $|g_n| \leqq |g|$. Тогава по теоремата на Фату – Лебег следва верността на (3) за g, понеже (3) е вярно за g_n (т. е. можем да извършим граничен преход под знака на интеграла).

4) В общия случай образуваме случайните величини

$$g_N^+(x) = \min(N, g^+(x)), \quad \widetilde{g}_N^+(\omega) = \min(N, g^+(\xi(\omega))).$$

Очевидно те са ограничени и положителни, като $g_N^+(x) \uparrow g^+(x)$ и $\widetilde{g}_N^+(\omega) \uparrow g^+(\xi(\omega))$ при $N \to \infty$. От друга страна, съгласно доказаното в 3), е изпълнено $\mathbf{E}_\mathbf{P}\widetilde{g}_N^+ = \mathbf{E}_{Q_\xi} g_N^+$. Тогава по лема 2.1

$$\mathbf{E}_\mathbf{P} g^+(\xi) = \lim_{N\to\infty} \uparrow \mathbf{E}_\mathbf{P}\widetilde{g}_N^+ = \lim_{N\to\infty} \uparrow \mathbf{E}_{Q_\xi} g_N^+ = \mathbf{E}_{Q_\xi} g^+.$$

Аналогично се показва, че $\mathbf{E}_\mathbf{P} g^-(\xi) = \mathbf{E}_{Q_\xi} g^-$. Следователно (3) е вярно за $g(x) = g^+(x) - g^-(x)$, което доказва теоремата.

Като имаме предвид съотношение (2), интегралът (3) се записва по следния начин:

$$\mathbf{E}g(\xi) = \int_{\mathbf{R}^1} g(x)dQ_\xi(x) = \int_{-\infty}^{+\infty} g(x)dF_\xi(x), \tag{4}$$

и се нарича *интеграл на Лебег – Стилтес.*

Специално при $g(x) = x$ от (3) и (4) получаваме

$$\mathbf{E}\xi = \int_{-\infty}^{+\infty} x dF_\xi(x). \tag{5}$$

При $g(x) = x^k$ имаме

$$\mathbf{E}\xi^k = \int_{-\infty}^{+\infty} x^k dF_\xi(x), \tag{6}$$

което се нарича *k-ти начален момент*, откъдето при $x = 1$ получаваме (5). Поради това математическото очакване се нарича често *първи начален момент* или просто *първи момент.*

При $g(x) \equiv (x-a)^k$ от (3) и (4) следва

$$\mathbf{E}(\xi - a)^k = \int_{-\infty}^{+\infty} (x-a)^k dF_\xi(x). \tag{7}$$

При $a = \mathbf{E}\xi$ от (7) имаме $\mu_k = \mathbf{E}(\xi - \mathbf{E}\xi)^k$, който се нарича *$k$-ти централен момент.* Специално при $k = 2$ централният момент от втори ред се нарича *дисперсия* (и както ще видим по-нататък, играе важна роля в стохастиката), т. е.

$$\mathbf{D}\xi = \mathbf{E}(\xi - \mathbf{E}\xi)^2 = \mathbf{E}\xi^2 - (\mathbf{E}\xi)^2. \tag{8}$$

Последното равенство се получава, като повдигнем на квадрат и използваме факта, че математическото очакване е линеен функционал.

Дисперсията на случайна величина е мярка за разсейването на стойностите ѝ около средната стойност, а $\sqrt{\mathbf{D}\xi}$ се нарича *средно-квадратично отклонение.*

Например, ако случайната величина ξ взема стойностите $x_1, \dots, x_n$ с една и съща вероятност $\frac{1}{n}$, както вече видяхме в § 1, $\mathbf{E}\xi = \frac{1}{n}\sum_{k=1}^{n} x_k = \overline{x}$, а непосредствено от (8) получаваме

$$\mathbf{D}\xi = \frac{1}{n}\sum_{k=1}^{n}(x_k - \overline{x})^2.$$

Не е трудно да се види, че функцията $h(a) = \mathbf{E}(\xi - a)^2$ има минимум при $a = \mathbf{E}\xi$. Действително

$$\begin{aligned}\mathbf{E}(\xi - a)^2 &= \mathbf{E}[(\xi - \mathbf{E}\xi) + (\mathbf{E}\xi - a)]^2 \\ &= \mathbf{E}[(\xi - \mathbf{E}\xi)^2 + 2(\xi - \mathbf{E}\xi)(\mathbf{E}\xi - a) + (\mathbf{E}\xi - a)^2] \\ &= \mathbf{D}\xi + 2(\mathbf{E}\xi - \mathbf{E}\xi)(\mathbf{E}\xi - a) + (\mathbf{E}\xi - a)^2 \\ &= \mathbf{D}\xi + (\mathbf{E}\xi - a)^2 \geqq \mathbf{D}\xi.\end{aligned}$$

Този факт обяснява защо се предпочита да се използват централните моменти.

Следните свойства следват директно от дефиницията, дадена с равенство (8):

$$\mathbf{D}c = 0, \quad c \in \mathbf{R}^1, \tag{9}$$

т.е. „разсейването“ на константите е нула;

$$\mathbf{D}c\xi = c^2\mathbf{D}\xi, \quad c \in \mathbf{R}^1, \tag{10}$$

т. е. за разлика от математическото очакване $\mathbf{E}$ дисперсията $\mathbf{D}$ е квадратичен функционал.

Да се върнем към съотношенията (3) и (4). Те показват, че математическото очакване $\mathbf{E}g(\xi)$, т. е. интегралът на Лебег – Стилтес, не зависи от вида на случайната величина $\xi(\omega)$, $\omega \in \Omega$, а само от нейната функция на разпределение $F_\xi(x)$.

Ако математическото очакване $\mathbf{E}g(\xi)$ е крайно, можем да интерпретираме $\int\limits_{-\infty}^{+\infty} g(x)dF_\xi(x)$ като несобствен интеграл на Риман – Стилтес (в този случай той е абсолютно сходящ).

Да напомним, че интегралът на Риман – Стилтес от $g(x)$ в краен интервал $(a, b]$ по функция с ограничена вариация $F(x)$ (т. е. която може да се представи като разлика на две монотонни функции) се дефинира по следния начин:

$$\int_a^b g(x)\,dF(x) = \lim_{n\to\infty} \sum_{k=0}^{n-1} g(x'_k)[F(x_{k+1}) - F(x_k)],$$

където $x_k = a + \dfrac{b-a}{n} \cdot k$, $\quad k = 0, 1, \ldots, n$, $x_k < x'_k \leqq x_{k+1}$.

Несобственият интеграл $\int\limits_{-\infty}^{+\infty} g(x)dF(x)$ се дефинира като

$$\lim_{a\to-\infty,\, b\to+\infty} \int_a^b g(x)dF(x).$$

Ако $F(x)$ има производна $f(x)$ и $F(x_2) - F(x_1) = \int\limits_{x_1}^{x_2} f(x)dx$ за всички $a \leqq x_1 < x_2 \leqq b$, то

$$\int_a^b g(x)dF(x) = \int_a^b g(x)f(x)dx.$$

Да отбележим, че интегралът на Лебег – Стилтес се определя по един и същ начин независимо от това в какво вероятностно пространство са дефинирани случайните величини, докато интегралът на Риман – Стилтес се определя само в $\mathbf{R}^1$ (респ. в $\mathbf{R}^n$). Ако например $g(x)$ е непрекъсната функция в $[a, b]$, интегралът на Лебег – Стилтес в $\mathbf{R}^1$ съвпада с интеграла на Риман – Стилтес.

Интегралът на Лебег – Стилтес притежава всички свойства, доказани в § 1 и § 2. В частност, ако $\xi = \sum_{k=1}^{n} x_k I_{A_k}$ е елементарна случайна величина с функция на разпределение $F_\xi(x) = \sum_{k:x_k<x} p_k$, където $p_k = \mathbf{P}(A_k) = \mathbf{P}\{\xi = x_k\}$, то от (1.1) следва

$$\mathbf{E}g(\xi) = \int_{-\infty}^{+\infty} g(x)dF_\xi(x) = \sum_{k=1}^{n} g(x_k)p_k. \tag{11}$$

Формула (11) е вярна и в общия случай, когато ξ взема изброимо много стойности $x_1, \dots, x_n, \dots$ със съответни вероятности $p_n = \mathbf{P}\{\xi = x_n\}$, $\sum_{n=1}^{\infty} p_n = 1$, т. е.

$$\mathbf{E}g(\xi) = \int_{-\infty}^{+\infty} g(x)dF_\xi(x) = \sum_{n=1}^{\infty} g(x_n)p_n. \tag{12}$$

Доказателството на (12) следва от факта, че то е вярно за елементарните случайни величини

$$\xi_n = \begin{cases} \xi & \text{при } |\xi| \leqq N, \\ 0 & \text{при } |\xi| > N \end{cases}$$

и $\xi = \lim_{N\to\infty} \uparrow \xi_N$, т.е. чрез лема 2.1.а). Естествено (12) има смисъл, когато редът отдясно е абсолютно сходящ, т. е.

$$\mathbf{E}|g(\xi)| = \sum_{n=1}^{\infty} |g(x_n)|p_n < \infty. \tag{13}$$

Теорема 2. *Ако случайната величина ξ има плътност $f_\xi(x) = F'_\xi(x)$ и $\int_{-\infty}^{+\infty} |x| f_\xi(x)dx < \infty$, то*

$$\mathbf{E}\xi = \int_{-\infty}^{+\infty} x dF_\xi(x) = \int_{-\infty}^{+\infty} x f_\xi(x)dx. \tag{14}$$

Доказателство. Нека първоначално $\xi \geqq 0$ и има функция на разпределение

$$(15)\qquad \mathbf{P}\{\xi < x\} = F_\xi(x) = \begin{cases} 0 & \text{при } x \leqq 0, \\ F_\xi(0) + \int\limits_0^x f_\xi(x)dx & \text{при } x > 0. \end{cases}$$

Въвеждаме следната редица от случайни величини:

$$\xi_n = \sum_{k=1}^{n.2^n} \frac{k-1}{2^n} I_{A_k},$$

където $A_k = \left\{ \frac{k-1}{2^n} < \xi \leqq \frac{k}{2^n} \right\}$. Тогава

$$\mathbf{E}\xi_n = \sum_{k=1}^{n.2^n} \frac{k-1}{2^n} \int\limits_{(k-1)/2^n}^{k/2^n} f_\xi(x)dx.$$

Оттук непосредствено се вижда, че

$$(16)\qquad \int\limits_0^n x f_\xi(x)dx - \frac{1}{2^n} \leqq \mathbf{E}\xi_n \leqq \int\limits_0^\infty x f_\xi(x)dx.$$

Тъй като $\xi_n \uparrow \xi$ и $\lim\limits_{n\to\infty} \mathbf{E}\xi_n = \mathbf{E}\xi$, от (16) при $n \to \infty$ получаваме (14).

В общия случай $\xi = \xi^+ - \xi^-$, където ξ^+ и ξ^- имат разпределения от вида (15) и плътности съответно $f_{\xi^+}(x) = f_\xi(x)$ и $f_{\xi^-}(x) = f_\xi(-x)$. Тогава

$$\mathbf{E}\xi = \mathbf{E}\xi^+ - \mathbf{E}\xi^- = \int\limits_0^\infty x f_\xi(x)dx - \int\limits_0^\infty x f_\xi(-x)dx = \int\limits_{-\infty}^{+\infty} x f_\xi(x)dx.$$

Като се използва същият принцип, може да се докаже следният по-общ резултат:

Теорема 3. *Ако ξ има плътност $f_\xi(x)$, функцията $g(x)$ е непрекъсната и $\int\limits_{-\infty}^{+\infty} |g(x)| f_\xi(x)dx < \infty$, то*

$$(17)\qquad \mathbf{E}g(\xi) = \int\limits_{-\infty}^{+\infty} g(x) f_\xi(x)dx.$$

Накрая да отбележим, че формулите (3), (4), (12) и (17) са верни и в общия случай, когато бореловата функция $g(x_1, \ldots, x_n)$ е изображение на $\mathbf{R}^n$ в $\mathbf{R}^1$.

Нека случайният вектор $\xi = (\xi_1, \ldots, \xi_n)$ има многомерна функция на разпределение $F_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n)$. Тогава е в сила следната формула за пресмятане на математическото очакване:

$$\mathbf{E}g(\xi_1, \ldots, \xi_n) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} g(x_1, \ldots, x_n) dF_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n). \tag{18}$$

Специално, ако ξ има плътност $f_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n)$, то

$$\mathbf{E}g(\xi_1, \ldots, \xi_n) \tag{19}$$

$$= \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} g(x_1, \ldots, x_n) f_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n) dx_1 \ldots dx_n.$$

Доказателствата са аналогични на тези в едномерния случай.

При пресмятане на математическите очаквания често се прибягва до различни начини, за да се избегнат (12), (17), (18) и (19). Например, ако случайната величина ξ може да се представи като сума от по-прости случайни величини $\xi = \sum_{i=1}^{m} \eta_i$, се използва адитивното свойство на функционала $\mathbf{E}$, откъдето $\mathbf{E}\xi = \mathbf{E}\eta_1 + \cdots + \mathbf{E}\eta_m$.

Пример 1. Нека $\nu_n \in Bi(n, p)$. Тогава от (3.4.3) следва

$$\mathbf{E}\nu_n = \sum_{k=1}^{n} \mathbf{E}\xi_k,$$

където $\xi_k \in Bi(1, p)$. Следователно $\mathbf{E}\xi_k = p$, откъдето $\mathbf{E}\nu_n = np$, т. е. получихме същия резултат, който в пример 1.1 бе намерен много по-трудно.

Пример 2. Нека $\xi \in \mathcal{N}(0, 1)$. Тогава

$$\mathbf{E}\xi = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-x^2/2} dx = 0,$$

$$\mathbf{D}\xi = \mathbf{E}\xi^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 e^{-x^2/2} dx = \frac{-x}{\sqrt{2\pi}} e^{-x^2/2} \Big|_{-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x^2/2} dx = 1.$$

Ако $\xi \in \mathcal{N}(0,1)$, то непосредствено се проверява, че за $\eta = \sigma\xi + a$ имаме $\eta \in \mathcal{N}(a, \sigma^2)$, $\mathbf{E}\eta = a$ и $\mathbf{D}\eta = \sigma^2$.

§ 4. Математическо очакване на произведение и дисперсия на сума от случайни величини

Теорема 1. *Нека ξ и η са независими случайни величини, дефинирани във вероятностното пространство $(\Omega, \mathfrak{F}, \mathbf{P})$. Ако $\mathbf{E}|\xi\eta| < \infty$, то*

$$\mathbf{E}\xi\eta = \mathbf{E}\xi\mathbf{E}\eta. \tag{1}$$

Доказателство. 1) Нека ξ и η са индикатори, т. е. $\xi = I_A$ и $\eta = I_B$, където събитията A и B са независими. Тогава

$$\mathbf{E}\xi\eta = \mathbf{E}I_A I_B = \mathbf{E}I_{A\cap B} = \mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B) = \mathbf{E}I_A\mathbf{E}I_B = \mathbf{E}\xi\mathbf{E}\eta.$$

2) Нека ξ и η са елементарни случайни величини, т. е.

$$\xi = \sum_{i=1}^{n} x_i I_{A_i}, \qquad \eta = \sum_{j=1}^{m} y_j I_{B_j},$$

където $\{A_i, B_j\}$ са независими в съвкупност случайни събития (понеже ξ и η са независими). Тогава

$$\begin{aligned}
\mathbf{E}\xi\eta &= \mathbf{E}\left(\sum_{i=1}^{n} x_i I_{A_i} \sum_{j=1}^{m} y_j I_{B_j}\right) = \mathbf{E}\left(\sum_{i,j} x_i y_j I_{A_i} I_{B_j}\right) \\
&= \sum_{i,j} x_i y_j \mathbf{E}I_{A_i \cap B_j} = \sum_{i,j} x_i y_j \mathbf{P}(A_i \cap B_j) \\
&= \sum_{i,j} x_i y_j \mathbf{P}(A_i)\mathbf{P}(B_j) = \sum_{i,j} x_i y_j \mathbf{E}I_{A_i}\mathbf{E}I_{B_j} \\
&= \sum_i x_i \mathbf{E}I_{A_i} \sum_j y_j \mathbf{E}I_{B_j} = \mathbf{E}\left(\sum_i x_i I_{A_i}\right)\mathbf{E}\left(\sum_j y_j I_{B_j}\right) \\
&= \mathbf{E}\xi\mathbf{E}\eta.
\end{aligned}$$

3) Нека $\xi, \eta \in \mathfrak{S}^+$. Тогава от § 2.2 следва, че съществуват такива $\xi_n, \eta_n \in \mathfrak{S}_0^+$, че $\xi_n \uparrow \xi$, $\eta_n \uparrow \eta$, откъдето $\mathbf{E}\xi_n \uparrow \mathbf{E}\xi$, $\mathbf{E}\eta_n \uparrow \mathbf{E}\eta$. Съгласно 2) имаме $\mathbf{E}\xi_n\eta_n = \mathbf{E}\xi_n\mathbf{E}\eta_n$ и оттук при $n \to \infty$ получаваме

$$\mathbf{E}\xi\eta = \lim \uparrow \mathbf{E}\xi_n\eta_n = \lim \uparrow \mathbf{E}\xi_n\mathbf{E}\eta_n = \mathbf{E}\xi\mathbf{E}\eta.$$

4) В общия случай от независимостта на ξ и η следва, че случайните величини $\xi^+, \eta^+, \xi^-, \eta^-$ са независими в съвкупност. Тогава от 3) получаваме

$$\begin{aligned}
\mathbf{E}\xi\eta &= \mathbf{E}(\xi^+ - \xi^-)(\eta^+ - \eta^-) \\
&= \mathbf{E}(\xi^+\eta^+ - \xi^+\eta^- - \xi^-\eta^+ + \xi^-\eta^-) \\
&= \mathbf{E}\xi^+\mathbf{E}\eta^+ - \mathbf{E}\xi^+\mathbf{E}\eta^- - \mathbf{E}\xi^-\mathbf{E}\eta^+ + \mathbf{E}\xi^-\mathbf{E}\eta^- \\
&= \mathbf{E}\xi^+(\mathbf{E}\eta^+ - \mathbf{E}\eta^-) - \mathbf{E}\xi^-(\mathbf{E}\eta^+ - \mathbf{E}\eta^-) \\
&= (\mathbf{E}\xi^+ - \mathbf{E}\xi^-)(\mathbf{E}\eta^+ - \mathbf{E}\eta^-) \\
&= \mathbf{E}\xi\mathbf{E}\eta.
\end{aligned}$$

Следствие 1. *Ако случайните величини $\xi_1, \dots, \xi_n$ са независими и имат крайни математически очаквания, то*

$$\mathbf{E}\prod_{k=1}^{n}\xi_k = \prod_{k=1}^{n}\mathbf{E}\xi_k. \tag{2}$$

Теорема 2. *Нека ξ и η са независими случайни величини в пространството $(\Omega, \mathfrak{F}, \mathbf{P})$. Ако $\mathbf{D}\xi < \infty$ и $\mathbf{D}\eta < \infty$, то*

$$\mathbf{D}(\xi + \eta) = \mathbf{D}\xi + \mathbf{D}\eta. \tag{3}$$

Доказателство. Непосредствено от дефиниция (3.8) за произволни случайни величини ξ и η получаваме

$$\begin{aligned}
\mathbf{D}(\xi + \eta) &= \mathbf{E}[\xi + \eta - (\mathbf{E}\xi + \mathbf{E}\eta)]^2 \\
&= \mathbf{E}[(\xi - \mathbf{E}\xi)^2 + (\eta - \mathbf{E}\eta)^2 + 2(\xi - \mathbf{E}\xi)(\eta - \mathbf{E}\eta)],
\end{aligned}$$

т. е.

$$\mathbf{D}(\xi + \eta) = \mathbf{D}\xi + \mathbf{D}\eta + 2\mathrm{cov}(\xi, \eta), \tag{4}$$

където

$$\mathrm{cov}(\xi, \eta) = \mathbf{E}[(\xi - \mathbf{E}\xi)(\eta - \mathbf{E}\eta)] = \mathbf{E}\xi\eta - \mathbf{E}\xi\mathbf{E}\eta. \tag{5}$$

Прието е (5) да се нарича *ковариация* на случайните величини ξ и η, а съотношението

$$\rho(\xi, \eta) = \frac{\mathrm{cov}(\xi, \eta)}{\sqrt{\mathbf{D}\xi\mathbf{D}\eta}} \tag{6}$$

се нарича *корелационен коефициент* или само *корелация* на ξ и η.

Ако $\rho(\xi,\eta) = 0$ (т. е. при $\mathrm{cov}(\xi,\eta) = 0$), случайните величини ξ и η се наричат *некорелирани.*

Ако ξ и η са независими, то от (5) съгласно (1) следва, че $\mathrm{cov}(\xi,\eta) = 0$, т. е. от независимостта следва некорелираност. Обратното в общия случай не е вярно.

Очевидно условията на теорема 2 могат да бъдат отслабени.

Следствие 2. *Ако случайните величини ξ и η са некорелирани, то в сила е* (3).

Следствие 3. *Ако случайните величини $\xi_1, \ldots, \xi_n$ са независими в съвкупност, то*

$$\mathbf{D}\left(\sum_{k=1}^{n} \xi_k\right) = \sum_{k=1}^{n} \mathbf{D}\xi_k. \tag{7}$$

Пример 1. Нека $\nu_n \in Bi(n,p)$. Както знаем от пример (3.4.14), $\nu_n = \xi_1 + \cdots + \xi_n$, където $\{\xi_k\}$ са независими и $\xi_k \in Bi(1,p)$, т. е. $\mathbf{P}\{\xi_k = 1\} = p$, $\mathbf{P}\{\xi_k = 0\} = q = 1 - p$. Следователно

$$\mathbf{E}\xi_k = p, \quad \mathbf{D}\xi_k = \mathbf{E}\xi_k^2 - (\mathbf{E}\xi_k)^2 = p - p^2 = pq.$$

Тогава от (7) следва $\mathbf{D}\nu_n = n\mathbf{D}\xi_1 = npq$. Очевидно в този случай пресмятането на дисперсията директно от дефиницията е доста трудоемко.

§ 5. Числови характеристики и неравенства

Математическото очакване и дисперсията в много случаи не дават достатъчна информация за изследваните случайни величини и затова за по-пълната им характеристика се използват и моменти от по-висок ред.

Специално третият централен момент $\mu_3 = \mathbf{E}(\xi - \mathbf{E}\xi)^3$ характеризира несиметричността на разпределението около математическото очакване. Затова за характеристика на несиметричността на разпределението се приема безразмерна величина — отношението на третия централен момент към куба на средно-квадратичното отклонение:

$$\gamma_1(\xi) = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mathbf{E}(\xi - \mathbf{E}\xi)^3}{(\sqrt{\mathbf{D}\xi})^3}, \tag{1}$$

която се нарича *асиметрия*.

Четвъртият централен момент $\mu_4 = \mathbf{E}(\xi - \mathbf{E}\xi)^4$ определя характера на максимума в точката $\mathbf{E}\xi$ при симетрични разпределения, като за такава характеристика се приема безразмерната величина

$$\gamma_2(\xi) = \frac{\mu_4}{\mu_2^2} - 3 = \frac{\mathbf{E}(\xi - \mathbf{E}\xi)^4}{(\mathbf{D}\xi)^2} - 3, \tag{2}$$

която се нарича *ексцес*.

Смисълът на (1) и (2) най-добре се разкрива при нормалното разпределение. Ако $\eta \in \mathcal{N}(a, \sigma^2)$, то $\mu_3 = 0$, $\mu_4 = 3\mu_2^2$ (вж. зад. 6) и следователно $\gamma_1(\eta) = \gamma_2(\eta) = 0$. При $\gamma_1(\xi) > 0$ (< 0) по-голяма част от „вероятностната маса" се намира надясно (наляво) от математическото очакване. При $\gamma_2 > 0$ максимумът на плътността $f_\xi(x)$ е „по-остър" (и следователно по-голям) от този на нормалната плътност $f_\eta(x)$ с параметри $a = \mathbf{E}\xi$ и $\sigma^2 = \mathbf{D}\xi$, а при $\gamma_2(\xi) < 0$ е съответно „по-тъп" (и следователно по-малък).

За всяка функция на разпределение $F(x)$ се дефинира *p-квантил* $x(p)$, $0 < p < 1$, по следния начин:

$$x(p) = \sup\{x : F(x) \leqq p\}. \tag{3}$$

Квантилът $x(0{,}5)$ се нарича *медиана*, $x(0{,}1), \ldots, x(0{,}9)$ се наричат *децили*, а $x(0{,}01), \ldots, x(0{,}99)$ — *процентили*.

За някои разпределения най-често срещаните квантили се дават във вид на таблици, към които ще се обръщаме впоследствие неединократно (вж. таблици 1 – 5 в края на книгата).

По-нататък често ще използваме следното **неравенство на Коши – Буняковски – Шварц**:

$$\mathbf{E}|\xi\eta| \leqq \sqrt{\mathbf{E}\xi^2\mathbf{E}\eta^2}, \tag{4}$$

което се получава от неравенството $2|ab| \leqq a^2 + b^2$, като положим в него $a^2 = \dfrac{\xi^2}{\mathbf{E}\xi^2}$, $b^2 = \dfrac{\eta^2}{\mathbf{E}\eta^2}$ и след това вземем математическото очакване на двете му части.

Теорема 1. *Нека $\rho(\xi, \eta)$ е корелационният коефициент, дефиниран в* (4.6). *Тогава:*

1) $|\rho(\xi, \eta)| \leqq 1$;

2) $|\rho(\xi, \eta)| = 1$ *само когато ξ и η са линейно зависими, т. е. съществуват константи a и b, такива че $\eta = a\xi + b$. Освен това $a > 0$ при $\rho = 1$ и $a < 0$ при $\rho = -1$.*

Доказателство. Твърдението 1) следва веднага от (4) и (4.6), тъй като

$$\text{(5)}\qquad \begin{aligned}|\mathrm{cov}(\xi,\eta)| &= |\mathbf{E}(\xi-\mathbf{E}\xi)(\eta-\mathbf{E}\eta)| \leqq \mathbf{E}|(\xi-\mathbf{E}\xi)(\eta-\mathbf{E}\eta)|\\ &\leqq (\mathbf{E}(\xi-\mathbf{E}\xi)^2\mathbf{E}(\eta-\mathbf{E}\eta)^2)^{1/2} = \sqrt{\mathbf{D}\xi\mathbf{D}\eta}.\end{aligned}$$

Нека сега разгледаме

$$\text{(6)}\qquad \begin{aligned}\mathbf{E}\left(\frac{\xi-\mathbf{E}\xi}{\sqrt{\mathbf{D}\xi}} \mp \frac{\eta-\mathbf{E}\eta}{\sqrt{\mathbf{D}\eta}}\right)^2 &= \frac{\mathbf{E}(\xi-\mathbf{E}\xi)^2}{\mathbf{D}\xi} + \frac{\mathbf{E}(\eta-\mathbf{E}\eta)^2}{\mathbf{D}\eta}\\ &\mp \frac{2\mathbf{E}(\xi-\mathbf{E}\xi)(\eta-\mathbf{E}\eta)}{\sqrt{\mathbf{D}\xi\mathbf{D}\eta}} = 2(1\mp\rho(\xi,\eta)).\end{aligned}$$

Да отбележим, че за всяка случайна величина $X \geqq 0$ с $\mathbf{E}X = 0$ имаме $\mathbf{P}\{X = 0\} = 1$ (казваме „$X = 0$ *почти сигурно*“, което се означава с „п. с.“).

Тогава от (6) следва, че при $\rho(\xi,\eta) = \pm 1$ имаме

$$\frac{\eta-\mathbf{E}\eta}{\sqrt{\mathbf{D}\eta}} = \pm\frac{\xi-\mathbf{E}\xi}{\sqrt{\mathbf{D}\xi}} \quad \text{п. с.},$$

което доказва твърдението 2).

Теорема 2. *Ако* $\mathbf{E}\xi$ *е крайно и функцията* $g(x)$ *е изпъкнала (гледана отдолу), в сила е следното неравенство на Йенсен:*

$$\text{(7)}\qquad g(\mathbf{E}\xi) \leqq \mathbf{E}g(\xi).$$

Доказателство. От изпъкналостта на $g(x)$ следва, че за всяко $x_0 \in \mathbf{R}^1$ съществува константа $c = c(x_0) \in \mathbf{R}^1$, такава че

$$\text{(8)}\qquad g(x) \geqq g(x_0) + c(x - x_0).$$

Като положим в (8) $x = \xi$, $x_0 = \mathbf{E}\xi$ и вземем математическото очакване на двете страни, получаваме (7).

З а б е л е ж к а. Очевидно в (7) имаме равенство само когато ξ е константа или $g(x)$ е линейна функция.

Следствие 1. *За всички* $0 < s < r$ *е в сила неравенството на Ляпунов*

$$\text{(9)}\qquad (\mathbf{E}|\xi|^s)^{1/s} \leqq (\mathbf{E}|\xi|^r)^{1/r}.$$

Доказателство. Тъй като $g(x) = |x|^t$ при $t \geqq 1$ е изпъкнала, от (7) следва

$$\text{(10)}\qquad |\mathbf{E}\eta|^t \leqq \mathbf{E}|\eta|^t, \quad t \geqq 1.$$

От (10) при $\eta = |\xi|^s$ и $t = \dfrac{r}{s} \geqq 1$ получаваме (9). От (9) следва, че ако съществува момент от ред r, то съществуват всички моменти от ред $s < r$.

Теорема 3. *Нека $\xi \geqq 0$. Тогава за $\varepsilon > 0$ е в сила неравенството на Чебишов*

$$\mathbf{P}\{\xi \geqq \varepsilon\} \leqq \frac{\mathbf{E}\xi}{\varepsilon}. \tag{11}$$

Доказателство. За всяко $\varepsilon > 0$ имаме

$$\mathbf{E}\xi = \int\limits_{\Omega} \xi d\mathbf{P} \geqq \int\limits_{\{\xi \geqq \varepsilon\}} \xi d\mathbf{P} \geqq \varepsilon \int\limits_{\{\xi \geqq \varepsilon\}} d\mathbf{P} = \varepsilon \mathbf{P}\{\xi \geqq \varepsilon\}.$$

Следствие 2. *За всяка случайна величина η и за всяко $s > 0$ е изпълнено*

$$\mathbf{P}\{|\eta| \geqq \varepsilon\} \leqq \frac{\mathbf{E}|\eta|^s}{\varepsilon^s}. \tag{12}$$

От (12) при $s = 2$ получаваме

$$\mathbf{P}\{|\xi - \mathbf{E}\xi| \geqq \varepsilon\} \leqq \frac{\mathbf{D}\xi}{\varepsilon^2}, \tag{13}$$

което в литературата се нарича също неравенство на Чебишов.

Задачи

1. Нека $\xi \in \mathcal{U}(0,1)$ (равномерно разпределение, вж. пример 2.3.1). Покажете, че $\mathbf{E}\xi = \dfrac{1}{2}$ и $\mathbf{D}\xi = \dfrac{1}{12}$.

2. Ако $\eta \in Po(\lambda)$ (поасоново разпределение с параметър $\lambda > 0$, вж. § 3.3), покажете, че $\mathbf{E}\eta = \lambda = \mathbf{D}\eta$.

3. Нека $\xi \in G(p)$ (геометрично разпределение с параметър $0 < p < 1$, вж. зад. 2.9). Докажете, че $\mathbf{E}\xi = \dfrac{q}{p}$, $\mathbf{D}\xi = \dfrac{q}{p^2}$.

4. Нека ξ е хипергеометрично разпределена случайна величина (вж. зад. 2.1), т.е. $\xi \in HG(N, M, n)$. Покажете, че

$$\mathbf{E}\xi = \frac{Mn}{N}, \quad \mathbf{D}\xi = \frac{(N-M)M(N-n)n}{N^2(N-1)}.$$

5. Ако $\xi \in Ex(\lambda)$ (експоненциално разпределение с параметър $\lambda > 0$, зад. 2.10), покажете, че $\mathbf{E}\xi = \lambda$, $\mathbf{D}\xi = \lambda^2$.

6. Случайната величина ξ е нормално разпределена с параметри (0, т. е. $\xi \in \mathcal{N}(0, \sigma^2)$. Докажете, че

$$\mathbf{E}\xi^{2n+1} = 0, \quad \mathbf{E}\xi^{2n} = \frac{(2n)!}{n!}\left(\frac{\sigma^2}{2}\right)^n.$$

7. Ако $\xi \in \mathcal{N}(0, \sigma^2)$, покажете, че $\mathbf{E}\exp(\xi) = \exp\left(\frac{\sigma^2}{2}\right)$.

8. Случайните величини $\xi_1, \ldots, \xi_n$ са независими и $\eta = \xi_1\xi_2\ldots\xi_n$. $\mathbf{E}\xi_k = a_k$ и $\mathbf{D}\xi_k = \sigma_k^2$, $k = 1, \ldots, n$, докажете, че

$$\mathbf{D}\eta = \prod_{i=1}^{n}(a_i^2 + \sigma_i^2) - \prod_{i=1}^{n} a_i^2.$$

9. Неотрицателните случайни величини $\xi_1, \ldots, \xi_n$ са независими и наkво разпределени, а

$$\eta_k = \frac{\xi_k + \alpha}{\sum\limits_{i=1}^{n} \xi_i + n\alpha}, \quad k = 1, \ldots, n,$$

където α е положителна константа. Покажете, че $\mathbf{E}\eta_k = \dfrac{1}{n}$.

10. Нека $\xi \in \mathcal{U}(0,1)$, а $\eta = \sin(2\pi\xi)$, $\zeta = \cos(2\pi\xi)$. Докажете, че случайн величини η и ξ са некорелирани.

11. Нека $\xi \in \mathcal{U}(0,1)$ и $\eta = \min(\xi, 1 - \xi)$. Покажете, че

$$\mathbf{E}_\eta = \frac{1}{4} \quad \text{и} \quad \mathbf{E}\left(\frac{\eta}{1-\eta}\right) = \ln\left(\frac{4}{e}\right).$$

П е т а г л а в а

Разпределения на функции от случайни величини

§ 1. Теорема за съвместната плътност. Композиции

Сега ще покажем как с помощта на апарата, развит в четвърта глава, могат да се намират разпределения на функции от случайни величини.

За всяко борелово множество $B \in \mathfrak{B}_1$ съгласно теорема 4.3.1 имаме

$$\mathbf{P}\{\xi \in B\} = Q_\xi(B) = \mathbf{E}_{Q_\xi} I_B = \int_{-\infty}^{+\infty} I_B(x) dF_\xi(x) = \int_B dF_\xi(x),$$

като в случай на дискретно разпределение

$$\int_B dF_\xi(x) = \sum_{x_k \in B} p_k,$$

а в случай на абсолютно непрекъснато разпределение

$$\int_B dF_\xi(x) = \int_B f_\xi(x) dx.$$

Специално при $B = \{x : a \leqq x < b\}$ имаме

$$\mathbf{P}\{a \leqq \xi < b\} = F_\xi(b) - F_\xi(a) = \int_{[a,b)} dF_\xi(x) = \int_a^b dF_\xi(x).$$

Ако $g(x)$ е борелова функция, а ξ — случайна величина в пространството $(\Omega, \mathfrak{F}, \mathbf{P})$, както вече знаем, $\eta = g(\xi)$ е случайна величина в същото вероятностно пространство. Тогава

$$\mathbf{P}\{\eta \in B\} = \mathbf{P}\{\xi \in C\} = \int_C dF_\xi(x),$$

където $C = \{x : g(x) \in B\}$. Оттук

$$F_\eta(x) = \mathbf{P}\{g(x) < x\} = \int\limits_{\{y: g(y) < x\}} dF_\xi(y).$$

Сега, ако съществува обратната функция $g^{-1}(x)$, получаваме

$$(1) \qquad F_\eta(x) = \int\limits_{\{y: y < g^{-1}(x)\}} dF_\xi(y) = \int\limits_{-\infty}^{g^{-1}(x)} dF_\xi(y) = F_\xi(g^{-1}(x)).$$

Пример 1. Нека случайната величина ξ има непрекъсната функция на разпределение $F_\xi(x)$; тогава очевидно съществува $F_\xi^{-1}(x)$. Да образуваме случайната величина $\eta = F_\xi(\xi)$, която очевидно приема стойности в интервала $[0, 1]$. От (1) за всяко $x \in [0, 1]$ получаваме $F_\eta(x) = F_\xi(F_\xi^{-1}(x)) = x$, т. е. $\eta \in \mathcal{U}(0, 1)$ (фиг. 6).

Изложеното дотук веднага може да се обобщи в n-мерния случай. Нека $\xi_1, \ldots, \xi_n$ са случайни величини, дефинирани във вероятностното пространство $(\Omega, \mathfrak{F}, \mathbf{P})$, а $g(x_1, \ldots, x_n)$ е борелова функция от $\mathbf{R}^n$ в $\mathbf{R}^1$. Тогава, както знаем, $\eta = g(\xi_1, \ldots, \xi_n)$ е случайна величина в $(\Omega, \mathfrak{F}, \mathbf{P})$.

За всяко $B \in \mathfrak{B}_1$ можем да дефинираме

$$(2) \quad \mathbf{P}(\eta \in B) = \mathbf{P}\{(\xi_1, \ldots, \xi_n) \in C\} = \underbrace{\int \ldots \int}_{C} dF_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n),$$

където $C = \{(x_1, \ldots, x_n) : g(x_1, \ldots, x_n) \in B\}$. За дискретни разпределения имаме

$$\mathbf{P}(\eta \in B) = \sum_{(x_1, \ldots, x_n) \in C} P_{x_1, \ldots, x_n},$$

а за абсолютно непрекъснати —

$$\mathbf{P}(\eta \in B) = \underbrace{\int \ldots \int}_{C} f_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n) dx_1 \ldots dx_n.$$

Теорема 1 (за съвместната плътност). *Нека $\xi_1, \ldots, \xi_n$ са случайни величини със съвместна плътност $f_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n)$. Нека функциите $u_1 = u_1(x_1, \ldots, x_n)$, $\ldots$, $u_n = u_n(x_1, \ldots, x_n)$ определят* *взаимноеднозначно съответствие на $\mathbf{R}^n$ в $\mathbf{R}^n$, т. е. съществува обратно изображение $x_1 = x_1(u_1, \ldots, u_n)$, $\ldots$, $x_n = x_n(u_1, \ldots, u_n)$, като и двете изображения са непрекъснати; освен това съществуват непрекъснати частни производни $\dfrac{\partial x_i}{\partial u_j}$, $i, j = 1, \ldots, n$, като якобианът на трансформацията не е изроден, т. е.*

$$J = \frac{\partial(x_1, \ldots, x_n)}{\partial(u_1, \ldots, u_n)} = \begin{bmatrix} \dfrac{\partial x_1}{\partial u_1} & \cdots & \dfrac{\partial x_1}{\partial u_n} \\ \cdots & \cdots & \cdots \\ \dfrac{\partial x_n}{\partial u_1} & \cdots & \dfrac{\partial x_n}{\partial u_n} \end{bmatrix} \neq 0.$$

Тогава случайните величини $\eta_1 = u_1(\xi_1, \ldots, \xi_n), \ldots, \eta_n = u_n(\xi_1, \ldots, \xi_n)$ имат съвместна плътност

$$\begin{aligned} (3) \qquad & f_{\eta_1, \ldots, \eta_n}(u_1, \ldots, u_n) \\ &= f_{\xi_1, \ldots, \xi_n}(x_1(u_1, \ldots, u_n), \ldots, x_n(u_1, \ldots, u_n))|J|. \end{aligned}$$

Доказателство. Да означим

$$S = \{(x_1, \ldots, x_n) \in \mathbf{R}^n : u_k(x_1, \ldots, x_n) < z_k, \ k = 1, \ldots, n\}.$$

Тогава

$$\begin{aligned} (4) \qquad & F_{\eta_1, \ldots, \eta_n}(z_1, \ldots, z_n) = \mathbf{P}\{\eta_1 < z_1, \ldots, \eta_n < z_n\} \\ &= \mathbf{P}\{(\xi_1, \ldots, \xi_n) \in S\} = \underbrace{\int \cdots \int}_{S} f_{\xi_1, \ldots, \xi_n}(x_1, \ldots, x_n) dx_1 \ldots dx_n. \end{aligned}$$

Условията на теоремата ни гарантират, че в последния интеграл можем да извършим смяна на променливите

$$x_i = x_i(u_1, \ldots, u_n), \quad i = 1, \ldots, n.$$

Следователно от (4) получаваме

$$\begin{aligned} & F_{\eta_1, \ldots, \eta_n}(z_1, \ldots, z_n) \\ (5) \qquad &= \int_{-\infty}^{z_1} \cdots \int_{-\infty}^{z_n} f_{\xi_1, \ldots, \xi_n}(x_1(u_1, \ldots, u_n), \ldots, x_n(u_1, \ldots, u_n)) \\ &\times \left| \frac{\partial(x_1, \ldots, x_n)}{\partial(u_1, \ldots, u_n)} \right| du_1 \ldots du_n. \end{aligned}$$

Съотношението (5) показва, че случайните величини $\eta_1, \ldots, \eta_n$ имат съвместна плътност, която се представя чрез (3).

Следствие 1. *Ако случайните величини* $\xi_1, \ldots, \xi_n$ *са независими, при условията на теорема 1 е изпълнено*

$$f_{\eta_1,\ldots,\eta_n}(u_1, \ldots, u_n) = \prod_{i=1}^{n} f_{\xi_i}(x_i(u_1, \ldots, u_n))|J|, \tag{6}$$

т. е. тръгвайки от едномерни плътности, можем да конструираме многомерни съвместни плътности.

Ще илюстрираме (6) като намерим разпределението на сумата от две независими случайни величини $\zeta = \xi + \eta$. Нека първоначално ξ и η имат плътности $f_\xi(x)$ и $f_\eta(x)$. Тогава за съвместната плътност получаваме $f_{\xi,\eta}(x,y) = f_\xi(x)f_\eta(y)$. Да направим трансформацията $u = x + y$, $v = y$. Тогава $x = u - v$, $y = v$ е обратната трансформация, якобианът на която е $J = \dfrac{\partial(x,y)}{\partial(u,v)} = 1$. Съгласно съотношение (6)

$$f_{\xi+\eta,\eta}(u,v) = f_{\xi,\eta}(u-v,v) = f_\xi(u-v)f_\eta(v).$$

Оттук намираме маргиналната плътност на $\xi + \eta$:

$$f_{\xi+\eta}(u) = \int_{-\infty}^{\infty} f_{\xi+\eta,\eta}(u,v)dv = \int_{-\infty}^{\infty} f_\xi(u-v)f_\eta(v)dv. \tag{7}$$

Последният интеграл се означава символично с $f_\xi * f_\eta$ и се нарича *композиция на функциите* f_ξ и f_η. Ако ξ и η са неотрицателни случайни величини, т. е. $f_\xi(x) = f_\eta(x) = 0$, $x < 0$, от (7) следва

$$f_{\xi+\eta}(u) = \int_0^u f_\xi(u-v)f_\eta(v)dv. \tag{8}$$

От (7), като интегрираме в граници от $-\infty$ до x, получаваме

$$F_{\xi+\eta}(x) = \int_{-\infty}^{+\infty} F_\xi(x-v)dF_\eta(v) = (F_\xi * F_\eta)(x). \tag{9}$$

Очевидно от (8) следва

$$F_{\xi+\eta}(x) = \int_0^x F_\xi(x-v)dF_\eta(v). \tag{10}$$

Съотношенията (9) и (10) са верни в общия случай, но ние няма да привеждаме прецизни доказателства, а ще дадем по-скоро едно „правдоподобно" обяснение:

От (2) получаваме

$$F_{\xi+\eta}(t) = \underbrace{\int \dots \int}_{C} dF_{\xi,\eta}(x,y), \tag{11}$$

където $C = \{(x,y) : x+y < t\} = \{-\infty < x < \infty,\ y < t-x\}$. Тъй като ξ и η са независими, то $F_{\xi,\eta}(x,y) = F_\xi(x)F_\eta(y)$. От друга страна,

$$\begin{aligned} dF_{\xi,\eta}(x,y) &= Q_{\xi,\eta}(dx,dy) = \mathbf{P}\{x < \xi < x+dx,\ y < \eta < y+dy\} \\ &= \mathbf{P}\{x < \xi < x+dx\}\mathbf{P}\{y < \eta < y+dy\} = dF_\xi(x)dF_\eta(y). \end{aligned}$$

Сега от (11) намираме

$$F_{\xi+\eta}(t) = \int_{-\infty}^{+\infty} \left\{ \int_{-\infty}^{t-x} dF_\eta(y) \right\} dF_\xi(x) = \int_{-\infty}^{+\infty} F_\eta(t-x)dF_\xi(x) = (F_\xi * F_\eta)(t).$$

Специално, ако ξ и η имат дискретни разпределения $\{p_n\}$ и $\{q_n\}$, $n = 0,\ 1,\ \dots,$ то

$$\mathbf{P}\{\xi+\eta = k\} = \sum_{i=0}^{k} p_i q_{k-i}, \quad k = 0,\ 1,\ 2,\ \dots, \tag{12}$$

което е аналог на (8).

Очевидно операцията композиция на разпределения „$*$" е комутативна. Освен това тя е асоциативна, т. е.

$$F_1 * (F_2 * F_3) = (F_1 * F_2) * F_3,$$

тъй като това е разпределение на сумата $\xi_1+\xi_2+\xi_3$, където ξ_1, ξ_2, ξ_3 са независими случайни величини.

Аналогично $F_1 * \dots * F_n$ е функция на разпределение на сумата $\xi_1 + \dots + \xi_n$, където $\xi_1,\ \dots,\ \xi_n$ са независими и $F_k(t) = \mathbf{P}\{\xi_k < t\}$, $k = 1, \dots, n$. В случай, че $\xi_1, \dots, \xi_n$ са еднакво разпределени, т. е. $F(t) = \mathbf{P}\{\xi_k < t\}$, $k = 1,\ \dots,\ n$, се използва означението

$$\mathbf{P}\{\xi_1 + \dots + \xi_n < t\} = F^{*n}(t).$$

Пример 2. Нека ξ и η са независими и равномерно разпределени в интервала $[0, a]$, $a > 0$, т. е. имат съответни плътности

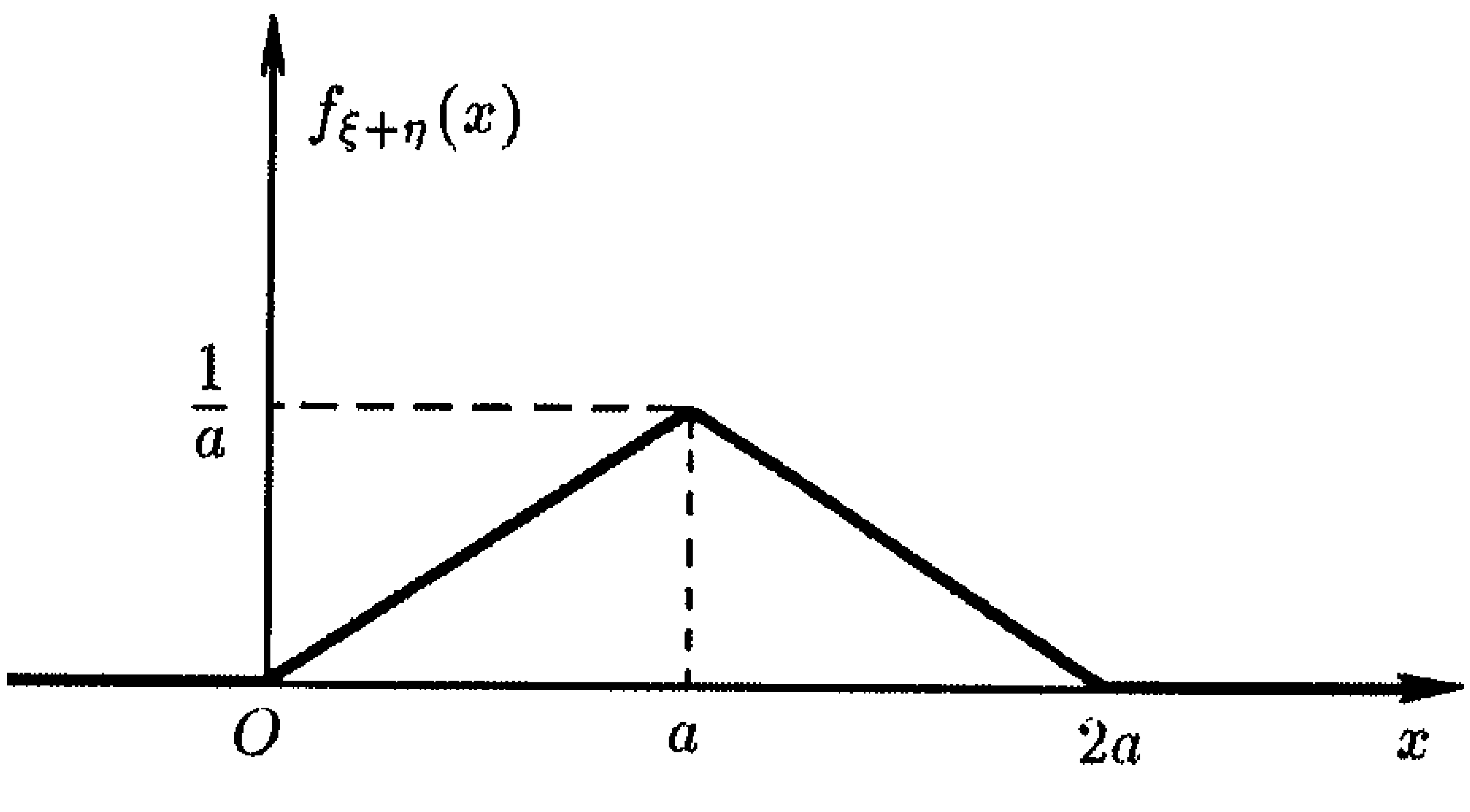
$$f_\xi(x) = f_\eta(x) = \begin{cases} \dfrac{1}{a} & \text{за } x \in [0, a], \\ 0 & \text{за } x \notin [0, a], \end{cases} \tag{13}$$

за което се използва означението $\xi \in \mathcal{U}(0, a)$. Тогава от (7) получаваме

$$f_{\xi+\eta}(x) = \int\limits_{-\infty}^{\infty} f_\xi(x-y) f_\eta(y) dy = \frac{1}{a} \int\limits_0^a f_\xi(x-y) dy = \frac{1}{a} \int\limits_{x-a}^{x} f_\xi(z) dz.$$

Последният интеграл очевидно представлява дължината на сечението на интервалите $[0, a]$ и $[x - a, x]$, разделена на a. Следователно

$$f_{\xi+\eta}(x) = \begin{cases} 0 & \text{за } x < 0, \\ \dfrac{x}{a^2} & \text{за } 0 \leqq x \leqq a, \\ \dfrac{2a - x}{a^2} & \text{за } a \leqq x \leqq 2a, \\ 0 & \text{за } x > 2a. \end{cases} \tag{14}$$



Фиг. 10

Графиката на плътността (14) представлява триъгълник (вж. фиг. 10) и се нарича *разпределение на Симпсън* или *триъгълно разпределение*. Ако въведем означението $x_+ = x$ при $x \geqq 0$ и $x_+ = 0$

при $x < 0$, за (14) получаваме

$$f_{\xi+\eta}(x) = u_2(x) = \frac{1}{a^2}(x_+ - 2(x-a)_+). \tag{15}$$

Нека $\xi_1, \ldots, \xi_n$ са независими случайни величини, $\xi_k \in \mathcal{U}(0,a)$, $k = 1, \ldots, n$. Нека $u_n(x)$ е плътността на сумата им, а $U_n(x)$ — функцията на разпределение. Тогава можем да докажем по индукция, че са в сила съотношенията

$$u_n(x) = \frac{1}{a^n(n-1)!}\left\{\sum_{k=0}^{n}(-1)^k\binom{n}{k}(x-ka)_+^{n-1}\right\}_+, \tag{16}$$

$$U_n(x) = \frac{1}{a^n n!}\sum_{k=0}^{n}(-1)^k\binom{n}{k}(x-ka)_+^{n}. \tag{17}$$

§ 2. Многомерно нормално разпределение

Сега ще направим едно важно приложение на следствие 1 от теорема 1.1.

Отсега нататък ще смятаме $\mathbf{y} = (y_1, \ldots, y_n)$ за $(n \times 1)$–матрица (т. е. вектор–стълб), транспонираният вектор $\mathbf{y}' = (y_1, \ldots, y_n)$ ще интерпретираме като $(1 \times n)$–матрица. Ако $\mathbf{X} = (X_{ij})$ е матрица от случайни величини, то $\mathbf{EX}$ е матрицата от математическите очаквания $(\mathbf{E}X_{ij})$.

Дефиниция 1. *Ще казваме, че компонентите на случайния вектор $\vec{\eta} = (\eta_1, \ldots, \eta_n)$ имат n-мерно съвместно нормално разпределение, ако съществуват n независими нормално разпределени $\mathcal{N}(0,1)$ случайни величини $\vec{\xi} = (\xi_1, \ldots, \xi_n)$, неизродена $(n \times n)$-матрица $\mathbf{A} = (a_{ij})$ и вектор от константи $\mathbf{m} = (m_1, \ldots, m_n)$, такива че*

$$\vec{\eta} = \mathbf{A}\vec{\xi} + \mathbf{m}, \tag{1}$$

т. е.

$$\eta_i = \sum_{j=1}^{n} a_{ij}\xi_j + m_i, \quad i = 1, \ldots, n. \tag{2}$$

От горната дефиниция следва, че

$$(3) \qquad f_{\xi_1,\dots,\xi_n}(x_1,\,\dots,\,x_n) = \prod_{k=1}^{n} f_{\xi_k}(x_k) = \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{(\mathbf{x},\mathbf{x})}{2}\right\},$$

където $\mathbf{x} = (x_1,\,\dots,\,x_n)$, $(\mathbf{x},\mathbf{x}) = \mathbf{x}'\mathbf{x} = \sum_{i=1}^{n} x_i^2$.

В (3) ще извършим трансформацията

$$(4) \qquad \mathbf{x} = \mathbf{A}^{-1}\cdot(\mathbf{u}-\mathbf{m}).$$

Ако положим $\mathbf{A}^{-1} = (b_{ij})$, то от (4) веднага се вижда, че $\dfrac{\partial x_i}{\partial u_j} = b_{ij}$, т. е. якобианът на трансформацията (4) е $\mathbf{J} = |\mathbf{A}^{-1}| = \det \mathbf{A}^{-1}$. Тогава съгласно (1.6) от (1), (3) и (4) получаваме

$$(5) \qquad \begin{aligned} &f_{\eta_1,\dots,\eta_n}(u_1,\,\dots,\,u_n) \\ &= \frac{|\det \mathbf{A}^{-1}|}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}(\mathbf{A}^{-1}\cdot(\mathbf{u}-\mathbf{m}),\quad \mathbf{A}^{-1}\cdot(\mathbf{u}-\mathbf{m}))\right\}, \end{aligned}$$

където $(\mathbf{A}^{-1}\cdot(\mathbf{u}-\mathbf{m}),\, \mathbf{A}^{-1}\cdot(\mathbf{u}-\mathbf{m})) = (\mathbf{u}-\mathbf{m})'(\mathbf{A}^{-1})'\mathbf{A}^{-1}(\mathbf{u}-\mathbf{m})$.

Да дефинираме матрица $\mathbf{C} = \mathbf{A}\mathbf{A}'$. Тъй като $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$, то $\mathbf{C}^{-1} = (\mathbf{A}\mathbf{A}')^{-1} = (\mathbf{A}')^{-1}(\mathbf{A})^{-1} = (\mathbf{A}^{-1})'\mathbf{A}^{-1}$. Следователно $|\mathbf{C}^{-1}| = |(\mathbf{A}^{-1})'||\mathbf{A}^{-1}| = |\mathbf{A}^{-1}|^2$, т. е. $|\det \mathbf{A}^{-1}| = \sqrt{|\mathbf{C}^{-1}|}$.

От друга страна, матрицата $\mathbf{C}^{-1}$ е симетрична и положително дефинитна. Действително

$$(\mathbf{C}^{-1})' = ((\mathbf{A}\mathbf{A}')^{-1})' = ((\mathbf{A}\mathbf{A}')')^{-1}) = (\mathbf{A}\mathbf{A}')^{-1} = \mathbf{C}^{-1}$$

и при $\mathbf{x} = (x_1,\,\dots,\,x_n) \neq 0$ имаме

$$\mathbf{x}'\mathbf{C}^{-1}\mathbf{x} = \mathbf{x}'(A^{-1})'\mathbf{A}^{-1}\mathbf{x} = (\mathbf{A}^{-1}\mathbf{x})'(\mathbf{A}^{-1}\mathbf{x}) > 0.$$

Тогава от (5) получаваме

$$(6) \qquad f_\eta(u) = \frac{\sqrt{|\mathbf{C}^{-1}|}}{(2\pi)^{(n/2)}} \exp\left\{-\frac{1}{2}(\mathbf{u}-\mathbf{m})'\mathbf{C}^{-1}(\mathbf{u}-\mathbf{m})\right\}.$$

От (2) непосредствено се вижда, че $\mathbf{E}\eta_i = m_i$, $i = 1,\,\dots,\,n$, т. е. $\mathbf{E}\vec{\eta} = \mathbf{m}$. От (1) за матрицата от вторите моменти на $\vec{\eta}$ получаваме

$$\begin{aligned} \mathbf{E}(\vec{\eta}-\mathbf{m})(\vec{\eta}-\mathbf{m})' &= (\mathbf{E}(\eta_i - m_i)(\eta_j - m_j)) = \mathbf{E}(\mathbf{A}\vec{\xi})(\mathbf{A}\vec{\xi})' \\ &= \mathbf{E}\mathbf{A}\vec{\xi}(\vec{\xi}'\mathbf{A}') = \mathbf{A}(\mathbf{E}\vec{\xi}\vec{\xi}')\mathbf{A}'. \end{aligned}$$

Тъй като случайните величини $\xi_i \in \mathcal{N}(0,1)$, $i = 1, \dots, n$, са независими помежду си, то $\mathbf{E}\xi_i\xi_j = \delta_{ij}$, където $\delta_{ij} = 1$ при $i = j$ и $\delta_{ij} = 0$ при $i \neq j$. Следователно $\mathbf{E}\vec{\xi}'\vec{\xi} = (\mathbf{E}\xi_i\xi_j) = (\delta_{ij}) = \mathbf{I}$, където $\mathbf{I}$ е единична $(n \times n)$-матрица. Тогава

$$\mathbf{E}(\vec{\eta} - \mathbf{m})(\vec{\eta} - \mathbf{m})' = \mathbf{AIA}' = \mathbf{AA}' = \mathbf{C}. \tag{7}$$

Получените резултати можем да оформим в следната теорема.

Теорема 1. *Ако* $\vec{\eta} = (\eta_1, \dots, \eta_n)$ *имат съвместно нормално разпределение, съвместната им плътност се определя от* (6), *където* $\mathbf{m} = \mathbf{E}\vec{\eta}$ *е векторът от първите моменти, а* $\mathbf{C} = \mathbf{E}(\vec{\eta} - \mathbf{m})(\vec{\eta} - \mathbf{m})'$ *е матрицата от ковариациите* $\operatorname{cov}(\eta_i, \eta_j) = \mathbf{E}(\eta_i - m_i)(\eta_j - m_j)$, $i, j = 1, \dots, n$.

Читателят може непосредствено да провери, че

$$f_{\eta_1,\eta_2}(u_1, u_2)$$

$$= \frac{\exp\left\{-\frac{1}{2(1-\rho^2)}\left(-\frac{(u_1-m_1)^2}{\sigma_1^2} - 2\rho\frac{(u_1-m_1)(u_2-m_2)}{\sigma_1\sigma_2} + \frac{(u_2-m_2)^2}{\sigma_2^2}\right)\right\}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}, \tag{8}$$

където

$$m_1 = \mathbf{E}\eta_1,\ m_2 = \mathbf{E}\eta_2,\ \sigma_1 = \mathbf{D}\eta_1,\ \sigma_2 = \mathbf{D}\eta_2,$$

а $\rho = \dfrac{\operatorname{cov}(\eta_1, \eta_2)}{\sigma_1\sigma_2}$ е корелационният коефициент.

При $\rho = 0$ от (8) следва, че $f_{\eta_1,\eta_2}(u_1, u_2) = f_{\eta_1}(u_1) f_{\eta_2}(u_2)$, където

$$f_{\eta_i}(u_i) = \frac{\exp\left\{-\frac{1}{2}(u_i - m_i)^2\right\}}{\sqrt{2\pi\sigma_i^2}},\quad i = 1, 2.$$

Следователно при съвместна нормалност от некорелираност (т. е. $\rho = 0$) следва независимост (което в общия случай не е вярно). С други думи, при нормално разпределени случайни величини понятията независимост и некорелираност са еквивалентни.

§ 3. Условно разпределение и условно математическо очакване

Нека (ξ, η) е случаен вектор във вероятностното пространство $(\Omega, \mathfrak{F}, \mathbf{P})$.

Ще въведем понятието *условно разпределение* първо в случая, когато случайните величини ξ и η са дискретни, т. е.

$$\text{(1)} \qquad \mathbf{P}\{\xi = x_i,\ \eta = y_j\} = p_{ij}, \qquad \sum_{i,j=1}^{\infty} p_{ij} = 1,$$

$$\mathbf{P}\{\xi = x_i\} = \sum_{j=1}^{\infty} p_{ij} = q_i, \qquad \mathbf{P}\{\eta = y_j\} = \sum_{i=1}^{\infty} p_{ij} = r_j.$$

Съгласно дефиниция (1.6.1)

$$\text{(2)} \qquad \mathbf{P}\{\xi = x_i \,|\, \eta = y_j\} = \frac{\mathbf{P}\{\xi = x_i,\ \eta = y_j\}}{\mathbf{P}\{\eta = y_j\}} = \frac{p_{ij}}{r_j}.$$

Ако фиксираме y_j, то (2) може да се разглежда като условно разпределение на ξ при условие, че $\eta = y_j$. В такъв случай условното математическо очакване на ξ при условие $\eta = y_j$ се дефинира по следния начин:

$$\text{(3)} \qquad \mathbf{E}(\xi \,|\, \eta = y_j) = \sum_{i=1}^{\infty} x_i \mathbf{P}\{\xi = x_i \,|\, \eta = y_j\} = \sum_{i=1}^{\infty} \frac{x_i p_{ij}}{r_j}.$$

Ако левите страни на (2) и (3) се разглеждат като функции на y_j, условното разпределение и условното математическо очакване са случайни величини в $(\Omega, \mathfrak{F}, \mathbf{P})$, т. е. за всяко $\omega \in \{\eta = y_j\}$, $j = 1, 2, \ldots,$ имаме

$$\text{(4)} \qquad \mathbf{P}\{\xi = x_i \,|\, \eta\} = \mathbf{P}\{\xi = x_i \,|\, \eta = y_j\}, \qquad \mathbf{E}(\xi \,|\, \eta) = \mathbf{E}(\xi \,|\, \eta = y_j),$$

като всяка от тези стойности е със съответната вероятност $\mathbf{P}\{\eta = y_j\}$.

Теорема 1 (формула за пълното математическо очакване).

$$\text{(5)} \qquad \mathbf{E}\xi = \mathbf{E}\{\mathbf{E}(\xi \,|\, \eta)\}.$$

Доказателство. От формулата за пресмятане на математическото очакване (4.3.12) следва

$$\text{(6)} \qquad \mathbf{E}\{\mathbf{E}(\xi \,|\, \eta)\} = \sum_{j=1}^{\infty} \mathbf{E}(\xi \,|\, \eta = y_j) \mathbf{P}\{\eta = y_j\}.$$

От (6), като приложим формула (3), получаваме

$$\begin{aligned}
\mathbf{E}\{\mathbf{E}(\xi \mid \eta)\} &= \sum_{j=1}^{\infty} \mathbf{P}\{\eta = y_j\} \sum_{i=1}^{\infty} \frac{x_i p_{ij}}{r_j} \\
&= \sum_{i,j=1}^{\infty} x_i \mathbf{P}\{\xi = x_i, \eta = y_j\} = \sum_{i=1}^{\infty} x_i \sum_{j=1}^{\infty} \mathbf{P}\{\xi = x_i, \eta = y_j\} \\
&= \sum_{i=1}^{\infty} x_i \mathbf{P}\{\xi = x_i\} = \mathbf{E}\xi.
\end{aligned}$$

Тук сме използвали факта, че

$$\sum_{j=1}^{\infty} \mathbf{P}\{\xi=x_i, \eta=y_j\} = \mathbf{P}\{\xi=x_i, \sum_{j=1}^{\infty}\{\eta=y_j\}\} = \mathbf{P}\{\xi=x_i, \Omega\} = \mathbf{P}\{\xi=x_i\}.$$

Следните свойства на условните математически очаквания се получават непосредствено от (3) и (4) и са верни за всяко $\omega \in \Omega$:

(7) $$\mathbf{E}(\xi \mid \eta) = \mathbf{E}\xi \quad \text{при независимост на } \xi \text{ и } \eta,$$

(8) $$\mathbf{E}(\xi_1 + \xi_2 \mid \eta) = \mathbf{E}(\xi_1 \mid \eta) + \mathbf{E}(\xi_2 \mid \eta),$$

(9) $$\mathbf{E}(\varphi(\eta) \mid \eta) = \varphi(\eta),$$

(10) $$\mathbf{E}(\varphi(\eta)\xi \mid \eta) = \varphi(\eta)\mathbf{E}(\xi \mid \eta),$$

за всяка борелова функция $\varphi(x)$.

Нека сега случайните величини ξ и η имат съвместно абсолютно непрекъснато разпределение, т. е. вместо (1) са в сила съотношенията

(11) $$\begin{gathered}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\xi,\eta}(x,y)\,dxdy = 1, \\
f_\xi(x) = \int_{-\infty}^{\infty} f_{\xi,\eta}(x,y)\,dy, \quad f_\eta(y) = \int_{-\infty}^{\infty} f_{\xi,\eta}(x,y)\,dx.
\end{gathered}$$

Тъй като в този случай $\mathbf{P}\{\eta = y\} = 0$ при всяко $y \in \mathbb{R}^1$, не можем да използваме дефиниция (1.6.1) за условна вероятност, както това беше направено в дискретния случай.

Дефиниция 1. *Условна плътност на случайната величина ξ при условие, че $\eta = y$, $f_\eta(y) > 0$, ще наричаме функцията*

$$f_\xi(x \mid \eta = y) = \frac{f_{\xi,\eta}(x,y)}{f_\eta(y)}. \tag{12}$$

Условно математическо очакване на ξ при условие, че $\eta = y$, $f_\eta(y) > 0$, се дефинира с формулата

$$\mathbf{E}(\xi \mid \eta = y) = \int_{-\infty}^{\infty} x f_\xi(x \mid \eta = y)dx. \tag{13}$$

Аналогично на дискретния случай (12) и (13) могат да се разглеждат като случайни величини, където за всяко $y \in \mathbf{R}^1$ имаме

$$\begin{aligned} f_\xi(x \mid \eta) &= f_\xi(x \mid \eta = y) \text{ за } \omega \in \{\eta = y\}, \\ \mathbf{E}(\xi \mid \eta) &= \mathbf{E}(\xi \mid \eta = y) \text{ за } \omega \in \{\eta = y\}. \end{aligned} \tag{14}$$

Тогава

$$\mathbf{P}\{a \leqq \xi \leqq b\} = \int_{-\infty}^{\infty} f_\eta(y) \left(\int_a^b f_\xi(x \mid \eta = y)dx \right) dy = \mathbf{E} \int_a^b f_\xi(x \mid \eta)dx. \tag{15}$$

Като въведем означението $F_\xi(x \mid \eta) = \int_{-\infty}^{x} f_\xi(t \mid \eta)dt$, от (15) получаваме $F_\xi(x) = \mathbf{E}F_\xi(x \mid \eta)$.

Ще покажем, че е в сила формула (5). Действително от равенствата (4.3.17), (11) и (12) следва

$$\begin{aligned} \mathbf{E}\{\mathbf{E}(\xi \mid \eta)\} &= \int_{-\infty}^{\infty} \mathbf{E}(\xi \mid \eta = y) f_\eta(y)dy \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x f_\xi(x \mid \eta = y)dx \right) f_\eta(y)dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{\xi,\eta}(x,y)dxdy = \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{\xi,\eta}(x,y)dy \right) dx \\ &= \int_{-\infty}^{\infty} x f_\xi(x)dx = \mathbf{E}\xi. \end{aligned}$$

Аналогично се запазват свойствата (7) — (10).

Да отбележим, че понякога $\mathbf{E}(\xi \mid \eta)$ се нарича *регресия на ξ спрямо η.*

Пример 1. Нека ξ_1, ξ_2, ... са еднакво разпределени случайни величини с крайно математическо очакване $a = \mathbf{E}\xi_1$, а ν е независима от тях целочислена положителна случайна величина. Тогава дефинираме случайната величина $\eta_\nu = \xi_1 + \cdots + \xi_\nu$, т. е. сума от случаен брой случайни величини. За да намерим $\mathbf{E}\eta_\nu$, ще използваме формули (5) и (6).

$$\mathbf{E}\eta_\nu = \mathbf{E}\{\mathbf{E}(\eta_\nu \mid \nu)\} = \sum_{k=1}^{\infty} \mathbf{P}\{\nu = k\}\mathbf{E}(\eta_\nu \mid \nu = k)$$

$$= \sum_{k=1}^{\infty} \mathbf{P}\{\nu = k\}\mathbf{E}(\xi_1 + \cdots + \xi_k) = \sum_{k=1}^{\infty} \mathbf{P}\{\nu = k\}ka = a\mathbf{E}\nu,$$

т. е.

$$\mathbf{E}(\xi_1 + \cdots + \xi_\nu) = \mathbf{E}\xi_1 \cdot \mathbf{E}\nu. \tag{16}$$

Формула (16) се нарича *тъждество на Валд.*

§ 4. Някои често срещани разпределения

В този раздел ще направим приложение на теоремата за съвместната плътност от § 1, като ще намерим някои важни за статистиката разпределения.

1. Гама-разпределение.

Дефиниция 1. *Ще казваме, че случайната величина ξ има Γ-разпределение с параметри $\alpha > 0$ и $\beta > 0$ (означаваме $\xi \in \Gamma(\alpha, \beta)$), ако ξ има плътност*

$$f_\xi(x) = \begin{cases} \dfrac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)} & \text{за } x \geqq 0, \\ 0 & \text{за } x < 0, \end{cases} \tag{1}$$

където гама-функцията се дефинира с $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$, и при $\alpha = n$ (n -цяло число) имаме $\Gamma(n) = (n-1)!$.

За математическото очакване от (1) по формула (4.3.14) получаваме

$$\mathbf{E}\xi = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty x^\alpha e^{-x/\beta} dx$$

$$= \frac{1}{\Gamma(\alpha)} \int_0^\infty \left(\frac{x}{\beta}\right)^\alpha e^{-x/\beta} dx = \frac{\beta\Gamma(\alpha+1)}{\Gamma(\alpha)} = \alpha\beta.$$

Аналогично по формула (4.3.17) намираме $\mathbf{E}\xi^2 = \alpha(\alpha+1)\beta^2$, откъдето

$$\mathbf{D}\xi = \mathbf{E}\xi^2 - (\mathbf{E}\xi)^2 = \alpha\beta^2.$$

2. χ^2-разпределение. Така се нарича разпределението на случайната величина $\chi_n^2 = \xi_1^1 + \cdots + \xi_n^2$ където $\{\xi_k\}$ са независими случайни величини, $\xi_k \in \mathcal{N}(0,1)$. Означава се $\chi_n^2 \in \chi^2(n)$ и се чете „хи-квадрат с n степени на свобода“. Съгласно (2.6.5) съвместната плътност на $\xi_1, \ldots, \xi_n$ се дава с

$$f_{\xi_1,\ldots,\xi_n}(x_1, \ldots, x_n) = (2\pi)^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^n x_i^2\right\}, \tag{2}$$

където $x_i \in \mathbf{R}^1$, $i = 1, \ldots, n$.

Да разгледаме трансформацията

$$\begin{aligned} x_1 &= \rho \prod_{k=1}^{n-1} \cos\theta_k, \\ x_i &= \rho \prod_{k=1}^{n-i} \cos\theta_k \sin\theta_{n-k+1}, \quad i = 2, \ldots, n-1, \\ x_n &= \rho \sin\theta_1, \end{aligned} \tag{3}$$

където $\rho > 0$, $0 \leqq \theta_1 \leqq 2\pi$, $0 \leqq \theta_k \leqq \pi$, $k = 2, \ldots, n-1$.

Непосредствено се проверява, че $\sum\limits_{k=1}^n x_k^2 = \rho^2$ и за якобиана на трансформацията имаме

$$\mathbf{J} = \frac{\partial(x_1, \ldots, x_n)}{\partial(\rho, \theta_1, \ldots, \theta_{n-1})} = \rho^{n-1} K(\theta_1, \ldots, \theta_{n-1}),$$

където $K(\theta_1, \ldots, \theta_{n-1})$ е функция само на $\theta_1, \ldots, \theta_{n-1}$.

Да разгледаме сега случайните величини $\chi_n, \eta_1, \ldots, \eta_{n-1}$, които се получават, като в (3) заместим ρ с χ_n, θ_i с η_i и x_i с ξ_i, при $i = 1, \ldots, n$. Тогава съгласно теорема 1.1 съвместната плътност на $\chi_n, \eta_1, \ldots, \eta_{n-1}$ се изразява чрез (2) и (1.3) по формулата

$$f_{\chi_n,\eta_1,\ldots,\eta_{n-1}}(\rho,\theta_1,\ldots,\theta_{n-1}) =$$

$$= \begin{cases} (2\pi)^{-n/2}\rho^{n-1}\exp\left\{-\frac{\rho^2}{2}\right\}|K(\theta_1,\ldots,\theta_{n-1})| & \text{при } \rho > 0, \\ & 0 \leqq \theta_1 \leqq 2\pi,\ 0 \leqq \theta_i \leqq \pi,\ i = 2, \ldots, n-1; \\ 0 & \text{в противен случай.} \end{cases} \tag{4}$$

От (4) за маргиналната плътност на χ_n намираме

$$\begin{aligned} f_{\chi_n}(\rho) &= \int_0^{2\pi}\int_0^{\pi}\cdots\int_0^{\pi} f_{\chi_n,\eta_1,\ldots,\eta_{n-1}}(\rho,\theta_1,\ldots,\theta_{n-1})d\theta_1\ldots d\theta_{n-1} \\ &= \begin{cases} C(2\pi)^{-n/2}\rho^{n-1}\exp\left\{-\frac{\rho^2}{2}\right\} & \text{при } \rho > 0, \\ 0 & \text{при } \rho \leqq 0, \end{cases} \end{aligned} \tag{5}$$

където $C = \int_0^{2\pi}\int_0^{\pi}\cdots\int_0^{\pi} |K(\theta_1, \ldots, \theta_{n-1})|d\theta_1 \ldots d\theta_{n-1}$.

Вместо да използваме горната формула, константата C най-лесно се пресмята от условието

$$\int_{-\infty}^{\infty} f_{\chi_n}(\rho)d\rho = 1,$$

откъдето следва, че

$$C = \frac{(2\pi)^{n/2}}{\int_0^{\infty} \rho^{n-1}\exp\left\{-\frac{\rho^2}{2}\right\}d\rho}.$$

В последния интеграл след смяната $u = \frac{\rho^2}{2}$ получаваме

$$C = \frac{(2\pi)^{n/2}}{2^{n/2-1}\Gamma\left(\frac{n}{2}\right)}.$$

Оттук и от (5) окончателно намираме

$$(6) \qquad f_{\chi_n}(\rho) = \begin{cases} \dfrac{\rho^{n-1}\exp\left\{-\dfrac{\rho^2}{2}\right\}}{2^{n/2-1}\Gamma\left(\dfrac{n}{2}\right)} & \text{при } \rho > 0, \\ 0 & \text{в противен случай.} \end{cases}$$

Тъй като

$$F_{\chi_n^2} = \mathbf{P}\{\chi_n^2 < x\} = \mathbf{P}\{\chi_n < \sqrt{x}\} = \int\limits_0^{\sqrt{x}} f_{\chi_n}(\rho)d\rho,$$

то за $f_{\chi_n^2}(x) = \dfrac{d}{dx}F_{\chi_n^2}(x)$ получаваме

$$(7) \qquad f_{\chi_n^2}(x) = \begin{cases} \dfrac{x^{n/2-1}e^{-x/2}}{2^{n/2}\Gamma\left(\dfrac{n}{2}\right)} & \text{при } x > 0, \\ 0 & \text{при } x \leqq 0. \end{cases}$$

Ако сравним (1) и (7), веднага се вижда, че $\chi^2(n) = \Gamma\left(\dfrac{n}{2}, 2\right)$, т. е. χ^2-разпределението е специален случай на Γ-разпределение. Тогава $\mathbf{E}\chi_n^2 = n$ и $\mathbf{D}\chi_n^2 = 2n$.

3. Разпределение на Стюдънт (t-разпределение). Така се нарича разпределението на случайната величина

$$T = \frac{\xi}{\sqrt{\dfrac{\zeta}{n}}},$$

където ξ и ζ са независими случайни величини, $\xi \in \mathcal{N}(0,1)$, $\zeta \in \chi^2(n)$. Тогава за случайната величина T се казва, че има t-разпределение с n степени на свобода и се означава $T \in t(n)$.

Ако означим $\eta = \sqrt{\dfrac{\zeta}{n}} = \dfrac{\chi_n}{\sqrt{n}}$, от (6), аналогично на (7), получаваме

$$(8) \qquad f_\eta(y) = \begin{cases} \dfrac{y^{n-1}\exp\left\{-\dfrac{ny^2}{2}\right\} 2\left(\dfrac{n}{2}\right)^{n/2}}{\Gamma\left(\dfrac{n}{2}\right)} & \text{при } y > 0, \\ 0 & \text{при } y \leqq 0. \end{cases}$$

Тогава съвместната плътност на ξ и ζ ще бъде

$$f_{\xi,\eta}(x,y) = f_\xi(x)f_\eta(y) =$$

$$(9)\quad = \begin{cases} \dfrac{2\left(\frac{n}{2}\right)^{n/2}}{\sqrt{2\pi}\Gamma\left(\frac{n}{2}\right)} y^{n-1}\exp\left\{-\dfrac{x^2+ny^2}{2}\right\} & \text{при } -\infty<x<\infty,\ y>0, \\ 0 & \text{в противен случай.} \end{cases}$$

За трансформацията $t = \dfrac{x}{y}$, $s = y$ (т. е. $x = ts$, $y = s$) лесно се вижда, че якобианът е $\dfrac{\partial(x,y)}{\partial(t,s)} = s$. Тогава по теорема 1.1 от (9) получаваме

$$f_{T,\eta}(t,s) = f_{\xi,\eta}(ts,s)s = \frac{2\left(\frac{n}{2}\right)^{n/2}}{\sqrt{2\pi}\Gamma\left(\frac{n}{2}\right)} s^n \exp\left\{-\frac{s^2(t^2+n^2)}{2}\right\}$$

при $-\infty < t < \infty$, $s > 0$ и $f_{T,\eta}(t,s) = 0$ в противен случай. Оттук за маргиналната плътност на T намираме

$$(10)\quad f_T(t) = \int_0^\infty f_{T,\eta}(t,s)ds = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}}\left(1+\frac{t^2}{n}\right)^{-(n+1)/2} \quad \text{за } t \in \mathbf{R}^1.$$

При $n = 1$ разпределението $t(1)$ е известно като *разпределение на Коши*.

Не е трудно да се покаже, че математическото очакване на $T \in t(n)$ съществува при $n \geqq 2$ и $\mathbf{E}T = 0$, а дисперсията съществува при $n \geqq 3$ и $\mathbf{D}T = \dfrac{n}{n-2}$.

4. Разпределение на Снедекор–Фишер (F-разпределение). Така се нарича разпределението на случайната величина

$$\zeta = \frac{\frac{\xi}{m}}{\frac{\eta}{n}},$$

където ξ и η са независими случайни величини, $\xi \in \chi^2(m)$, $\eta \in \chi^2(n)$. Означава се $\zeta \in F(m,n)$ и се казва, че има F-разпределение с m и n степени на свобода.

От (7) следва, че съвместната плътност на ξ и η е

$$
(11) \qquad f_{\xi,\eta}(x,y) = \begin{cases} \dfrac{\left(\dfrac{x}{2}\right)^{\frac{m}{2}-1}\left(\dfrac{y}{2}\right)^{\frac{n}{2}-1}\exp\left\{-\dfrac{x+y}{2}\right\}}{4\Gamma\left(\dfrac{m}{2}\right)\Gamma\left(\dfrac{n}{2}\right)} & \text{при } x,y>0, \\ 0 & \text{при } x,y \leqq 0. \end{cases}
$$

Разглеждаме трансформацията

$$
u = \frac{nx}{my}, \; v = y \iff x = \frac{uvm}{n}, \; y = v,
$$

якобианът на която е $\dfrac{\partial(x,y)}{\partial(u,v)} = \dfrac{m}{n}v$. Тогава от (11) по теорема 1.1 получаваме

$$
\begin{aligned}
f_{\zeta,\eta}(u,v) &= f_{\xi,\eta}\left(\frac{m}{n}uv, v\right)\frac{m}{n}v \\
&= \frac{\left(\dfrac{m}{n}\right)^{m/2} u^{m/2-1}}{2^{(m+n)/2}\Gamma\left(\dfrac{m}{2}\right)\Gamma\left(\dfrac{n}{2}\right)} \exp\left\{-\frac{v}{2}\left(1+\frac{mu}{n}\right)\right\} v^{(m+n)/2-1}
\end{aligned}
$$

при $u,v>0$ и $f_{\zeta,\eta}(u,v)=0$ в противен случай.

Оттук за маргиналната плътност на ζ при $u>0$ получаваме

$$
f_\zeta(u) = \int_{-\infty}^{\infty} f_{\zeta,\eta}(u,v)dv.
$$

В последния интеграл чрез смяната $z = \dfrac{1}{2}v\left(1+\dfrac{mu}{n}\right)$ и като се използва представянето на Г-функцията, се получава окончателно

$$
(12) \quad f_\zeta(u) = \begin{cases} \dfrac{m^{\frac{m}{2}} n^{\frac{n}{2}}\Gamma\left(\dfrac{m+n}{2}\right)}{\Gamma\left(\dfrac{m}{2}\right)\Gamma\left(\dfrac{n}{2}\right)} u^{\frac{m}{2}-1}(mu+n)^{-\frac{m+n}{2}} & \text{за } u>0, \\ 0 & \text{за } u \leqq 0. \end{cases}
$$

Не е трудно да се покаже, че математическото очакване на $\zeta \in F(m,n)$ съществува при $n \geqq 3$ и $\mathbf{E}\zeta = \frac{n}{n-2}$.

Поради широката им употреба в практиката (което е показано във втората част на книгата) разпределенията χ^2, t и F са подробно табулирани. Част от най-често употребяваните квантили на тези разпределения са приведени в таблиците на края на книгата.

5. В–разпределение.

Дефиниция 2. *Ще казваме, че случайната величина ζ има* В-*разпределение (четем „бета-разпределение") с параметри $p > 0$ и $q > 0$, ако има плътност*

$$(13) \qquad f_\zeta(x) = \begin{cases} \dfrac{x^{p-1}(1-x)^{q-1}}{\mathrm{B}(p,q)} & \text{за } 0 \leqq x \leqq 1, \\ 0 & \text{в противен случай,} \end{cases}$$

където

$$\mathrm{B}(p,q) = \int_0^1 x^{p-1}(1-x)^{q-1}dx = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$$

е познатата от анализа бета-функция на Ойлер.

Не е трудно да се пресметне, че

$$(14) \qquad \mathbf{E}\zeta = \frac{p}{p+q}, \qquad \mathbf{D}\zeta = \frac{pq}{(p+q)^2(p+q+1)}.$$

Задачи

1. Докажете формулите (1.16) и (1.17).

2. Нека

$$f_{\xi,\eta}(x,y) = \begin{cases} 4xy & \text{за } 0 < x < 1,\ 0 < y < 1, \\ 0 & \text{в противен случай.} \end{cases}$$

Покажете, че съвместната плътност на случайните величини ξ^2 и η^2 е

$$f_{\xi^2,\eta^2}(u,v) = \begin{cases} 1 & \text{за } 0 < u < 1,\ 0 < v < 1, \\ 0 & \text{в противен случай.} \end{cases}$$

3. Нека $f_{\xi,\eta}(x,y)$ е съвместната плътност на ξ и η. Намерете плътността на случайната величина $\zeta = \xi - \eta$.

4. Нека
$$f_{\xi,\eta}(x,y) = \begin{cases} 3x & \text{за } 0 < x < 1,\ 0 < y < 1, \\ 0 & \text{в противен случай.} \end{cases}$$
Покажете, че плътността на $\zeta = \xi - \eta$ е
$$f_\zeta(z) = \begin{cases} \dfrac{3(1-z^2)}{2} & \text{за } 0 < z < 1, \\ 0 & \text{в противен случай.} \end{cases}$$

5. Случайните величини ξ, η и ζ имат съвместна плътност
$$f_{\xi,\eta,\zeta}(x,y,z) = \begin{cases} \exp\{-(x+y+z)\} & \text{за } x > 0,\ y > 0,\ z > 0, \\ 0 & \text{в противен случай.} \end{cases}$$
Покажете, че случайната величина $u = \dfrac{1}{3}(\xi + \eta + \zeta)$ има плътност
$$f_u(u) = \begin{cases} \dfrac{27}{2}u^2 \exp(-3u) & \text{за } u > 0, \\ 0 & \text{в противен случай.} \end{cases}$$

6. Нека $\xi \in Bi(n,p)$ и $\eta \in Bi(m,p)$ са независими случайни величини. Покажете, че $\xi + \eta = \zeta \in Bi(n+m,p)$.

7. Нека $\xi \in Po(\lambda_1)$ и $\eta \in Po(\lambda_2)$ са независими случайни величини. Покажете, че $\xi + \eta = \zeta \in Po(\lambda_1 + \lambda_2)$.

8. Нека ξ и η са независими и еднакво разпределени случайни величини с плътност
$$f(x) = \begin{cases} \exp\left(-\dfrac{x}{\theta}\right) & \text{за } x \geqq 0, \\ 0 & \text{в противен случай.} \end{cases}$$
Докажете, че плътността на случайната величина $\zeta = \xi + \eta$ има следния вид:
$$f_\zeta(z) = \begin{cases} z \exp\left(-\dfrac{z}{\theta}\right) & \text{за } z \geqq 0, \\ 0 & \text{в противен случай.} \end{cases}$$

9. При хвърляне на зар са се паднали ν точки. След това зарът се хвърля ν пъти. Да означим с η сумата от точките при тези ν хвърляния. Покажете, че $\mathbf{E}\eta = \dfrac{49}{4}$.

10. Докажете, че
$$\mathbf{D}\xi = \mathbf{E}\{\mathbf{D}(\xi \,|\, \eta)\} + \mathbf{D}\{\mathbf{E}(\xi \,|\, \eta)\}.$$

11. Случайните величини ξ_1, ξ_2 и ξ_3 са независими и $\xi_i \in \mathcal{N}(0,1)$, $i = 1,2,3$. Покажете, че случайната величина $\eta = \dfrac{\xi_1 + \xi_2\xi_3}{\sqrt{1+\xi_3^2}}$ има плътност и $\eta \in \mathcal{N}(0,1)$.

Упътване. Намерете условната плътност на η при условие, че е фиксирана ξ_3.

Шеста глава

Аналитичен апарат на теория на вероятностите

§ 1. Пораждащи функции

Наред с реални случайни величини можем да разглеждаме и комплексни случайни величини $\zeta(\omega) = \xi(\omega) + i\eta(\omega)$, където ξ и η са реални случайни величини. Всичко доказано за реални случайни величини по очевиден начин се пренася за комплексни случайни величини, например $\mathbf{E}\zeta = \mathbf{E}\xi + i\mathbf{E}\eta$. Случайните величини $\zeta_1 = \xi_1 + i\eta_1$ и $\zeta_2 = \xi_2 + i\eta_2$ са независими, ако са независими векторите (ξ_1, η_1) и (ξ_2, η_2). При това директно се проверява, че $\mathbf{E}\zeta_1\zeta_2 = \mathbf{E}\zeta_1\mathbf{E}\zeta_2$, и т. н.

В този параграф ще изучаваме дискретни случайни величини, приемащи цели неотрицателни стойности, т. е.

$$p_k = \mathbf{P}\{\xi = k\}, \qquad \sum_{k=0}^{\infty} p_k = 1. \tag{1}$$

Целочислените разпределения (1) лесно се изучават с помощта на *пораждащи* (*производящи*) *функции*, които се дефинират по следния начин:

$$h_\xi(s) = \mathbf{E}s^\xi = \sum_{n=0}^{\infty} p_n s^n. \tag{2}$$

Редът (2) е абсолютно сходящ при $|s| \leqq 1$, тъй като

$$|h_\xi(s)| \leqq \sum_{n=0}^{\infty} p_n |s|^n \leqq \sum_{n=0}^{\infty} p_n = 1 = h_\xi(1), \tag{3}$$

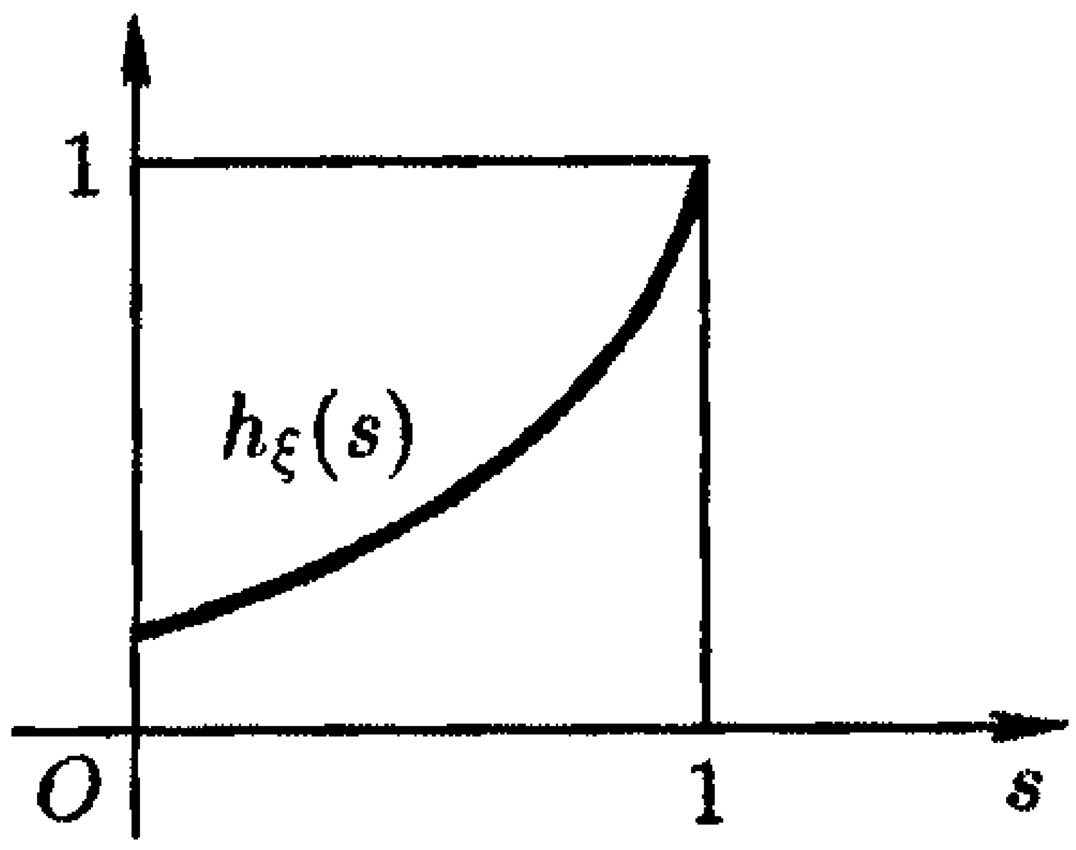
т. е. $h_\xi(s)$ е аналитична функция в единичния кръг. Тъй като

$$p_n = \frac{h_\xi^{(n)}(0)}{n!}, \quad n = 0, 1, 2, \dots, \tag{4}$$

между разпределенията (1) и пораждащите функции (2) съществува взаимноеднозначно съответствие.

В интервала $[0,1]$ пораждащата функция $h_\xi(s)$ и всичките ѝ производни $h_\xi^{(k)}(s)$, $k = 1, 2, \ldots$, са неотрицателни, ненамаляващи и изпъкнали функции (фиг. 11), тъй като при $0 \leqq s \leqq 1$ е изпълнено

$$(5) \qquad h^{(k)}(s) = \sum_{n=k}^{\infty} n(n-1)\cdots(n-k+1)p_n s^{n-k} \geqq 0.$$



Фиг. 11

Ако при $s > 1$ пораждащата функция $h_\xi(s)$ не е дефинирана, производните ѝ при $s = 1$ се определят като леви производни, т. е.

$$h_\xi^{(k)}(1) = \lim_{s\uparrow 1} \frac{h_\xi^{(k-1)}(1) - h_\xi^{(k-1)}(s)}{1-s}, \qquad k = 1, 2, \ldots$$

При това, ако

$$(6) \qquad m_k = \sum_{n=k}^{\infty} n(n-1)\cdots(n-k+1)p_n < \infty,$$

то по теоремата на Абел при $s \uparrow 1$ имаме $\lim h_\xi^{(k)}(s) = h_\xi^{(k)}(1) = m_k$, където съгласно (6) е изпълнено $m_k = \mathbf{E}\{\xi(\xi-1)\ldots(\xi-k+1)\}$ и m_k се нарича *k-ти факториален момент.*

Тъй като от $\mathbf{E}\xi^k < \infty$, съгласно (4.5.9), следва $\mathbf{E}\xi^m < \infty$ за всички $m < k$, то факториалният момент m_k е също краен, защото се изразява линейно чрез моментите $\mathbf{E}\xi^m$, $m \leqq k$. Специално

$$(7) \qquad \mathbf{E}\xi = m_1 = h'_\xi(1),$$

$$(8) \qquad \mathbf{D}\xi = m_2 + m_1 - m_1^2 = h_\xi''(1) + h_\xi'(1) - [h_\xi'(1)]^2.$$

Пример 1. Нека $\xi \in Bi(n,p)$. Тогава

$$h_\xi(s) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} s^k = (ps+q)^n,$$

$$h_\xi'(s) = np(ps+q)^{n-1}, \quad h_\xi''(s) = n(n-1)p^2(ps+q)^{n-2},$$
$$m_1 = h_\xi'(1) = np = \mathbf{E}\xi, \quad m_2 = h_\xi''(1) = n(n-1)p^2 = \mathbf{E}\xi(\xi-1),$$
$$\mathbf{D}\xi = n(n-1)p^2 + np - (np)^2 = npq.$$

Пример 2. Нека $\eta \in Po(\lambda)$. Тогава

$$h_\eta(s) = \sum_{n=0}^{\infty} \frac{e^{-\lambda}\lambda^n}{n!} s^n = e^{\lambda(s-1)},$$

$$h_\eta'(s) = \lambda e^{\lambda(s-1)}, \quad h_\eta''(s) = \lambda^2 e^{\lambda(s-1)},$$
$$\mathbf{E}\xi = \lambda, \quad \mathbf{D}\xi = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Пример 3. Нека $\zeta \in G(p)$. Тогава

$$h_\zeta(s) = \sum_{n=0}^{\infty} pq^n s^n = \frac{p}{1-qs},$$

$$h_\zeta'(s) = \frac{pq}{(1-qs)^2}, \quad h_\zeta''(s) = \frac{2pq^2}{(1-qs)^3},$$
$$\mathbf{E}\zeta = \frac{q}{p}, \quad \mathbf{D}\zeta = \frac{2q^2}{p^2} + \frac{q}{p} - \left(\frac{q}{p}\right)^2 = \frac{q}{p^2}.$$

Теорема 1 (мултипликативно свойство). *Ако $\xi_1, \ldots, \xi_n$ са независими случайни величини с пораждащи функции $h_{\xi_k}(s)$, където $k = 1, 2, \ldots, n$, то за пораждащата функция на тяхната сума $S_n = \xi_1 + \cdots + \xi_n$ имаме*

$$(9) \qquad h_{S_n}(s) = \prod_{k=1}^{n} h_{\xi_k}(s).$$

Доказателство. Тъй като $\xi_1, \ldots, \xi_n$ са независими случайни величини, то по следствие 2.6.1 случайните величини $s^{\xi_1}, s^{\xi_2}, \ldots, s^{\xi_n}$

са също независими. Тогава от мултипликативното свойство на математическото очакване (вж. (4.4.2)) получаваме

$$\mathbf{E}s^{S_n} = \mathbf{E}s^{\xi_1+\xi_2+\cdots+\xi_n} = \mathbf{E}s^{\xi_1}\dots s^{\xi_n} = \prod_{k=1}^{n} \mathbf{E}s^{\xi_k},$$

което е еквивалентно на (9).

Ако ξ и η са независими случайни величини с разпределения $p_k = \mathbf{P}\{\xi = k\}$ и $q_k = \mathbf{P}\{\eta = k\}$, разпределението на тяхната сума се дефинира по формулата за пълната вероятност

$$(10) \qquad r_n = \mathbf{P}\{\xi + \eta = n\} = \sum_{k=0}^{n} \mathbf{P}\{\xi = k\}\mathbf{P}\{\eta = n-k\} = \sum_{k=0}^{n} p_k q_{n-k}$$

и, както знаем от § 5.1, се нарича композиция на $\{p_k\}$ и $\{q_k\}$ (вж. (5.1.12)).

Сега ще покажем как теорема 1 позволява да се намери композицията на разпределения, без да се прибягва до сложни формули от типа на (10).

Пример 4. От равенствата

$$(ps+q)^n(ps+q)^m = (ps+q)^{n+m}$$

следва, че композицията на две разпределения $Bi(n,p)$ и $Bi(m,p)$ е пак биномно разпределена — $Bi(n+m,p)$.

Пример 5. От

$$e^{\lambda(s-1)}e^{\mu(s-1)} = e^{(\lambda+\mu)(s-1)}$$

следва, че композицията на две поасонови разпределения $Po(\lambda)$ и $Po(\mu)$ е пак поасоново разпределена — $Po(\lambda+\mu)$.

Пример 6. В схемата на Бернули да означим с ζ_r броя на опитите до r-тия успех. Тогава, както знаем от § 3.2, случайната величина ζ_1 е геометрично разпределена, т. е. $\zeta_1 \in G(p)$. Не е трудно да се съобрази, че ζ_r може да се представи като сума

$$(11) \qquad \zeta_r = r - 1 + \xi_1 + \cdots + \xi_r,$$

където $\{\xi_r\}$ са независими и еднакво разпределени (като ζ от пример 3) случайни величини (ξ_k се интерпретира като брой на опитите между $(k-1)$-вия и k-тия успех). Тогава от пример 3 и от равенствата (9) и (11) следва

$$(12) \qquad h_{\zeta_r}(s) = \frac{s^{r-1}p^r}{(1-qs)^r}.$$

Като разложим (12) в степенен ред, получаваме

$$h_{\zeta_r}(s) = p^r s^{r-1} \sum_{k=0}^{\infty} \binom{-r}{k} (-1)^k q^k s^k. \tag{13}$$

Тъй като

$$\binom{-r}{k}(-1)^k = \frac{-r(-r-1)\dots(-r+k-1)}{k!}(-1)^k$$
$$= \frac{(r+k-1)(r+k-2)\dots r}{k!} = C_{r+k-1}^k,$$

от (13) следва

$$h_{\zeta_r}(s) = p^r \sum_{k=0}^{\infty} C_{r+k-1}^k q^k s^{k+r-1} = p^r \sum_{j=r-1}^{\infty} C_j^{r-1} q^{j-r+1} s^j,$$

т. е.

$$\mathbf{P}\{\zeta_r = j\} = C_j^{r-1} p^r q^{j-r+1}, \quad j = r-1,\, r,\, r+1,\, \dots \tag{14}$$

Понякога вместо (11) се разглежда случайната величина $\eta_r = \zeta_r - r + 1$, която очевидно е броят на неуспехите, предшестващи r-тия успех. Тогава от (14) следва

$$\mathbf{P}\{\zeta_r = k\} = C_{r+k-1}^k p^r q^k, \quad k = 0,\, 1,\, 2,\, \dots \tag{15}$$

За η_r с (15) ще казваме, че има отрицателно биномно разпределение, което ще бележим $\eta_r \in NB(r, p)$.

Теорема 2. *Нека $\{\xi_k\}$ са независими и еднакво разпределени случайни величини с пораждаща функция $h_\xi(s)$ и ν е независима от тях целочислена случайна величина с пораждаща функция $h_\nu(s)$. Тогава пораждащата функция на сумата $\zeta_\nu = \xi_1 + \dots + \xi_\nu$ от случаен брой случайни величини се определя от суперпозицията*

$$h_{\zeta_\nu}(s) = h_\nu(h_\xi(s)). \tag{16}$$

Доказателство. За да пресметнем $h_{\zeta_\nu}(s) = \mathbf{E}s^{\zeta_\nu}$, ще използваме теорема 5.3.1. Тъй като по теорема 1

$$\mathbf{E}(s^{\zeta_\nu} \mid \nu = n) = \mathbf{E}s^{\xi_1 + \dots + \xi_n} = [h_\xi(s)]^n,$$

от (5.3.5) и (5.3.6) намираме

$$\mathbf{E}s^{\zeta_\nu} = \mathbf{E}\{\mathbf{E}(s^{\zeta_\nu} \mid \nu)\} = \mathbf{E}[h_\xi(s)]^\nu = h_\nu(h_\xi(s)),$$

което доказва теоремата.

Сега ще покажем, че съответствието между разпределенията (1) и пораждащите функции (2) е не само взаимноеднозначно, но и взаимнонепрекъснато. Това твърдение се съдържа в следващата теорема за непрекъснатост на съответствието между множеството на пораждащите функции и множеството на целочислените разпределения.

Теорема 3 (теорема за непрекъснатост). *Нека $\xi_1, \xi_2, \ldots$ е редица от целочислени случайни величини с разпределения*

$$p_k(n) = \mathbf{P}\{\xi_n = k\}, \quad \sum_{k=0}^{\infty} p_k(n) = 1,$$

и съответни пораждащи функции

$$h_n(s) = \mathbf{E}s^{\xi_n} = \sum_{k=0}^{\infty} p_k(n)s^k, \quad n = 1, 2, \ldots$$

За да бъде изпълнено при всяко $k = 0, 1, \ldots$

(17) $$\lim_{n\to\infty} p_k(n) = p_k,$$

необходимо и достатъчно е при всички $0 \leqq s < 1$ да имаме

(18) $$\lim_{n\to\infty} h_n(s) = h(s).$$

Доказателство. *Необходимост.* Нека е изпълнено (17). За всяко $\varepsilon > 0$ и $|s| < 1$ съществува $N = N(\varepsilon, s)$, така че $\dfrac{|s|^N}{1-|s|} < \dfrac{\varepsilon}{2}$. От (17) следва, че при така избраното N съществува n_0, такова че при $n \geqq n_0$ имаме $|p_k(n) - p_k| < \dfrac{\varepsilon}{2N}$ при $k = 0, 1, \ldots, N-1$. Тогава при $|s| < 1$ от неравенствата

$$|h_n(s) - h(s)| \leqq \sum_{k=0}^{\infty} |p_k(n) - p_k||s|^k$$

$$\leqq \sum_{k=0}^{N-1} |p_k(n) - p_k| + \sum_{k=N}^{\infty} |s|^k = \sum_{k=0}^{N-1} |p_k(n) - p_k| + \frac{|s|^N}{1-|s|}$$

получаваме $|h_n(s) - h(s)| < \varepsilon$, което доказва (18) за $|s| < 1$.

Достатъчност. Нека е в сила (18). Да допуснем, че (17) не е вярно. Тъй като $0 \leqq p_k(n) \leqq 1$ за всички k и n, то могат да се намерят две подредици $\{n_i\}$ и $\{m_r\}$, такива че $p_k(n_i) \underset{n_i \to \infty}{\longrightarrow} \widetilde{p}_k$ и $p_k(m_r) \underset{m_r \to \infty}{\longrightarrow} \widehat{p}_k$ за всяко $k = 0, 1, 2, \ldots$, и поне за едно k да имаме $\widetilde{p}_k \not\equiv \widehat{p}_k$, т. е. $\{\widetilde{p}_k\}$ и $\{\widehat{p}_k\}$ да не съвпадат. Това може да се докаже по диагоналния принцип на Кантор. От $0 \leqq p_k(n) \leqq 1$ следва, че можем да изберем такава сходяща подредица $\{n_0\}$, че $p_0(n_0) \to \widetilde{p}_0$ при $n_0 \to \infty$. Разглеждаме редицата $\{p_1(n_0)\}$ и избираме подредица $\{n_1\}$, за която $p_1(n_1) \to \widetilde{p}_1$ при $n_1 \to \infty$. Но тогава и $p_0(n_1) \to \widetilde{p}_0$ при $n_1 \to \infty$. Аналогично избираме подредица $\{n_2\}$ от $\{n_1\}$, такава че $p_2(n_2) \to \widetilde{p}_2$ при $n_2 \to \infty$, и т. н. По този начин се показва, че за всяко i съществува подредица $\{n_i\}$, такава че $p_k(n_i) \to \widetilde{p}_k$ при $n_i \to \infty$ за всички $k \leqq i$. Аналогично се разсъждава за подредицата $\{m_r\}$. Следователно от доказаната необходимост ще имаме

$$\lim_{n_i \to \infty} h_{n_i}(s) = \widetilde{h}(s) = \sum_{k=0}^{\infty} \widetilde{p}_k s^k,$$

$$\lim_{m_r \to \infty} h_{m_r}(s) = \widehat{h}(s) = \sum_{k=0}^{\infty} \widehat{p}_k s^k,$$

за $|s| < 1$, като $\widetilde{h}(s) \not\equiv \widehat{h}(s)$ при $|s| < 1$. Но от (18) имаме

$$\lim_{n \to \infty} h_n(s) = h(s)$$

при $0 \leqq s < 1$ и следователно $h(s) = \widetilde{h}(s) = \widehat{h}(s)$. Тъй като функциите $\widetilde{h}(s)$, $\widehat{h}(s)$ са аналитични, оттук следва, че $\widetilde{h}(s) \equiv \widehat{h}(s)$ при $|s| < 1$. Полученото противоречие показва, че нашето допускане не е вярно, което доказва (17).

З а б е л е ж к а 1. Всъщност доказахме по-силен вариант на теорема 1, а именно, че от (17) следва (18) за $|s| < 1$, а (17) следва от (18) при $s \in U$, където U е подмножество на единичния кръг с точка на сгъстяване. Във формулировката на теорема 1 фигурира $U = [0, 1)$.

З а б е л е ж к а 2. Тъй като $0 \leqq p_k(n) \leqq 1$, за техните граници имаме $0 \leqq p_k \leqq 1$. От друга страна, за всяко $N < \infty$ при $n \to \infty$ е в сила

$$1 \geqq \sum_{k=0}^{N} p_k(n) \to \sum_{k=0}^{N} p_k$$

и следователно $\sum\limits_{k=0}^{\infty} p_k \leqq 1$. За да бъде изпълнено $\sum\limits_{k=0}^{\infty} p_k = 1$, трябва да се поиска допълнителното условие $h(1) = 1$, т. е. (18) да бъде изпълнено и при $s = 1$.

З а б е л е ж к а 3. Тъй като $F_n(x) = \mathbf{P}\{\xi_n < x\} = \sum\limits_{k<x} p_k(n)$ и $F(x) = \sum\limits_{k<x} p_k = \mathbf{P}\{\xi < x\}$, то (17) е еквивалентно на

$$\lim_{n\to\infty} F_n(x) = F(x) \tag{19}$$

във всички точки на непрекъснатост на $F(x)$. В този случай казваме, че ξ_n клони по разпределение към ξ, и означаваме $\xi_n \xrightarrow{d} \xi$. По-подробно с тази и други видове сходимости ще се запознаем в седма глава.

Нека случайният вектор $\eta = (\xi_1, \ldots, \xi_k)$ с целочислени неотрицателни компоненти ξ_i, $i = 1, \ldots, k$, има разпределение $p_\alpha = \mathbf{P}\{\eta = \alpha\}$, където $\alpha = (\alpha_1, \ldots, \alpha_k)$ са възможните стойности на вектора η. *Многомерна пораждаща функция* се дефинира по следния начин:

$$h_\eta(s_1, \ldots, s_k) = \mathbf{E}\{s_1^{\xi_1} \ldots s_k^{\xi_k}\} = \sum_\alpha p_\alpha s_1^{\alpha_1} \ldots s_k^{\alpha_k}. \tag{20}$$

Многомерните пораждащи функции имат свойства, аналогични на едномерните пораждащи функции. От тях лесно се пресмятат смесените факториални моменти. Например

$$\mathbf{E}\{\xi_1 \ldots \xi_k\} = \left.\frac{\partial^k h_\eta(s_1, \ldots, s_k)}{\partial s_1 \ldots \partial s_k}\right|_{s_1=\cdots=s_k=1}.$$

Пример 7. Полиномното разпределение

$$\mathbf{P}\{\zeta = \alpha\} = \frac{n!}{\alpha_1! \ldots \alpha_k!} p_1^{\alpha_1} \ldots p_k^{\alpha_k}, \quad \sum_{i=1}^{k} \alpha_i = n, \quad \sum_{i=1}^{k} p_i = 1$$

има пораждаща функция $h_\zeta(s_1, \ldots, s_k) = (p_1 s_1 + \cdots + p_k s_k)^n$.

§ 2. Трансформации на Лаплас – Стилтес

Като аналог на пораждащите функции за произволни неотрицателни случайни величини служат трансформациите на Лаплас - Стилтес:

$$\varphi_\xi(\lambda) = \mathbf{E}e^{-\lambda\xi} = \int\limits_0^\infty e^{-\lambda x} dF_\xi(x),$$

дефинирани при $\mathrm{Re}\lambda \geqq 0$ (в частност при $\lambda \geqq 0$).

Лема 1. *Нека* $\zeta = \xi + i\eta$, *където* $\mathbf{E}|\xi| < \infty$ и $\mathbf{E}|\eta| < \infty$. *Тогава*

$$|\mathbf{E}\zeta| \leqq \mathbf{E}|\zeta|. \tag{1}$$

Доказателство. Ще разгледаме първо случая, където ζ взема краен брой стойности $z_k = x_k + iy_k$ с вероятности $p_k = \mathbf{E}\{\zeta = z_k\}$, $k = 1, \ldots, n$. Тогава

$$|\mathbf{E}\zeta| = |\sum_{k=1}^n (x_k + iy_k)p_k| \leqq \sum_{k=1}^n |z_k| p_k = \mathbf{E}|\zeta|. \tag{2}$$

В общия случай $\xi = \xi^+ - \xi^-$, $\eta = \eta^+ - \eta^-$ и съществуват елементарни случайни величини $\xi_n^\pm \uparrow \xi^\pm$, $\eta_n^\pm \uparrow \eta^\pm$, такива че $\xi_n = \xi_n^+ - \xi_n^- \to \xi$, $\eta_n = \eta_n^+ - \eta_n^- \to \eta$. Следователно $\xi_n + i\eta_n = \zeta_n \to \zeta$ и $\mathbf{E}\zeta_n \to \mathbf{E}\zeta$.

От друга страна,

$$|\zeta_n| \leqq |\xi_n| + |\eta_n| = \xi_n^+ + \xi_n^- + \eta_n^+ + \eta_n^- \leqq \xi^+ + \xi^- + \eta^+ + \eta^- = |\xi| + |\eta|.$$

Оттук по теоремата на Фату – Лебег получаваме $\mathbf{E}|\zeta_n| \to \mathbf{E}|\zeta|$. Тъй като от (2) имаме $|\mathbf{E}\zeta_n| \leqq \mathbf{E}|\zeta_n|$, като извършим граничен преход при $n \to \infty$, получаваме (1).

От лема 1 следва, че при $\mathrm{Re}\lambda \geqq 0$

$$|\varphi_\xi(\lambda)| = |\mathbf{E}e^{-\lambda\xi}| \leqq \mathbf{E}|e^{-\lambda\xi}| \leqq 1,$$

като $\varphi_\xi(0) = 1$.

Ако ξ е целочислена случайна величина, т. е.

$$\mathbf{P}\{\xi = k\} = p_k, \ \sum_{k=0}^\infty p_k = 1,$$

то

$$\varphi_\xi(\lambda) = \mathbf{E}e^{-\lambda\xi} = \sum_{k=0}^\infty p_k e^{-\lambda k} = h_\xi(e^{-\lambda}),$$

т. е. трансформацията на Лаплас – Стилтес се получава от пораждащата функция чрез смяната $s = e^{-\lambda}$. С това се обяснява и голямото сходство между техните свойства.

Например, ако ξ и η са независими неотрицателни случайни величини, то

$$(3) \qquad \begin{aligned} \varphi_{\xi+\eta}(\lambda) &= \mathbf{E}e^{-\lambda(\xi+\eta)} = \mathbf{E}(e^{-\lambda\xi} \cdot e^{-\lambda\eta}) \\ &= \mathbf{E}e^{-\lambda\xi}\mathbf{E}e^{-\lambda\eta} = \varphi_\xi(\lambda)\varphi_\eta(\lambda), \end{aligned}$$

което е аналог на теорема 1.1.

От (5.1.10) знаем, че в този случай

$$F_{\xi+\eta}(x) = \int_0^x F_\xi(x-t)dF_\eta(t).$$

Мултипликативното свойство (3) дава възможност да се намират композиции на разпределения, без да се използва горната сложна формула.

Пример 1. Нека $\xi \in Ex(\theta)$, т. е. $F_\xi(x) = 1 - e^{-x/\theta}$, $x \geqq 0$, $\theta > 0$. Тогава при $\lambda \geqq 0$

$$(4) \qquad \varphi_\xi(\lambda) = \int_0^\infty e^{-\lambda x}dF_\xi(x) = \frac{1}{\theta}\int_0^\infty e^{-\lambda x - x/\theta}dx = \frac{1}{1+\theta\lambda}.$$

От дефиницията на $\varphi_\xi(\lambda)$, като диференцираме формално под знака на интеграла при $\lambda > 0$, получаваме

$$(5) \qquad (-1)^n\varphi_\xi^{(n)}(\lambda) = \mathbf{E}(\xi^n e^{-\lambda\xi}) = \int_0^\infty e^{-\lambda x}x^n dF_\xi(x).$$

Всъщност равенството в (5) е напълно законно, тъй като $\zeta = \xi^n e^{-\lambda\xi}$ е ограничена случайна величина. По теоремата на Фату – Лебег следва, че $\mathbf{E}\xi^n$ съществува тогава и само тогава, когато съществува $\varphi_\xi^{(n)}(0) = \lim \varphi_\xi^{(n)}(\lambda)$ при $\lambda \downarrow 0$. В този случай, ако $\mathbf{E}\xi^n < \infty$, то

$$(6) \qquad \mathbf{E}\xi^n = (-1)^n\varphi_\xi^{(n)}(0).$$

В частност $\mathbf{E}\xi = -\varphi'_\xi(0)$, $\mathbf{E}\xi^2 = \varphi''_\xi(0)$.

Следващата теорема показва, че аналогично на пораждащите функции трансформацията (1) възстановява напълно функцията на разпределение.

Теорема 1. *На различни неотрицателни разпределения съответстват различни трансформации на Лаплас – Стилтес. При това във всички точки на непрекъснатост на функцията на разпределение $F_\xi(x)$ е вярна следната формула за обръщане:*

$$F_\xi(x) = \lim_{\lambda\to\infty} \sum_{k<\lambda x} \frac{(-\lambda)^k}{k!}\varphi_\xi^{(k)}(\lambda). \tag{7}$$

Следващата теорема е аналог на теоремата за непрекъснатост на пораждащите функции (теорема 1.3).

Теорема 2 (теорема за непрекъснатост). *Нека*

$$\varphi_n(\lambda) = \int_0^\infty e^{-\lambda x} dF_n(x), \quad \varphi(\lambda) = \int_0^\infty e^{-\lambda x} dF(x)$$

са трансформациите на Лаплас – Стилтес за неотрицателните случайни величини $\{\xi_n\}$ и ξ, т. е.

$$F_n(x) = \mathbf{P}\{\xi_n < x\},\ n = 1, 2, \ldots, \quad F(x) = \mathbf{P}\{\xi < x\}.$$

Тогава при $\lambda > 0$ имаме $\lim\limits_{n\to\infty} \varphi_n(\lambda) = \varphi(\lambda)$ тогава и само тогава, когато $\lim\limits_{n\to\infty} F_n(x) = F(x)$ в точките на непрекъснатост на $F(x)$.

Пример 2. Нека $\xi_1, \ldots, \xi_n$ са независими експоненциално разпределени случайни величини, т. е. $\xi_k \in Ex(\theta)$, $k = 1, \ldots, n$. Нека $\varphi_n(\lambda) = \mathbf{E}e^{-\lambda\zeta_n}$, където $\zeta_n = \xi_1 + \cdots + \xi_n$. Съгласно (3) и (4) при $\lambda > 0$ е изпълнено

$$\varphi_n(\lambda) = (1+\theta\lambda)^{-n}. \tag{8}$$

От друга страна, непосредствено се проверява, че

$$f(x) = \frac{x^{n-1}e^{-x/\theta}}{\Gamma(n)\theta^n}, \quad x \geqq 0, \tag{9}$$

е плътност на разпределение и при $\lambda > 0$

$$\int_0^\infty e^{-\lambda x} f(x)dx = (1+\theta\lambda)^{-n}. \tag{10}$$

Съотношенията (8) – (10) показват, че случайната величина ζ_n има плътност $f(x)$. За ζ_n се казва, че има разпределение на Ерланг, което от своя страна е частен случай на Г-разпределение. Означава се $\zeta_n \in \Gamma(n,\theta)$.

§ 3. Характеристични функции

Аналог на пораждащите функции и на трансформациите на Лаплас – Стилтес за произволни случайни величини ξ са т. нар. характеристични функции

$$(1)\qquad \psi_\xi(t) = \mathbf{E}e^{it\xi} = \int_{-\infty}^{\infty} e^{itx}dF_\xi(x),$$

дефинирани за $t \in \mathbf{R}^1$, където $i = \sqrt{-1}$. Понякога (1) се нарича трансформация на Фурие – Стилтес. Очевидно характеристичните функции са дефинирани за всяко реално t, тъй като подинтегралната случайна величина

$$(2)\qquad e^{it\xi} = \cos t\xi + i\sin t\xi$$

е ограничена. В частност от (2) следва представянето

$$(3)\qquad \psi_\xi(t) = \int_{-\infty}^{\infty} \cos tx dF_\xi(x) + i\int_{-\infty}^{\infty} \sin tx dF_\xi(x) = \mathbf{E}\cos t\xi + i\mathbf{E}\sin t\xi.$$

От (1) и (3) се вижда, че характеристичната функция е комплексна функция от реален аргумент, при това реалната ѝ част е четна, а имагинерната — нечетна функция.

Съгласно лема 2.1 от (1) следва, че

$$(4)\qquad |\psi_\xi(t)| \leqq \mathbf{E}|e^{it\xi}| = 1 \quad \text{и} \quad \psi_\xi(0) = 1.$$

Да отбележим, че ако ξ е целочислена случайна величина, т. е. $p_k = \mathbf{P}\{\xi = k\}$, $\sum\limits_{k=0}^{\infty} p_k = 1$, то

$$\psi_\xi(t) = \mathbf{E}e^{it\xi} = \sum_{k=0}^{\infty} p_k e^{itk} = h_\xi(e^{it}),$$

т. е. характеристичната функция получаваме от пораждащата, като положим $s = e^{it}$, което обяснява сходството между двете функции.

Аналогично, ако $\xi \geqq 0$ и $\varphi_\xi(\lambda) = \mathbf{E}e^{-\lambda\xi}$ е трансформацията на Лаплас – Стилтес, то

$$\psi_\xi(t) = \int_0^\infty e^{itx} dF_\xi(x) = \varphi_\xi(-it),$$

т. е. характеристичната функция се получава от трансформацията на Лаплас – Стилтес чрез полагането $\lambda = -it$.

Ако случайната величина ξ има плътност $\varphi_\xi(x)$, от (1) следва

$$\psi_\xi(t) = \int_{-\infty}^\infty e^{itx} f_\xi(x) dx, \tag{5}$$

т. е. в този случай характеристичната функция е добре познатата от анализа трансформация на Фурие.

Аналогично на (1.9) и (2.3) е в сила мултипликативното свойство при независими случайни величини $\xi_1, \ldots, \xi_n$:

$$\begin{aligned} \psi_{\xi_1+\cdots+\xi_n}(t) &= \mathbf{E}e^{it(\xi_1+\cdots+\xi_n)} \\ &= \mathbf{E}\prod_{k=1}^n e^{it\xi_k} = \prod_{k=1}^n \mathbf{E}e^{it\xi_k} = \prod_{k=1}^n \psi_{\xi_k}(t). \end{aligned} \tag{6}$$

Следователно на композицията от функции на разпределение (вж. § 5.1) се съпоставя произведение на съответните характеристични функции. Тъй като съответствието между функциите на разпределение и характеристичните функции е взаимноеднозначно (това ще бъде доказано в следващата глава), то (6) позволява да се избягват сложните формули за композиция като (5.1.9).

В частност, ако a и b са константи и $\eta = a\xi + b$, то

$$\psi_\eta(t) = \mathbf{E}e^{it(a\xi+b)} = e^{itb}\psi_\xi(at). \tag{7}$$

По-нататък ще използваме следното помощно твърдение:

Лема 1. *За всяко $t \in \mathbf{R}^1$ и за всяко цяло $n \geqq 1$ е в сила неравенството*

$$\left| e^{it} - \sum_{k=0}^{n-1} \frac{(it)^k}{k!} \right| \leqq \frac{|t|^n}{n!}. \tag{8}$$

Доказателство. При $n = 1$ получаваме директно

$$|e^{it} - 1| = \left| i \int_0^t e^{ix} dx \right| \leqq \int_0^{|t|} |e^{ix}| dx = |t|.$$

Да допуснем, че неравенството (8) е в сила за някакво $n \geqq 1$. Тогава

$$\left| e^{it} - \sum_{k=0}^{n} \frac{(it)^k}{k!} \right| = \left| i \int_0^t \left(e^{ix} - \sum_{k=0}^{n-1} \frac{(ix)^k}{k!} \right) dx \right| \leqq \int_0^{|t|} \frac{x^n}{n!} dx = \frac{|t|^{n+1}}{(n+1)!},$$

което доказва (8) по индукция.

Теорема 1. *Всяка характеристична функция $\psi_\xi(t)$, дефинирана с (1), е равномерно непрекъсната по $t \in \mathbf{R}^1$.*

Доказателство. Тъй като $\mathbf{P}\{|\xi| \leqq x\} = F(x) - F(-x) \to 1$ при $x \to \infty$, то за всяко $\varepsilon > 0$ съществува x_0, такова че

$$\mathbf{P}(A) = \mathbf{P}\{|\xi| \leqq x_0\} \geqq 1 - \frac{\varepsilon}{4}. \tag{9}$$

Тогава, като използваме (8), при $n = 1$ получаваме

$$\begin{aligned} |\psi_\xi(t+h) - \psi_\xi(t)| &= |\mathbf{E}e^{it\xi}(e^{ih\xi} - 1)| \\ &\leqq \mathbf{E}|e^{ih\xi} - 1| I_A + \mathbf{E}|e^{ih\xi} - 1| I_{\overline{A}} \\ &\leqq |h| \mathbf{E}|\xi| I_A + 2\mathbf{E}I_{\overline{A}} \leqq |h| x_0 + 2\mathbf{P}(\overline{A}). \end{aligned}$$

Оттук, като положим $\delta = \dfrac{\varepsilon}{2x_0}$ и използваме (9), следва, че $|\psi_\xi(t+h) - \psi_\xi(t)| < \varepsilon$ при $|h| < \delta$, което доказва теоремата.

Теорема 2. *Ако $\mathbf{E}|\xi^n| < \infty$, то съществуват всички производни при $k \leqq n$, като*

$$\psi_\xi^{(k)}(0) = i^k \mathbf{E}\xi^k \tag{10}$$

и е в сила представянето

$$\psi_\xi(t) = \sum_{k=0}^{n} \frac{(it)^k}{k!} \mathbf{E}\xi^k + R_n(t), \tag{11}$$

където $R_n(t) = o(t^n)$ при $t \to \infty$.

Доказателство. От неравенството $|x|^k \leqq |x|^n + 1$ при $k \leqq n$ следва $\mathbf{E}|\xi|^k \leqq \mathbf{E}|\xi|^n + 1 < \infty$.

От друга страна, като диференцираме формално (1), получаваме

$$\psi_\xi^{(k)}(t) = i^k \mathbf{E}\xi^k e^{it\xi} = i^k \int_{-\infty}^{\infty} x^k e^{itx} dF_\xi(x), \tag{12}$$

откъдето при $t = 0$ намираме (10).

За да обосновем (12), ще разсъждаваме по индукция. Нека (12) е вярно при $k < n$. Тогава

$$\frac{\psi_\xi^{(k)}(t+h) - \psi_\xi^{(k)}(t)}{h} = i^k \mathbf{E}\xi^k \frac{e^{it\xi}(e^{ih\xi} - 1)}{h}. \tag{13}$$

Тъй като от (8) следват

$$\left| \xi^k e^{it\xi} \frac{e^{ih\xi} - 1}{h} \right| \leqq |\xi|^{k+1} \quad \text{и} \quad \mathbf{E}|\xi|^{k+1} < \infty,$$

по теоремата на Фату – Лебег можем да извършим граничен преход под знака на математическото очакване в (13) при $h \to 0$. Следователно (12) е вярно при $k + 1$.

Като имаме предвид (10), всъщност (11) е тейлоровото развитие на $\psi_\xi(t)$ около точката $t = 0$, където за оценката на остатъчния член $R_n(t)$ получаваме

$$|R_n(t)| = \left| \mathbf{E}\left(e^{it\xi} - \sum_{k=0}^{n} \frac{(it\xi)^k}{k!} \right) \right| \leqq \mathbf{E}\left| e^{it\xi} - \sum_{k=0}^{n} \frac{(it\xi)^k}{k!} \right|. \tag{14}$$

Нека $\zeta = e^{it\xi} - \sum\limits_{k=0}^{n} \dfrac{(it\xi)^k}{k!}$. Тогава от лема 1 намираме оценките

$$|\zeta| \leqq \frac{|t\xi|^{n+1}}{(n+1)!},$$

$$|\zeta| \leqq \left| e^{it\xi} - \sum_{k=0}^{n-1} \frac{(it\xi)^k}{k!} \right| + \frac{|t\xi|^n}{n!} \leqq \frac{2|t\xi|^n}{n!}.$$

Оттук и от (14), като въведем събитието $A = \{|\xi| \leqq N\}$, получаваме

$$
(15) \qquad \begin{aligned} |R_n(t)| &\leqq \frac{|t|^{n+1}}{(n+1)!}\mathbf{E}|\xi|^{n+1}I_A + 2\frac{|t|^n}{n!}\mathbf{E}|\xi|^n I_{\overline{A}} \\ &\leqq \frac{|t|^{n+1}}{(n+1)!}N^{n+1} + 2\frac{|t|^n}{n!}\int_N^\infty |x|^n dF_\xi(x). \end{aligned}
$$

Тъй като $\mathbf{P}(\overline{A}) = \mathbf{P}\{|\xi| \geqq N\} \to 0$ при $N \to \infty$, то за всяко $\varepsilon > 0$ съществува такова N, че $\mathbf{E}|\xi|^n I_{\overline{A}} \leqq \frac{\varepsilon}{4}$. Да положим $\delta = \frac{(n+1)\varepsilon}{2N^{n+1}}$. Тогава от (15) следва, че при $|t| < \delta$ имаме $|R_n(t)| \leqq \frac{|t|^n}{n!}\varepsilon$, което доказва теоремата.

Както при пораждащите функции и преобразуванията на Лаплас – Стилтес съответствието между характеристичните функции и функциите на разпределение е взаимноеднозначно и взаимнонепрекъснато. Доказателството на този факт ще отложим за следващата глава, а сега ще разгледаме някои примери.

Пример 1. Характеристичната функция на изродено разпределение $\mathbf{P}\{\xi = C\} = 1$ е $\psi_\xi(t) = e^{itC}$.

Пример 2. Ако $\xi \in \mathcal{N}(0,1)$, то

$$
(16) \qquad \begin{aligned} \psi_\xi(t) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{itx - x^2/2}dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(\cos tx)e^{-x^2/2}dx + \frac{i}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(\sin tx)e^{-x^2/2}dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x(\cos tx)e^{-x^2/2}dx = u(t), \end{aligned}
$$

тъй като подинтегралната функция в имагинерната част е нечетна.

Диференцирайки (16) по t, получаваме

$$
u'(t) = -\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x(\sin tx)e^{-x^2/2}dx.
$$

Оттук, като интегрираме по части, намираме

$$u'(t) = \frac{\sin tx}{\sqrt{2\pi}} e^{-x^2/2}\Big|_{-\infty}^{\infty} - \frac{t}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\cos tx) e^{-x^2/2} dx = -tu(t).$$

Решаваме това уравнение при начално условие $u(0) = 1$:

(17) $$\psi_\xi(t) = u(t) = e^{-t^2/2}.$$

В общия случай, когато $\eta \in \mathcal{N}(a, \sigma^2)$, от $\eta = \sigma\xi + a$ по свойство (7) следва

(18) $$\psi_\eta(t) = \exp\left\{ita - \frac{\sigma^2 t^2}{2}\right\}.$$

Пример 3. При $\xi \in \mathcal{U}(a, b)$ имаме

(19) $$\psi_\xi(t) = \frac{1}{b-a} \int_a^b e^{itx} dx = \frac{e^{itb} - e^{ita}}{it(b-a)},$$

откъдето при $a = -c$, $b = c$ получаваме

(20) $$\psi_\xi(t) = \frac{(\sin tc)}{tc}.$$

Теорема 3. *Ако случайната величина ξ е симетрична (т. е. ξ и $-\xi$ са еднакво разпределени), то $\psi_\xi(t)$ е четна и реална функция.*

Доказателство. За произволна случайна величина η е в сила

(21) $$\overline{\psi_\eta(t)} = \overline{\mathbf{E}e^{it\eta}} = \mathbf{E}\overline{e^{it\eta}} = \mathbf{E}e^{-it\eta} = \psi_\eta(-t).$$

От симетричността на ξ се получава

(22) $$\psi_\xi(t) = \mathbf{E}e^{it\xi} = \mathbf{E}e^{-it(-\xi)} = \psi_\xi(-t).$$

От (21) и (22) следва, че $\psi_\xi(t) = \psi_\xi(-t) = \overline{\psi_\xi(t)}$, което доказва теоремата.

Съотношенията (17) и (20) илюстрират добре теорема 3.

Нека случайният вектор $\zeta = (\xi_1, \ldots, \xi_k)$ има многомерна функция на разпределение

(23) $$F_\zeta(x) = \mathbf{P}\{\xi_1 < x_1, \ldots, \xi_k < x_k\},$$

където $\mathbf{x} = (x_1, \ldots, x_k) \in \mathbf{R}^k$. Многомерна характеристична функция на случайния вектор ζ наричаме

$$(24) \qquad \psi_\zeta(t) = \psi_{\xi_1,\ldots,\xi_k}(t_1,\ldots, t_n) = \mathbf{E}^{i(t,\zeta)} = \int_{\mathbf{R}^k} e^{i(t,x)} dF_\zeta(x),$$

където $\mathbf{t} = (t_1, \ldots, t_k)$, $(\mathbf{t}, \zeta) = \sum\limits_{m=1}^{k} t_m \xi_m$.

Многомерните характеристични функции имат свойства аналогични на едномерните характеристични функции:

1) при всички $t \in \mathbf{R}^k$ имаме $|\psi_\zeta(t)| \leqq 1$ и $\psi_\zeta(0) = 1$;

2) характеристичната функция $\psi_\zeta(t)$ е равномерно непрекъсната по $t \in \mathbf{R}^k$;

3) характеристичната функция на вектора $(\xi_1, \ldots, \xi_m)$ при $m < k$ се получава от (24) чрез полагането

$$\psi_{\xi_1,\ldots,\xi_m}(t_1, \ldots, t_m) = \psi_{\xi_1,\ldots,\xi_k}(t_1, \ldots, t_k, 0, \ldots, 0);$$

4) $\psi_{\xi_1+\cdots+\xi_k}(\tau) = \psi_{\xi_1,\ldots,\xi_k}(\tau, \ldots, \tau)$ за всяко $\tau \in \mathbf{R}^1$;

5) по характеристичната функция (24) еднозначно се възстановява функцията на разпределение (23);

6) за независимостта на случайните величини $\xi_1, \ldots, \xi_k$ е необходимо и достатъчно условието

$$(25) \qquad \psi_{\xi_1,\ldots,\xi_k}(t_1, \ldots, t_k) = \prod_{m=1}^{k} \psi_{\xi_m}(t_m);$$

7) ако смесените моменти $m_{\alpha_1,\ldots,\alpha_k} = \mathbf{E}\{\xi_1^{\alpha_1}, \ldots, \xi_k^{\alpha_k}\}$, за които $\alpha_1 + \cdots + \alpha_k = r$, са крайни, то за всички $\alpha = \alpha_1 + \cdots + \alpha_k \leqq r$ изпълнено

$$m_{\alpha_1,\ldots,\alpha_k} = i^\alpha \frac{\partial^\alpha \psi_\zeta(t_1, \ldots, t_k)}{\partial t_1^{\alpha_1}, \ldots, \partial t_k^{\alpha_k}}\Big|_{t_1=\cdots=t_k=0}.$$

Пример 4. Нека $\xi_1, \ldots, \xi_k$ са независими нормално разпределени случайни величини, като $\xi_m \in \mathcal{N}(a_m, \sigma_m^2)$, $m = 1, \ldots, k$. Тогава за многомерната характеристична функция от (24), (25) и (18) получаваме

$$(26) \qquad \psi_{\xi_1,\ldots,\xi_k}(t_1, \ldots, t_k) = \exp\left\{ i \sum_{m=1}^{k} t_m a_m - \frac{1}{2} \sum_{m=1}^{k} \sigma_m^2 t_m^2 \right\}.$$

Задачи

1. Нека $h_\xi(s) = \sum_{k=0}^{\infty} p_k s^k$ е пораждащата функция на случайната величина ξ. Покажете, че за производящата функция на редицата $q_n = \mathbf{P}\{\xi > n\}$ имаме

$$Q(s) = \sum_{k=0}^{\infty} q_k s^k = \frac{1 - h_\xi(s)}{1 - s}.$$

Докажете, че $Q(1) = \mathbf{E}\xi$.

2. Нека $\xi \in NB(r, p)$, т. е. ξ има отрицателно биномно разпределение (вж. (1.15)). Покажете, че $\mathbf{E}\xi = \dfrac{rq}{p}$ и $\mathbf{D}\xi = \dfrac{rq}{p^2}$.

3. Нека $\mathbf{E}\xi \in Bi(n, p)$ и $\lim\limits_{n \to \infty} np_n = \lambda$, $0 < \lambda < \infty$. Докажете, че $\xi_n \xrightarrow{d} \eta$ при $n \to \infty$, където $\eta \in Po(\lambda)$.

Упътване. Използвайте теоремата за непрекъснатост — теорема 1.3 (сравнете този метод с теорема 3.3.1).

4. Нека ξ и η са независими случайни величини, $\xi \in NB(n, p)$, $\eta \in NB(m, p)$. Докажете, че $\xi + \eta \in NB(m + n, p)$.

5. Нека $\{\xi_k\}$ са независими случайни величини, като $\xi_k \in \mathcal{N}(\mu_k, \sigma_k^2)$, $k = 1, \ldots, n$. Докажете, че за случайната величина $\eta = \sum_{k=1}^{n} a_k \xi_k$ имаме $\eta \in N(\mu, \sigma^2)$, където $\mu = \sum_{k=1}^{n} a_k \mu_k$ и $\sigma^2 = \sum_{k=1}^{n} a_k^2 \sigma_k^2$.

6. Нека $\xi \in Ex(\theta)$. Покажете, че ξ има характеристична функция

$$\psi_\xi(t) = \frac{1}{1 - it\theta}.$$

Упътване. Интегрирайте по части.

7. Нека ξ има плътност $f_\xi(x) = \dfrac{1}{2} e^{-|x|}$, $-\infty < x < \infty$ (разпределение на Лаплас). Покажете, че характеристичната функция на ξ е $\psi_\xi(t) = (1 + t^2)^{-1}$.

8. Нека $\xi \in \Gamma(n, \theta)$, т. е. има разпределение на Ерланг. Покажете, че $\mathbf{E}\xi = n\theta$ и $\mathbf{D}\xi = n\theta^2$.

9. Нека ξ и η са независими случайни величини, $\xi, \eta \in \mathcal{N}(a, \sigma^2)$. С помощта на характеристични функции докажете, че случайните величини $\xi + \eta$ и $\xi - \eta$ са независими.

Упътване. Използвайте (3.25) и (3.26).

10. Нека $\{\xi_n\}$ и τ са независими случайни величини, $\xi_n \in Ex(\theta)$, т.

$$\mathbf{P}\{\xi_n < x\} = 1 - \exp\left(-\frac{x}{\theta}\right), \quad x \geqq 0,$$

а τ има геометрично разпределение $\mathbf{P}\{\tau = k\} = p(1-p)^{k-1}$, $k = 1$,

Покажете, че за случайната величина $S_\tau = \xi_1 + \cdots + \xi_\tau$ имаме $S_\tau \in Ex\Big($

Седма глава

Сходимост на редици от случайни величини

§ 1. Видове сходимост

Навсякъде в този параграф ще предполагаме, че ξ_1, ξ_2, ... е редица от случайни величини, дефинирани във вероятностното пространство $(\Omega, \mathfrak{F}, \mathbf{P})$. Сега ще пристъпим към въпроса за изучаване на различните видове сходимост.

Дефиниция 1. *Ще казваме, че редицата $\{\xi_n\}$ е сходяща с вероятност единица или почти сигурно (п. с.) към случайна величина ξ, ако*

$$\mathbf{P}\{\lim_{n\to\infty} \xi_n = \xi\} = \mathbf{P}\{\lim_{n\to\infty} \sup \xi_n = \lim_{n\to\infty} \inf \xi_n\} = 1. \tag{1}$$

Означава се $\xi_n \xrightarrow{\text{п. с.}} \xi$ или $\lim\limits_{n\to\infty} \xi_n = \xi$ ($\mathbf{P}$ – п. с.). Да отбележим, че (1) е еквивалентно на съотношението

$$\mathbf{P}\{\omega : \lim_{n\to\infty} \xi_n(\omega) \neq \xi(\omega)\} = 0. \tag{2}$$

Всъщност (1) или (2) дефинират добре познатата в анализа сходимост почти навсякъде.

Теорема 1. *Необходимо и достатъчно условие за сходимост почти сигурно е за всяко $\varepsilon > 0$ да бъде изпълнено*

$$\mathbf{P}\{\sup_{k \geqq n} |\xi_k - \xi| > \varepsilon\} \to 0 \quad \textit{при} \quad n \to \infty. \tag{3}$$

Доказателство. За $r = 1, 2, \ldots$ да въведем събитията

$$A_r = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \left\{|\xi_m - \xi| > \frac{1}{r}\right\} = \inf_{n \geqq 1} \left(\sup_{m \geqq n} \left\{|\xi_m - \xi| > \frac{1}{r}\right\}\right). \tag{4}$$

Очевидно $A_r \subset A_{r+1}$ и $\{\lim\limits_{n\to\infty} \xi_n \neq \xi\} = \bigcup\limits_{r=1}^{\infty} A_r$. Следователно

$$\mathbf{P}\{\lim_{n\to\infty} \xi_n \neq \xi\} = \lim_{r\to\infty} \uparrow \mathbf{P}(A_r) = 0$$

тогава и само тогава, когато за всяко $r \geqq 1$ имаме $\mathbf{P}(A_r) = 0$, т. е.

$$\mathbf{P}\{\inf_{n \geqq 1}(\sup_{m \geqq n} |\xi_m - \xi| > \frac{1}{r})\} = 0,$$

което е еквивалентно на

$$\lim_{n\to\infty} \downarrow \mathbf{P}\{\sup_{m \geqq n} |\xi_m - \xi| > \frac{1}{r}\} = 0$$

за всяко $r \geqq 1$. Последното съотношение от своя страна е равносилно на (3).

Да отбележим, че сходимостта на случайните величини навсякъде, т. е. за всяко $\omega \in \Omega$, която използвахме досега, може да бъде заменена със сходимост почти сигурно, в частност в теоремите за монотонна сходимост под знака на интеграла, теоремата на Фату – Лебег и т. н. Преформулировката на всички тези резултати не би трябвало да затрудни особено читателя.

Дефиниция 2. *Казваме, че редицата $\{\xi_n\}$ клони по вероятност към ξ, ако за всяко $\varepsilon > 0$*

$$\mathbf{P}\{|\xi_n - \xi| > \varepsilon\} \to 0 \quad \textit{при} \quad n \to \infty. \tag{5}$$

Означава се $\xi_n \xrightarrow{\mathbf{P}} \xi$ или $\lim\limits_{n\to\infty} \xi_n = \xi\ (\mathbf{P})$.

Сходимостта по вероятност е всъщност добре познатата от анализа сходимост по мярка. С този вид сходимост вече се сблъскахме в трета глава (вж. (3.4.15)).

Тъй като от $\omega \in \{|\xi_n - \xi| > \varepsilon\}$ следва $\omega \in \{\sup\limits_{k \geqq n} |\xi_k - \xi| > \varepsilon\}$, то

$$\mathbf{P}\{|\xi_n - \xi| > \varepsilon\} \leqq \mathbf{P}\{\sup_{k \geqq n} |\xi_k - \xi| > \varepsilon\},$$

което показва, че от сходимост почти сигурно следва сходимост по вероятност. Обратното не е вярно, в което се убеждаваме от следния

Контрапример 1. За вероятностното пространство $(\Omega, \mathfrak{F}, \mathbf{P})$ нека $\Omega = [0,1]$, $\mathfrak{F} = \mathfrak{B} \cap \Omega$, а $\mathbf{P}$ е лебеговата мярка (т. е. дължината). За всяко $n = 1, 2, \ldots$ да въведем събитията $A_n^i = \left[\frac{i-1}{n}, \frac{i}{n}\right]$, $i = 1, 2, \ldots, n$, и нека $\xi_{n,i} = I(A_n^i)$ са техните индикатори.

Да отбележим, че

$$(6)\qquad \mathbf{P}\{\xi_{n,i}=1\}=\mathbf{P}\{A_n^i\}=\frac{1}{n},\quad \mathbf{P}\{\xi_{n,i}=0\}=\mathbf{P}(\overline{A_n^i})=1-\frac{1}{n},$$

при $i=1,\,2,\,\ldots,\,n$; $n=1,\,2,\,\ldots$

Да разгледаме редицата

$$(7)\qquad \xi_{1,1};\ \xi_{2,1},\ \xi_{2,2};\ \xi_{3,1},\ \xi_{3,2},\ \xi_{3,3};\ \ldots$$

Не е трудно да се види, че редицата (7) клони по вероятност към 0, тъй като за всяко $\varepsilon>0$ и $i=1,\,2,\,\ldots,\,n$ от (6) получаваме

$$\mathbf{P}\{\xi_{n,i}>\varepsilon\}=\mathbf{P}\{\xi_{n,i}=1\}=\frac{1}{n}\to 0\quad \text{при}\quad n\to\infty.$$

От друга страна, редицата (7) има две точки на сгъстяване 0 и 1, т. е. $\lim\limits_{n\to\infty}\inf\xi_n^i=0$, $\lim\limits_{n\to\infty}\sup\xi_n^i=1$. Следователно (7) не е сходяща за нито едно $\omega\in[0,1]$, откъдето $\mathbf{P}\{\lim\limits_{n\to\infty}\xi_{n,i}=0\}=0$.

Дефиниция 3. *Редицата* $\{\xi_n\}$ *с* $\mathbf{E}\xi_n^2<\infty$ *се нарича сходяща в средноквадратичен смисъл* ($\mathbf{L}_2$*-сходяща) към случайна величина* ξ *с* $\mathbf{E}\xi^2<\infty$*, ако*

$$(8)\qquad \lim_{n\to\infty}\mathbf{E}(\xi_n-\xi)^2=0.$$

Означава се $\lim\xi_n=\xi$ или $\xi_n\xrightarrow{\mathbf{L}_2}\xi$.

Случайните величини $\{\xi\}$ в $(\Omega,\mathfrak{F},\mathbf{P})$, за които $\mathbf{E}|\xi|^p<\infty$ $(p\geqq 1)$, образуват $\mathbf{L}_p$-пространство с норма $||\xi||_p=(\mathbf{E}|\xi|^p)^{1/p}$, т. е. L_p*-сходимост* означава, че $\mathbf{E}|\xi_n-\xi|^p\to 0$ при $n\to\infty$.

Тъй като по неравенството на Чебишов (4.5.12) имаме

$$(9)\qquad \mathbf{P}\{|\xi_n-\xi|>\varepsilon\}\leqq\frac{\mathbf{E}|\xi_n-\xi|^p}{\varepsilon^p}$$

за всяко $\varepsilon>0$, то (9) показва, че от $\mathbf{L}_p$-сходимост следва сходимост по вероятност.

Обратното не е вярно, тъй като ако $\mathbf{E}|\xi|^p=\infty$, то редицата $\{\xi_n\}$ с общ член $\xi_n=\xi+\dfrac{1}{n}$ е сходяща дори почти сигурно.

Пример 1. Нека $\nu_n\in Bi(n,p)$. Тогава

$$\mathbf{E}\left(\frac{\nu_n}{n}-p\right)^2=\frac{\mathbf{E}(\nu_n-np)^2}{n^2}=\frac{\mathbf{D}\nu_n}{n^2}=\frac{npq}{n^2}=\frac{pq}{n}\to 0$$

при $n \to \infty$, т. е. $\dfrac{\nu_n}{n} \xrightarrow{\mathrm{L}_2} p$.

Дефиниция 4. *Редицата $\{\xi_n\}$ се нарича сходяща по разпределение към случайната величина ξ, ако*

$$\lim_{n\to\infty} F_{\xi_n}(x) = F_\xi(x) \tag{10}$$

във всички точки на непрекъснатост на функцията на разпределение $F_\xi(x)$.

Означава се $\xi_n \xrightarrow{d} \xi$, както вече отбелязахме при интегралната теорема на Моавър – Лаплас (вж. (3.4.16)). Използва се също означението $F_n(x) \xrightarrow{d} F(x)$.

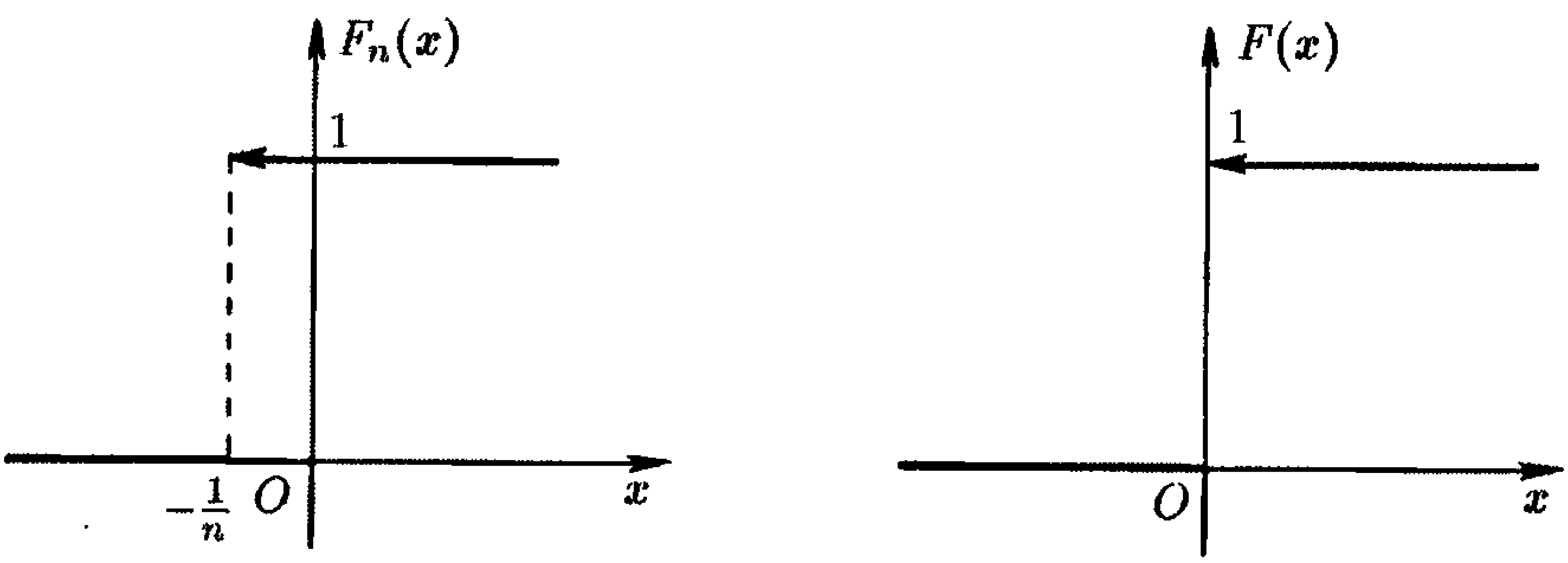
Условието за сходимост на граничната функция в точките на непрекъснатост е съществено, както се вижда от следния

Пример 2. Нека $\xi_n = -\dfrac{1}{n}$, $n \geqq 1$. Тогава очевидно $\xi_n \xrightarrow{\text{п. с.}} \xi \equiv 0$ при $n \to \infty$. Тъй като

$$\mathrm{P}\{\xi_n < x\} = F_n(x) = \begin{cases} 0 & \text{при} \quad x \leqq -\dfrac{1}{n}, \\ 1 & \text{при} \quad x > -\dfrac{1}{n}, \end{cases}$$

$$\mathrm{P}\{\xi < x\} = F(x) = \begin{cases} 0 & \text{при} \quad x \leqq 0, \\ 1 & \text{при} \quad x > 0, \end{cases}$$

то $\lim\limits_{n\to\infty} F_n(x) = F(x)$ при $x \neq 0$ (фиг. 12).



Фиг. 12

От друга страна, при $x = 0$ имаме $F_n(0) = 1$ за всички $n \geqq 1$, а $F(0) = 0$, т. е. условието за сходимост в точката 0 е нарушено.

З а б е л е ж к а. За сходимост по разпределение има смисъл да говорим и тогава, когато случайните величини ξ_n са зададени в различни вероятностни пространства $(\Omega_n, \mathfrak{F}_n, \mathbf{P}_n)$, което е принципно невъзможно в дефинициите 1 – 3. Следователно изглежда естествено, че $\xi_n \xrightarrow{d} \xi$, но няма сходимост в никакъв друг смисъл.

Теорема 2. *Ако* $\xi_n \xrightarrow{p} \xi$, *то* $\xi_n \xrightarrow{d} \xi$, *т. е. от сходимост по вероятност следва сходимост по разпределение.*

Доказателство. По условие за всяко $\varepsilon > 0$ при $n \to \infty$ имаме

(11) $$\mathbf{P}\{|\xi_n - \xi| > \varepsilon\} = \mathbf{P}\{\xi_n - \xi < -\varepsilon\} + \mathbf{P}\{\xi_n - \xi > \varepsilon\} \to 0.$$

Ще покажем, че от (11) следва

(12) $$F_n(x) = \mathbf{P}\{\xi_n < x\} \longrightarrow \mathbf{P}\{\xi < x\} = F(x)$$

за всяко $x \in \mathbf{R}^1$, което е точка на непрекъснатост на $F(x)$.

За всяко $x \in \mathbf{R}^1$ не е трудно да се провери директно, че

(13) $$\{\xi_n < x\} \subset \{\xi_n - \xi < -\varepsilon\} \cup \{\xi < x + \varepsilon\}.$$

Следователно

(14) $$F_n(x) \leqq \mathbf{P}\{\xi_n - \xi < -\varepsilon\} + F(x + \varepsilon).$$

Аналогично на (13) се проверява, че

(15) $$\{\xi_n \geqq x\} \subset \{\xi_n - \xi > \varepsilon\} \cup \{\xi \geqq x - \varepsilon\},$$

откъдето следва

(16) $$1 - F_n(x) \leqq \mathbf{P}\{\xi_n - \xi > \varepsilon\} + 1 - F(x - \varepsilon).$$

Сега от (14) и (16) получаваме неравенствата

(17) $$F(x - \varepsilon) - \mathbf{P}\{\xi_n - \xi > \varepsilon\} \leqq F_n(x) \leqq \mathbf{P}\{\xi_n - \xi < -\varepsilon\} + F(x + \varepsilon).$$

От (11) следва, че за всяко $\varepsilon > 0$ съществува такова N, че $\mathbf{P}\{\xi_n - \xi > \varepsilon\} < \varepsilon$ и $\mathbf{P}\{\xi_n - \xi < -\varepsilon\} < \varepsilon$ при $n \geqq N$. Тогава от (17) намираме

(18) $$F(x - \varepsilon) - \varepsilon < F_n(x) < F(x + \varepsilon) + \varepsilon.$$

Нека сега x е точка на непрекъснатост на функцията $F(x)$. Тогава $F(x) - \delta < F(x - \varepsilon)$ и $F(x + \varepsilon) < F(x) + \delta$, където $\delta > 0$ и $\varepsilon > 0$

могат да бъдат избирани произволно малки. По този начин от (18) при $n \geqq N$ получаваме

$$F(x) - (\varepsilon + \delta) < F_n(x) < F(x) + \varepsilon + \delta,$$

което доказва (12).

Следващият контрапример показва, че обратното на теорема 2 твърдение не е вярно.

Контрапример 2. Нека ξ, ξ_1, ξ_2, ... са независими и еднакво разпределени случайни величини, т. е.

$$F(x) = \mathbf{P}\{\xi < x\} = \mathbf{P}\{\xi_n < x\}, \quad n = 1, 2, \ldots$$

Очевидно $\xi_n \xrightarrow{d} \xi$. Директно се проверява, че за всяко $\varepsilon > 0$ имаме

$$\left\{\xi_n \geqq \frac{\varepsilon}{2},\ \xi < -\frac{\varepsilon}{2}\right\} \subset \{\xi_n - \xi > \varepsilon\}.$$

Следователно

$$\begin{aligned}\mathbf{P}\{\xi_n - \xi > \varepsilon\} &\geqq \mathbf{P}\left\{\xi_n \geqq \frac{\varepsilon}{2},\ \xi < -\frac{\varepsilon}{2}\right\} = \mathbf{P}\left\{\xi_n \geqq \frac{\varepsilon}{2}\right\}\mathbf{P}\left\{\xi < -\frac{\varepsilon}{2}\right\} \\ &= \left[1 - F\left(\frac{\varepsilon}{2}\right)\right] F\left(-\frac{\varepsilon}{2}\right).\end{aligned} \tag{19}$$

Нека сега $F(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-t^2/2}dt$. Тогава от (19) получаваме

$$\mathbf{P}\{\xi_n - \xi > \varepsilon\} \geqq \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{-\varepsilon/2} e^{-t^2/2}dt \left[1 - \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\varepsilon/2} e^{-t^2/2}dt\right]. \tag{20}$$

При $\varepsilon \to 0$ дясната страна в (20) клони към $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$, което показва, че $\mathbf{P}\{\xi_n - \xi > \varepsilon\}$ не клони към нула при $n \to \infty$, т. е. ξ_n не клони по вероятност към ξ.

И така, доказахме верността на следните вериги от импликации:

$$\xi_n \xrightarrow{\text{п. с.}} \xi \Longrightarrow \xi_n \xrightarrow{p} \xi \Longrightarrow \xi_n \xrightarrow{d} \xi$$

$$\xi_n \xrightarrow{\mathbf{L}_p} \xi \Longrightarrow \xi_n \xrightarrow{p} \xi.$$

Сега ще покажем, че в някои случаи са верни и обратните импликации.

Теорема 3. *За монотонна редица от случайни величини сходимостите п. с. и по вероятност съвпадат.*

Доказателство. Нека $\xi_n \uparrow \xi$. Тогава $\eta_n = \xi - \xi_n \downarrow 0$ и за всяко $\varepsilon > 0$ имаме $\{\eta_{n+1} \geqq \varepsilon\} \subset \{\eta_n \geqq \varepsilon\}$, откъдето следва

$$\{\lim_{n\to\infty} \eta_n \geqq \varepsilon\} = \bigcap_{n=1}^{\infty} \{\eta_n \geqq \varepsilon\}.$$

Оттук веднага получаваме

$$\mathbf{P}\{\lim_{n\to\infty} \eta_n \geqq \varepsilon\} = \lim_{n\to\infty} \mathbf{P}\{\eta_n \geqq \varepsilon\},$$

което доказва нашето твърдение.

Теорема 4. *Ако редица от случайни величини клони по разпределение към константа, тя е сходяща и по вероятност.*

Доказателство. Нека $\xi_n \xrightarrow{d} C$, т. е. при $n \to \infty$ е изпълнено

$$\mathbf{P}\{\xi_n < x\} = F_n(x) \longrightarrow F(x) = \begin{cases} 0 & \text{за } x \leqq C, \\ 1 & \text{за } x > C \end{cases}$$

за всяко $x \neq C$. Тогава за всяко $\varepsilon > 0$ имаме

$$\begin{aligned} &\mathbf{P}\{|\xi_n - C| \geqq \varepsilon\} = \mathbf{P}\{\xi_n \geqq C + \varepsilon\} + \mathbf{P}\{\xi_n \leqq C - \varepsilon\} \\ &= 1 - \mathbf{P}\{\xi < C + \varepsilon\} + \mathbf{P}\{\xi_n \leqq C - \varepsilon\} \\ &= 1 - F_n(C+\varepsilon) + F_n(C-\varepsilon) \longrightarrow 1 - F(C+\varepsilon) + F(C-\varepsilon) = 0, \end{aligned}$$

което доказва теоремата.

Теорема 5 (теорема на Слуцки). *Нека $\xi_n^i \xrightarrow{p} C_i$ при $n \to \infty$ за $i = 1, 2, \ldots, k$. Ако функцията $g(x_1, \ldots, x_k)$ е непрекъсната в точката $C = (C_1, \ldots, C_k)$, то при $n \to \infty$ имаме*

$$g(\xi_n^1, \ldots, \xi_n^k) \xrightarrow{p} g(C_1, \ldots, C_k). \tag{21}$$

Доказателство. От условието на теоремата следва, че за всяко $\delta > 0$ и за всяко $\varepsilon > 0$, и при достатъчно голямо n имаме

$$\mathbf{P}\{|\xi_n^i - C_i| > \delta\} < \frac{\varepsilon}{k}, \quad i = 1, \ldots, k. \tag{22}$$

Тъй като $g(x)$ е непрекъсната в точката C, то за всяко $\varepsilon_1 > 0$ съществува $\delta_1 > 0$, такова че $|g(x) - g(y)| < \varepsilon_1$ при $|x_i - y_i| < \delta_1$ и $i = 1, \ldots, k$. Това означава, че

$$\bigcap_{i=1}^{k}\{|\xi_n^i - C_i| < \delta_1\} \subset \{|g(\xi_n^1, \ldots, \xi_n^k) - g(C_1, \ldots, C_k)| < \varepsilon\}.$$

Като преминем към допълнителните събития и използваме тъждествата на де Морган (1.2.1), получаваме

$$\{|g(\xi_n^1, \ldots, \xi_n^k) - g(C_1, \ldots, C_k)| > \varepsilon\} \subset \bigcup_{i=1}^{k}\{|\xi_n^i - C_i| > \delta_1\}.$$

Оттук и от (22) намираме

$$\mathbf{P}\{|g(\xi_n^1, \ldots, \xi_n^k) - g(C_1, \ldots, C_k)| > \varepsilon\} \leqq \sum_{i=1}^{k}\mathbf{P}\{|\xi_n^i - C_i| > \delta_1\} < \varepsilon$$

при достатъчно големи n, което доказва (21).

Теорема 6. *Нека* $\zeta_n = \xi_n + \eta_n$, $n = 1, 2, \ldots$ *Ако* $\eta_n \xrightarrow{p} 0$ *и* $\xi_n \xrightarrow{d} \xi$ *при* $n \to \infty$, *то* $\zeta_n \xrightarrow{d} \xi$.

Доказателство. Нека x е точка на непрекъснатост на функцията $F(x) = \mathbf{P}\{\xi < x\}$. Да положим $A_n = \{\xi_n + \eta_n < x\}$, $B_n = \{|\eta_n| < \varepsilon\}$. Тъй като $A_n = A_nB_n + A_n\overline{B}_n$ и $A_n\overline{B}_n \subset \overline{B}_n$, то

$$\mathbf{P}(A_nB_n) \leqq \mathbf{P}(A_n) \leqq \mathbf{P}(A_nB_n) + \mathbf{P}(\overline{B}_n). \tag{23}$$

От друга страна, $\{\xi_n < x - \varepsilon,\ |\eta_n| < \varepsilon\} \subset A_nB_n \subset \{\xi_n < x + \varepsilon\}$ и от (23) получаваме

$$\mathbf{P}\{\xi_n < x - \varepsilon,\ |\eta_n| < \varepsilon\} \leqq \mathbf{P}\{\xi_n + \eta_n < x\} \leqq \mathbf{P}\{\xi_n < x + \varepsilon\} + \mathbf{P}\{|\eta_n| \geqq \varepsilon\}.$$

Оттук, като имаме предвид неравенството

$$\mathbf{P}\{\xi_n < x - \varepsilon,\ |\eta_n| < \varepsilon\} \geqq \mathbf{P}\{\xi_n < x - \varepsilon\} - \mathbf{P}\{|\eta_n| \geqq \varepsilon\},$$

получаваме

$$F_n(x - \varepsilon) - \mathbf{P}\{|\eta_n| \geqq \varepsilon\} \leqq \mathbf{P}\{\xi_n + \eta_n < x\} \leqq F_n(x + \varepsilon) + \mathbf{P}\{|\eta_n| \geqq \varepsilon\}, \tag{24}$$

където сме означили $F_n(x) = \mathbf{P}\{\xi_n < x\}$.

Тъй като по условие $\mathbf{P}\{|\eta_n| \geqq 0\} \to 0$ при $n \to \infty$, от (24) следва

$$F(x - \varepsilon) \leqq \underline{\lim}\,\mathbf{P}\{\zeta_n < x\} \leqq \overline{\lim}\,\mathbf{P}\{\zeta_n < x\} \leqq F(x + \varepsilon).$$

Оттук при $\varepsilon \to 0$ получаваме твърдението на теоремата, понеже x е точка на непрекъснатост на $F(x)$.

Оказва се, че ако $\mathbf{P}\{|\xi_n - \xi| > \varepsilon\}$ клони към нула достатъчно бързо, от сходимостта по вероятност следва сходимост почти сигурно.

Теорема 7. *Ако за всяко $\varepsilon > 0$ редът*

$$\sum_{n=1}^{\infty} \mathbf{P}\{|\xi_n - \xi| > \varepsilon\} \tag{25}$$

е сходящ, то $\xi_n \xrightarrow{\text{п. с.}} \xi$.

Доказателство. Като използваме свойството полуадитивност на вероятностните мерки, получаваме

$$\mathbf{P}\{\sup_{k \geqq n} |\xi_k - \xi| > \varepsilon\} = \mathbf{P}\left(\bigcup_{k=n}^{\infty} \{|\xi_k - \xi| > \varepsilon\right) \leqq \sum_{k=n}^{\infty} \mathbf{P}\{|\xi_n - \xi| > \varepsilon\} \longrightarrow 0$$

при $n \to \infty$, тъй като редът (25) е сходящ. Оттук съгласно (3) следва твърдението на теоремата.

Дефиниция 5. *Казваме, че редицата от случайни величини $\{\xi_n\}$ клони слабо към случайната величина ξ, ако за всяка ограничена и непрекъсната функция $g(x)$, $x \in \mathbf{R}^1$, е в сила*

$$\lim_{n \to \infty} \int_{-\infty}^{\infty} g(x) dF_n(x) = \int_{-\infty}^{\infty} g(x) dF(x), \tag{26}$$

където $F_n(x) = \mathbf{P}\{\xi_n < x\}$ и $F(x) = \mathbf{P}\{\xi < x\}$.

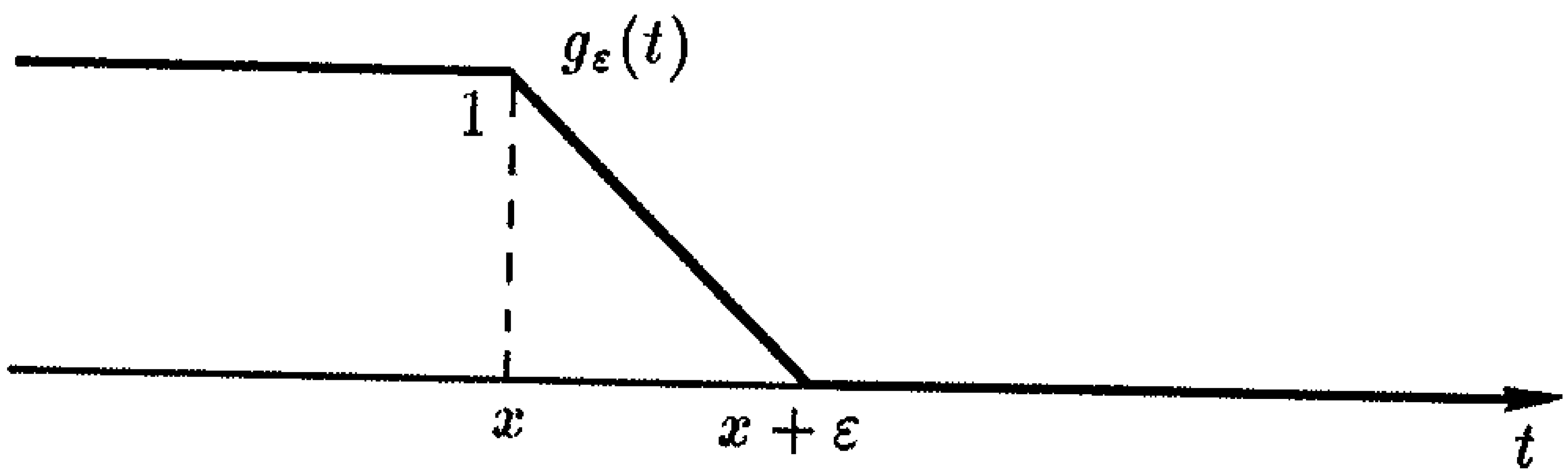
Означава се $\xi_n \Rightarrow \xi$ или $F_n \Rightarrow F$.

Да отбележим, че (26) може да се запише като

$$\lim_{n \to \infty} \mathbf{E} g(\xi_n) = \mathbf{E} g(\xi). \tag{27}$$

Понятието *слаба сходимост*, както ще видим по-нататък, играе важна роля в теория на вероятностите, а следващата теорема най-добре разкрива неговия смисъл.

Теорема 8. *Понятията слаба сходимост и сходимост по разпределение са еквивалентни, т. е. $\xi_n \Rightarrow \xi$ тогава и само тогава, когато $\xi_n \xrightarrow{d} \xi$.*



Фиг. 13

Доказателство. Нека да е в сила (26). За всяко $\varepsilon > 0$ и $x \in \mathbf{R}^1$ определяме непрекъсната функция $g_\varepsilon(t)$, равна на 1 при $t < x$ и на 0 при $t \geqq x + \varepsilon$ и линейна в интервала $[x, x+\varepsilon)$ (фиг. 13).

Тъй като

$$F_n(x) = \int_{-\infty}^{x} g_\varepsilon(t)dF_n(t) \leqq \int_{-\infty}^{\infty} g_\varepsilon(t)dF_n(t),$$

от (26) следва, че при $n \to \infty$

$$\limsup_{n\to\infty} F_n(x) \leqq \int_{-\infty}^{\infty} g_\varepsilon(t)dF(t) \leqq F(x+\varepsilon). \tag{28}$$

Ако x е точка на непрекъснатост на F, от (28) при $\varepsilon \to 0$ получаваме

$$\limsup_{n\to\infty} F_n(x) \leqq F(x). \tag{29}$$

Съвсем аналогично с помощта на функцията $h_\varepsilon(t) = g_\varepsilon(t+\varepsilon)$ се вижда, че

$$\liminf_{n\to\infty} F_n(x) \geqq F(x),$$

което заедно с (29) показва, че $\xi_n \xrightarrow{d} \xi$.

Нека сега $F_n(x) \to F(x)$ при $n \to \infty$ във всички точки на непрекъснатост на $F(x)$. Да предположим, че $|g(x)| \leqq C$ и $g(x)$ е непрекъсната функция. За всяко $\varepsilon > 0$ съществуват точки $-u$ и v на непрекъснатост на $F(x)$, такива че $F(-u) < \frac{\varepsilon}{5}C$ и $1-F(v) < \frac{\varepsilon}{5}C$. Тогава при достатъчно големи n имаме $F_n(-u) < \frac{\varepsilon}{4}C$ и $1-F_n(v) < \frac{\varepsilon}{4}C$.

Следователно при достатъчно големи n е изпълнено

$$(30)\qquad \begin{aligned} \left|\int\limits_{-\infty}^{\infty} g(x)dF_n(x) - \int\limits_{-u}^{v} g(x)dF_n(x)\right| &< \frac{\varepsilon}{2}, \\ \left|\int\limits_{-\infty}^{\infty} g(x)dF(x) - \int\limits_{-u}^{v} g(x)dF(x)\right| &< \frac{\varepsilon}{2}. \end{aligned}$$

В интервала $(-u, v]$ построяваме стъпаловидната функция $h(x)$ със скокове в точки на непрекъснатост на $F(x)$, за която $|h(x) - g(x)| < \frac{\varepsilon}{2}$ и $h(x) = 0$ при $x \notin (-u, v]$. Такава е например функцията

$$h(x) = \sum_{i=1}^{k} g(x_i)\delta_i(x),$$

където $-u = x_0 < x_1 < \cdots < x_k = v$ са подходящо избрани точки на непрекъснатост на $g(x)$, а $\delta_i(x) = I_{(x_{i-1}, x_i]}$, т. е. е индикатор на интервала $(x_{i-1}, x_i]$.

Тогава от (30) следва, че при достатъчно големи n е изпълнено

$$(31)\qquad \begin{aligned} \left|\int\limits_{-\infty}^{\infty} g(x)dF_n(x) - \int\limits_{-\infty}^{\infty} h(x)dF_n(x)\right| &< \varepsilon, \\ \left|\int\limits_{-\infty}^{\infty} g(x)dF(x) - \int\limits_{-\infty}^{\infty} h(x)dF(x)\right| &< \varepsilon. \end{aligned}$$

От друга страна, при $n \to \infty$ имаме

$$\begin{aligned} \lim \int\limits_{-\infty}^{\infty} h(x)dF_n(x) &= \lim \sum_{i=1}^{k} g(x_i)(F_n(x_i) - F_n(x_{i-1})) \\ &= \sum_{i=1}^{k} g(x_i)(F(x_i) - F(x_{i-1})) = \int\limits_{-\infty}^{\infty} h(x)dF(x). \end{aligned}$$

Оттук и от (31) получаваме

$$\lim_{h\to\infty}\sup\int_{-\infty}^{\infty} g(x)dF_n(x) \leqq \varepsilon + \lim_{n\to\infty}\sup\int_{-\infty}^{\infty} h(x)dF_n(x)$$

$$= \varepsilon + \int_{-\infty}^{\infty} h(x)dF(x) \leqq 2\varepsilon + \int_{-\infty}^{\infty} g(x)dF(x),$$

$$\lim_{h\to\infty}\inf\int_{-\infty}^{\infty} g(x)dF_n(x) \geqq \lim_{h\to\infty}\inf\int_{-\infty}^{\infty} h(x)dF_n(x) - \varepsilon$$

$$= \int_{-\infty}^{\infty} h(x)dF(x) - \varepsilon \geqq \int_{-\infty}^{\infty} g(x)dF(x) - 2\varepsilon.$$

При $\varepsilon \to 0$ следва (26), което завършва доказателството на теоремата.

§ 2. Еднозначност и непрекъснатост на съответствието между функции на разпределение и характеристични функции

В този параграф ще направим връзка между въпросите за сходимост на редица от случайни величини, разработени в § 1, и апарата на характеристичните функции, развит в § 6.3.

Теорема 1 (формула за обръщане). *Нека случайната величина ξ има функция на разпределение $F(x) = \mathbf{P}\{\xi < x\}$ и характеристична функция $\psi(t) = \mathbf{E}e^{it\xi}$. Тогава във всички точки на непрекъснатост x и y на функцията F е в сила*

$$F(x) - F(y) = \frac{1}{2\pi}\lim_{\sigma\to 0}\int_{-\infty}^{\infty}\frac{e^{-ity} - e^{-itx}}{it}\psi(t)e^{-\sigma^2t^2}dt. \tag{1}$$

Доказателство. Да предположим първо, че ξ има плътност $f(t)$ и съответната характеристична функция $\psi(t) = \int_{-\infty}^{\infty} e^{itx}f(x)dx$

е абсолютно интегруема, т. е.

$$\int_{-\infty}^{\infty} |\psi(t)|dt < \infty. \tag{2}$$

Според известната от анализа формула за обръщане на трансформации на Фурие имаме

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{itx}\psi(t)dt. \tag{3}$$

От (3) получаваме

$$F(x) - F(y) = \int_{y}^{x} f(z)dz$$

$$= \int_{y}^{x} \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itz}\psi(t)dt \right) dz = \frac{1}{2\pi} \int_{-\infty}^{\infty} \psi(t) \left(\int_{y}^{x} e^{-itz}dz \right) dt.$$

Тук сме използвали (2), за да сменим реда на интегриране, откъдето окончателно намираме

$$F(x) - F(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-ity} - e^{-itx}}{it} \psi(t)dt. \tag{4}$$

Да разгледаме сега общия случай, където $\psi(t)$ е характеристична функция на произволна случайна величина ξ с функция на разпределение $F(x)$. Нека случайната величина η, зададена в същото вероятностно пространство с ξ, е независима от нея и $\eta \in \mathcal{N}(0, 2\sigma^2)$. Съгласно (6.3.18) за нейната характеристична функция имаме $\psi_\eta(t) = e^{-\sigma^2 t^2}$. Тогава от (6.3.6) за характеристичната функция на $\xi + \eta$ намираме $\psi_{\xi+\eta}(t) = \psi(t)e^{-\sigma^2 t^2}$.

Тъй като $|\psi(t)| \leqq 1$, следва, че

$$\int_{-\infty}^{\infty} |\psi_{\xi+\eta}(t)|dt \leqq \int_{-\infty}^{\infty} e^{-\sigma^2 t^2} dt < \infty. \tag{5}$$

Съотношението (5) показва, че са в сила равенства от типа (3) и (4), т. е.

$$(6) \qquad F_{\xi+\eta}(x) - F_{\xi+\eta}(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-ity} - e^{-itx}}{it} \psi(t) e^{-\sigma^2 t^2} dt.$$

От друга страна, за всяко $\varepsilon > 0$ е изпълнено

$$\mathbf{P}\{|\eta| > \varepsilon\} = \mathbf{P}\{\eta > \varepsilon\} + \mathbf{P}\{\eta < -\varepsilon\}$$

$$= \frac{1}{\sqrt{4\pi\sigma^2}} \int_{\varepsilon}^{\infty} e^{-\frac{z^2}{4\sigma^2}} dz + \frac{1}{\sqrt{4\pi\sigma^2}} \int_{-\infty}^{-\varepsilon} e^{-\frac{z^2}{4\sigma^2}} dz = \frac{1}{\sqrt{\pi\sigma^2}} \int_{\varepsilon}^{\infty} e^{-\frac{z^2}{4\sigma^2}} dz.$$

Следователно

$$\mathbf{P}\{|\eta| > \varepsilon\} \leqq \frac{e^{-\varepsilon^2/4\sigma^2}}{\sqrt{\pi\sigma^2}} \to 0 \quad \text{при} \quad \sigma \to 0,$$

т. е. случайната величина $\eta = \eta_\sigma$ клони по вероятност към нула при $\sigma \to 0$. Оттук съгласно теорема 1.6 следва, че $F_{\xi+\eta}(x) \to F(x)$ при $\sigma \to 0$ във всяка точка на непрекъснатост на $F(x)$. От (6) окончателно получаваме

$$F(x) - F(y) = \lim_{\sigma \to 0} [F_{\xi+\eta}(x) - F_{\xi+\eta}(y)] = \frac{1}{2\pi} \lim_{\sigma \to 0} \int_{-\infty}^{\infty} \frac{e^{-ity} - e^{-itx}}{it} \psi(t) e^{-\sigma^2 t^2} dt.$$

Следствие 1. *Ако* $\int_{-\infty}^{\infty} \left| \frac{\psi(t)}{t} \right| dt < \infty$, *можем да извършим граничен преход под знака на интеграла в* (1), *откъдето получаваме*

$$(7) \qquad F(x) - F(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-ity} - e^{-itx}}{it} \psi(t) dt.$$

При доказателството на теорема 1 използвахме метода на изглаждане, т. е. вместо първоначалното разпределение $F_\xi(x)$ разглеждаме разпределението $F_{\xi+\eta}(x)$, където $F_\eta(x)$ е някаква „гладка“ функция.

Теорема 2 (теорема за единственост). *На всяка характеристична функция съответства само една функция на разпределение.*

Доказателство. От (1) следва, че разликата $F(x_1)-F(x_2)$ в точките на непрекъснатост на функцията на разпределение $F(x)$ еднозначно се определя от характеристичната функция $\psi(t)$. Ако в разликата $F(x_1)-F(x_2)$ положим $x_2 \to -\infty$ и x_2 са измежду точките на непрекъснатост на $F(x)$, то $F(x)$ е еднозначно дефинирана в точките на непрекъснатост x_1. Тъй като за всяко x имаме $F(x)=\lim F(x_1)$ при $x_1 \downarrow x$, като границата е по точките на непрекъснатост x_1, следва, че $F(x)$ еднозначно се определя от $\psi(t)$.

Теорема 3. *Нека* $p_k = \mathbf{P}\{\xi = a+bk\}$, $\sum\limits_{k=-\infty}^{\infty} p_k = 1$ *и съответната характеристична функция е* $\psi_\xi(t) = \sum\limits_{k=-\infty}^{\infty} p_k e^{i(a+bk)t}$. *Тогава за всяко цяло число k е изпълнено*

$$p_k = \frac{b}{2\pi} \int\limits_{-\pi/b}^{\pi/b} \psi_\xi(x) e^{-i(a+bk)x} dx. \tag{8}$$

Доказателство. Подинтегралният израз може да бъде представен във вид на ред, където коефициентът пред p_k е равен на $\exp\{i(n-k)bx\}$, интегралът от който е 0 при $n \neq k$ и е равен на $\dfrac{2\pi}{b}$ при $n=k$, което доказва (8).

Дефиниция 1. *Функцията $G(x)$ ще наричаме финитна, ако е равна на нула извън някакъв краен интервал и има непрекъсната втора производна.*

Лема 1. *Нека* $F_n(x) = \mathbf{P}\{\xi_n < x\}$ *и* $\psi_n(t) = \mathbf{E}e^{it\xi_n}$. *Ако за всяко* $t \in \mathbf{R}^1$ *имаме* $\psi_n(t) \longrightarrow \psi(t) = \mathbf{E}e^{it\xi}$ *при* $n \to \infty$, *то*

$$\lim_{n\to\infty} \int\limits_{-\infty}^{\infty} G(x)dF_n(x) = \int\limits_{-\infty}^{\infty} G(x)dF(x) \tag{9}$$

за всяка финитна функция $G(x)$, *където* $F(x) = \mathbf{P}\{\xi < x\}$.

Доказателство. За всяка финитна функция $G(x)$ трансформацията на Фурие

$$g(t) = \int\limits_{-\infty}^{\infty} e^{itx} G(x) dx$$

е абсолютно интегруема, в което се убеждаваме, като интегрираме по части

$$g(t) = -\frac{1}{it}\int_{-\infty}^{\infty} e^{itx}G'(x)dx = -\frac{1}{t^2}\int_{-\infty}^{\infty} e^{itx}G''(x)dx.$$

Следователно е в сила формулата за обръщане, т. е.

$$G(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}g(t)dt.$$

От друга страна,

$$\int_{-\infty}^{\infty} G(x)dF(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} g(t)\left\{\int_{-\infty}^{\infty} e^{-itx}dF(x)\right\}dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} g(t)\psi(-t)dt, \tag{10}$$

където смяната на реда на интегриране е законна поради абсолютната сходимост на интегралите.

Тъй като $|g(t)\psi_n(-t)| \leqq |g(t)|$, където $|g(t)|$ е интегруема, то от $\psi_n(t) \to \psi(t)$, като използваме представянето (10), при $n \to \infty$ получаваме

$$\int_{-\infty}^{\infty} G(x)dF_n(x) = \int_{-\infty}^{\infty} g(t)\psi_n(-t)dt \longrightarrow \int_{-\infty}^{\infty} g(t)\psi(-t)dt = \int_{-\infty}^{\infty} G(x)dF(x).$$

Лемата е доказана.

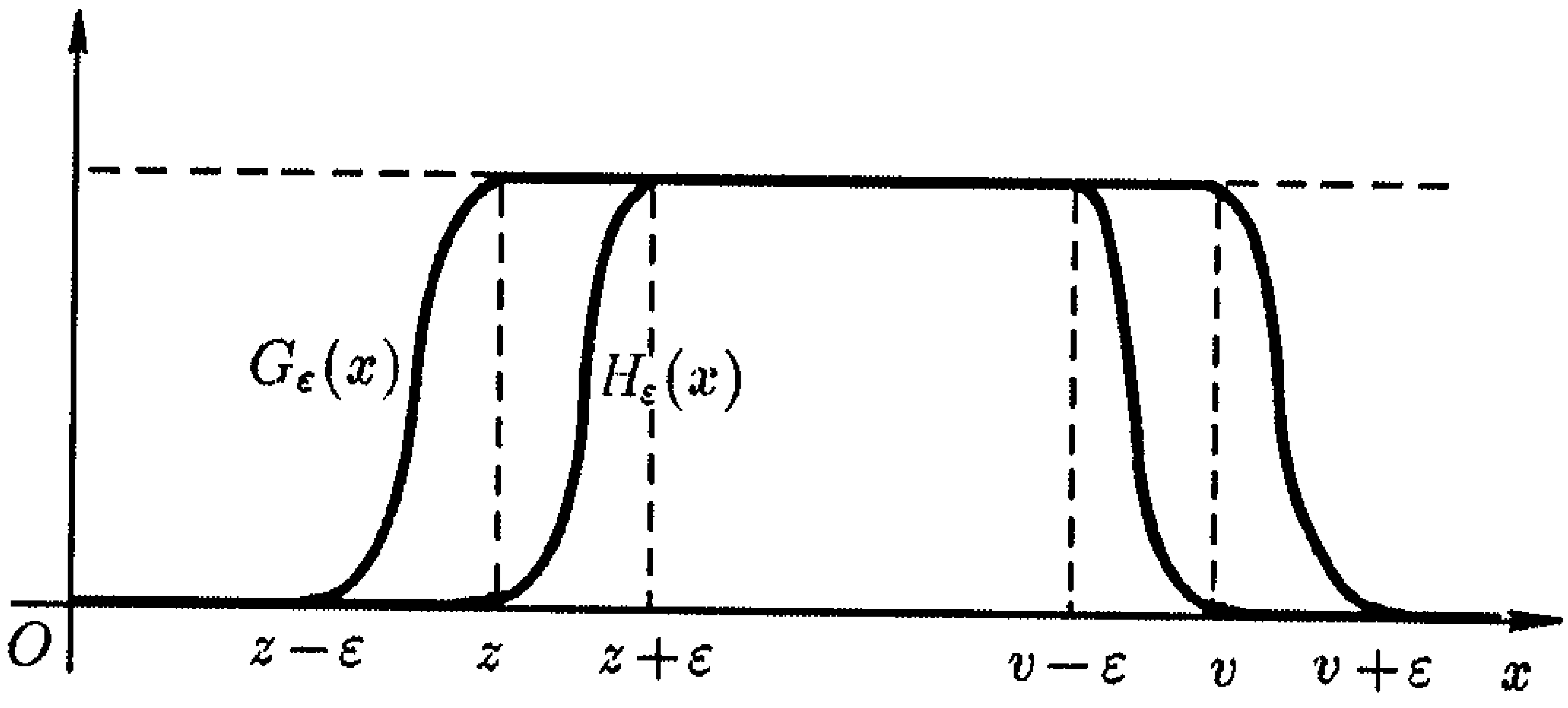
Теорема 4 (теорема за непрекъснатост). *Нека са дадени* $F_n(x) = \mathbf{P}\{\xi_n < x\}$ *и* $\psi_n(t) = \mathbf{E}e^{it\xi_n}$, $n = 1, 2, \dots$ *За сходимостта* $\xi_n \xrightarrow{d} \xi$ *е необходимо и достатъчно да бъде изпълнено* $\psi_n(t) \to \psi(t)$ *за всяко* $t \in \mathbf{R}^1$, *където* $\psi(t) = \mathbf{E}e^{it\xi}$.

Доказателство. От $\xi_n \xrightarrow{d} \xi$ по теорема 1.7 следва $\xi_n \Rightarrow \xi$, т. е. съгласно (1.27) имаме $\psi_n(t) = \mathbf{E}g(\xi_n) \to \mathbf{E}g(\xi) = \psi(t)$, тъй като $g(x) = e^{itx}$ е ограничена и непрекъсната функция.

Ще докажем достатъчността. За всяко $\varepsilon > 0$ и точки z, v на непрекъснатост на $F(x)$ построяваме финитна функция $G_\varepsilon(x)$, такава че $0 \leqq G_\varepsilon(x) \leqq 1$, като $G_\varepsilon(x) = 1$ при $x \in [z, v]$ и $G_\varepsilon(x) = 0$, когато

$x \notin [z-\varepsilon, v+\varepsilon]$ (фиг. 14). Съгласно лема 1 при $n \to \infty$ имаме

$$\text{(11)}\quad \lim_{n\to\infty} \sup(F_n(v)-F_n(z)) \leqq \lim_{n\to\infty} \sup \int_{-\infty}^{\infty} G_\varepsilon(x)dF_n(x) = \int_{-\infty}^{\infty} G_\varepsilon(x)dF(x)$$
$$\leqq F(v+\varepsilon) - F(z-\varepsilon).$$



Фиг. 14

Аналогично построяваме финитната функция $H_\varepsilon(x)$, такава че $0 \leqq H_\varepsilon(x) \leqq 1$, където $H_\varepsilon(x) = 1$ при $x \in [z+\varepsilon, v-\varepsilon]$ и $H_\varepsilon(x) = 0$ при $x \notin [z, v]$ (фиг. 14). Тогава от лема 1 при $n \to \infty$ следва

$$\text{(12)}\quad \lim_{n\to\infty} \inf(F_n(v)-F_n(z)) \geqq \lim_{n\to\infty} \inf \int_{-\infty}^{\infty} H_\varepsilon(x)dF_n(x) = \int_{-\infty}^{\infty} H_\varepsilon(x)dF(x)$$
$$\geqq F(v-\varepsilon) - F(z+\varepsilon).$$

От (11) и (12), тъй като z и v са точки на непрекъснатост, а ε може да бъде направено произволно малко, получаваме

$$\lim_{n\to\infty} (F_n(v) - F_n(z)) = F(v) - F(z),$$

което доказва теоремата.

Задачи

1. Нека $g(x)$ е непрекъсната и ограничена функция и $\xi_n \xrightarrow{d} \xi$ при $n \to \infty$. Докажете, че $g(\xi_n) \xrightarrow{d} g(\xi)$.

Упътване. Използвайте характеристични функции и слаба сходимост.

2. Нека $\{\xi_n\}$ е редица от равномерно ограничени случайни величини, т. е. $|\xi_n| \leqq C < \infty$ п. с. за $n = 1, 2, \ldots$ Докажете, че $\xi_n \xrightarrow{p} 0$ тогава и само тогава, когато $\xi_n \xrightarrow{\mathbf{L}_1} 0$.

3. Нека

$$\mathbf{P}\{\xi_n = -n-4\} = \frac{1}{n+4}, \ \mathbf{P}\{\xi_n = -1\} = 1 - \frac{4}{n+4}, \ \mathbf{P}\{\xi_n = n+4\} = \frac{3}{n+4}.$$

Покажете, че $\xi_n \xrightarrow{p} \xi$, където $\mathbf{E}\xi \neq \lim\limits_{n\to\infty} \mathbf{E}\xi_n$.

4. Нека $\xi_n \in Bi(n,p)$ и $\lim\limits_{n\to\infty} nP_n = \lambda > 0$. Докажете, че $\xi_n \xrightarrow{d} \eta \in Po(\lambda)$.

Упътване. Използвайте характеристични функции.

5. Случайната величина $\xi \in \Gamma(\alpha, \beta)$, $\alpha > 0$, $\beta > 0$, при $x \geqq 0$ има плътност $f_\xi(x) = \dfrac{x^{\alpha-1}\, e^{-x/\beta}}{\Gamma(\alpha)\, \beta^\alpha}$. Покажете, че характеристичната функция на ξ е $\psi_\xi(t) = (1 - i\,t\beta)^{-\alpha}$.

6. Ако $\xi_n \in \Gamma\left(\alpha_n, \dfrac{\beta}{\alpha_n}\right)$, докажете, че $\xi_n \xrightarrow{\mathbf{P}} \beta$ при $\alpha_n \to \infty$.

Упътване. Използвайте характеристични функции.

7. Нека $\xi_n \in \Gamma(\alpha_n, \beta)$. Ако $\alpha_n \to \infty$ при $n \to \infty$, покажете, че

$$\eta_n = \frac{\xi_n - \alpha_n \beta}{\beta\sqrt{\alpha_n}} \xrightarrow{d} \eta \in \mathcal{N}(0, 1).$$

Упътване. Използвайте теоремата за непрекъснатост на характеристични функции.

8. Нека $\xi_n = \max\{\eta_1, \ldots, \eta_n\}$, където $\{\eta_n\}$ са независими и $\eta_k \in \mathcal{U}(0, 1)$, $k = 1, \ldots, n$. Покажете, че $\zeta_n = n(1 - \xi_n) \xrightarrow{d} \zeta \in Ex(1)$.

Упътване. Намерете функцията на разпределение на ζ_n.

9. Нека $\mathbf{P}\{\xi_n = n^{\alpha/r}\} = \dfrac{1}{n}$ и $\mathbf{P}\{\xi_n = 0\} = 1 - \dfrac{1}{n}$, $n = 1, 2, \ldots$ Ако α, $r \geqq 1$, докажете, че $\xi_n \xrightarrow{\mathbf{P}} 0$, но ξ_n не е сходяща в $\mathbf{L}_r$.

10. Нека $\eta_n = \max\{\xi_1, \ldots, \xi_n\}$, където $\{\xi_k\}$ са независими и еднакво разпределени случайни величини с функция на разпределение $F(x) = 1 - x^{-\alpha}$ при $x \geqq 1$ и 0 в противен случай. Покажете, че $\zeta_n = \dfrac{\eta_n}{n^{1/\alpha}} \xrightarrow{d} \zeta$, където $F_\zeta(x) = \exp\{-x^{-\alpha}\}$ при $x > 0$.

Осма глава

Закони за големите числа

§ 1. Слаби закони за големите числа

Както вече отбелязахме в трета глава (вж. (3.4.15)), теоремата на Бернули е пример за слаб закон за големите числа, т. е. ако $\{\xi_k\}$ са независими случайни величини, $\xi_k \in Bi(1,p)$, то при $n \to \infty$

$$\frac{\nu_n}{n} = \frac{1}{n}\sum_{k=1}^{n}\xi_k \xrightarrow{p} p = \mathbf{E}\xi_1. \tag{1}$$

Сега ще изследваме сходимостта на редици от типа (1) в общия случай, когато случайните величини ξ_k имат произволно разпределение. Да отбележим, че (1) може да се представи в еквивалентна форма

$$\eta_n = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - \mathbf{E}\xi_i) \xrightarrow{p} 0 \quad \text{при} \quad n \to \infty. \tag{2}$$

Дефиниция 1. *Ако е в сила граничното съотношение* (2), *ще казваме, че редицата* $\{\xi_k\}$ *удовлетворява слаб закон за големите числа* (СЗГЧ). *Ако в* (2) *сходимостта е почти сигурно, ще казваме, че е в сила усилен закон за големите числа* (УЗГЧ).

Да отбележим, че в (2) случайните величини $\{\xi_k\}$ не е задължително да бъдат еднакво разпределени. Ако $\{\xi_k\}$ са еднакво разпределени, от (2) следва

$$\frac{1}{n}\sum_{k=1}^{n}\xi_k \xrightarrow{p} a = \mathbf{E}\xi_1 \quad \text{при} \quad n \to \infty. \tag{3}$$

Тъй като от сходимост п. с. следва сходимост по вероятност, то очевидно от УЗГЧ следва СЗГЧ.

Законите за големите числа (ЗГЧ) се явяват обобщение на огромния опит, натрупан от човечеството. Впрочем всеки интуитивно чувства, че ако разполага например с няколко измервания над даден обект, ще получи нещо по-добро, ако вземе средното аритметично от тези измервания. (В този смисъл може да се тълкува и народната поговорка: „Седем пъти мери — един път режи“). Естествено е в рамките на съответния математически модел да си поставим въпроса: Защо е така и кога всъщност това е възможно. Отговор на тези въпроси дават ЗГЧ, посочващи общите условия, което влече след себе си статистическата устойчивост на средните аритметични.

Като илюстрация за емпиричните предпоставки на ЗГЧ може да се приведе следният пример. Съгласно съвременните физически представи всеки газ се състои от много отделни молекули, които се движат хаотично, като за всяка отделна молекула не могат да се посочат нито скоростта, нито местоположението. Но характеристиките на газа (налягане, температура) не се определят от поведението на една отделна молекула, а от сумарното им действие. Така например съгласно ЗГЧ при определени условия налягането на газа трябва да бъде постоянно и на практика наистина е така, независимо че различните молекули имат различно поведение.

Теорема 1 (теорема на Марков). *Ако*

$$\frac{1}{n^2}\mathbf{D}\sum_{i=1}^{n}\xi_i \longrightarrow 0 \quad \textit{при} \quad n\to\infty, \tag{4}$$

то е в сила СЗГЧ, *т. е.* (2).

Доказателство. От неравенството на Чебишов (4.5.13) за всяко $\varepsilon > 0$ при $n\to\infty$ получаваме

$$\mathbf{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}(\xi_k - \mathbf{E}\xi_k)\right| \geqq \varepsilon\right\} \leqq \frac{\mathbf{E}\left(\sum\limits_{k=1}^{n}\xi_k - \mathbf{E}\sum\limits_{k=1}^{n}\xi_k\right)^2}{n^2\varepsilon^2} = \frac{\mathbf{D}\left(\sum\limits_{k=1}^{n}\xi_k\right)}{n^2\varepsilon^2} \longrightarrow 0.$$

Следствие 1 (теорема на Чебишов). *Ако случайните величини $\xi_1, \xi_2, \ldots$ са независими и дисперсиите им са ограничени, т. е. $\mathbf{D}\xi_n \leqq C$, то е в сила* (2), *т. е.* СЗГЧ.

Доказателство. Директно се проверява условието (4):

$$\frac{1}{n^2}\mathbf{D}\left(\sum_{k=1}^{n}\xi_k\right) = \frac{1}{n^2}\sum_{k=1}^{n}\mathbf{D}\xi_k \leqq \frac{nC}{n^2} = \frac{C}{n} \longrightarrow 0 \quad \text{при} \quad n\to\infty.$$

Следствие 2. *Ако* $\xi_1, \xi_2, \ldots$ *са независими и еднакво разпределени случайни величини с крайна дисперсия* $\mathbf{D}\xi_1 = d$, *то е в сила* СЗГЧ, *т. е.* (3).

Доказателство. Проверява се верността на условието (4):

$$\frac{1}{n^2}\mathbf{D}\left(\sum_{k=1}^{n}\xi_k\right) = \frac{1}{n^2}\sum_{k=1}^{n}\mathbf{D}\xi_k = \frac{d}{n} \longrightarrow 0 \quad \text{при} \quad n \to \infty.$$

Както видяхме в теорема 1, в този случай

$$\frac{1}{n}\left(\sum_{k=1}^{n}\xi_k - \mathbf{E}\xi_k\right) = \frac{1}{n}\sum_{k=1}^{n}\xi_k - a \xrightarrow{p} 0 \quad \text{при} \quad n \to \infty.$$

Теорема 2 (теорема на Хинчин). *Ако* $\{\xi_n\}$ *са независими и еднакво разпределени случайни величини с крайно математическо очакване* $a = \mathbf{E}\xi_n$, *то е в сила* (3), *т. е.* СЗГЧ.

Доказателство. Нека $\psi(t) = \mathbf{E}e^{it\xi_1}$ и $\eta_n = \frac{1}{n}\sum_{k=1}^{n}\xi_k$. Тогава от мултипликативното свойство (6.3.6) получаваме

$$\psi_n(t) = \mathbf{E}e^{it\eta_n} = \left[\psi\left(\frac{t}{n}\right)\right]^n.$$

Съгласно теорема 6.3.2 е в сила представянето $\psi(t) = 1 + iat + o(t)$. Следователно

$$\psi_n(t) = \left[1 + ia\frac{t}{n} + o\left(\frac{t}{n}\right)\right]^n \longrightarrow e^{iat} \quad \text{при} \quad n \to \infty. \tag{5}$$

Очевидно $\psi_a(t) = \int\limits_{-\infty}^{\infty} e^{itx}dF_a(x) = e^{iat}$, където $F_a(x)$, равна на 0 при $x \leqq a$ и равна на 1 при $x \geqq a$, е функция на разпределение на константата a.

От (5) по теоремата за непрекъснатост на характеристични функции (вж. теорема 7.2.4) следва, че $F_n(x) = \mathbf{P}\{\eta_n < x\} \to F_a(x)$ при $n \to \infty$ във всички точки на непрекъснатост на $F_a(x)$, т. е. при $x \neq a$. Съгласно теорема 7.1.4 сходимостта $\eta_n \xrightarrow{d} a$ е еквивалентна на $\eta_n \xrightarrow{p} a$, което доказва теоремата.

Теорема 3. *За да бъде сходяща по вероятност към* 0 *една редица от случайни величини* $\{\eta_n\}$, *е необходимо и достатъчно условието*

$$\mathbf{E}\left\{\frac{\eta_n^2}{1+\eta_n^2}\right\} \longrightarrow 0 \quad \text{при} \quad n \to \infty. \tag{6}$$

Доказателство. *Достатъчност.* Нека е в сила (6). Да означим $G_n(x) = \mathbf{P}\{\eta_n < x\}$. Тогава за всяко $\varepsilon > 0$ при $n \to \infty$ имаме

$$\mathbf{P}\{|\eta_n| \geqq \varepsilon\} = \int\limits_{\{|x| \geqq \varepsilon\}} dG_n(x) \leqq \frac{1+\varepsilon^2}{\varepsilon^2} \int\limits_{\{|x| \geqq \varepsilon\}} \frac{x^2}{1+x^2}\, dG_n(x)$$

$$\leqq \frac{1+\varepsilon^2}{\varepsilon^2} \int\limits_{-\infty}^{\infty} \frac{x^2}{1+x^2}\, dG_n(x) = \frac{1+\varepsilon^2}{\varepsilon^2}\, \mathbf{E}\{\eta_n^2 | (1+\eta_n^2)\} \longrightarrow 0.$$

Тук сме използвали лесно проверяемия факт, че за функцията $g(t) = \dfrac{t^2}{1+t^2}$ при $|x| \geqq \varepsilon$ имаме $g(\varepsilon) \leqq g(x)$, т. е. $\dfrac{x^2}{1+x^2} \cdot \dfrac{\varepsilon^2}{1+\varepsilon^2} \geqq 1$.

Необходимост. Нека $\eta_n \xrightarrow{p} 0$. За всяко $\varepsilon > 0$ получаваме

$$\mathbf{P}\{|\eta_n| \geqq \varepsilon\} = \int\limits_{\{|x| \geqq \varepsilon\}} dG_n(x) \geqq \int\limits_{\{|x| \geqq \varepsilon\}} \frac{x^2}{1+x^2} dG_n(x)$$

$$= \int\limits_{-\infty}^{\infty} \frac{x^2}{1+x^2}\, dG_n(x) - \int\limits_{\{|x|<\varepsilon\}} \frac{x^2}{1+x^2}\, dG_n(x)$$

$$\geqq \mathbf{E}\{\eta_n^2 | (1+\eta_n^2)\} - \int\limits_{\{|x|<\varepsilon\}} x^2 dG_n(x) \geqq \mathbf{E}\{\eta_n^2 | (1+\eta_n^2)\} - \varepsilon^2.$$

Следователно

$$0 \leqq \mathbf{E}\{\eta_n^2 | (1+\eta_n^2)\} \leqq \mathbf{P}\{|\eta_n| \geqq \varepsilon\} + \varepsilon^2,$$

което доказва (6).

Следствие 3. *За редица от случайни величини* $\{\xi_n\}$ *условието* (6) *е необходимо и достатъчно за* СЗГЧ, *където случайните величини* $\{\eta_n\}$ *се определят в* (2).

Това е най-обща форма на СЗГЧ, но за съжаление условието (6) е трудно проверяемо на практика.

§ 2. Лема на Борел – Кантели и неравенство на Колмогоров

Резултатите в този параграф, освен че са интересни сами за себе си, играят и важна роля при изследване на усилените закони за големите числа.

Лема 1 (Борел – Кантели). *Ако за редица от събития* $\{A_n\}$ *е в сила*

$$\sum_{n=1}^{\infty} \mathbf{P}(A_n) < \infty, \tag{1}$$

то $\mathbf{P}\left(\lim_{n\to\infty} \sup A_n\right) = 0$.

Ако събитията $\{A_n\}$ *са независими и*

$$\sum_{n=1}^{\infty} \mathbf{P}(A_n) = \infty, \tag{2}$$

то $\mathbf{P}\left(\lim_{n\to\infty} \sup A_n\right) = 1$.

Доказателство. Имайки предвид (1), последователно получаваме

$$\mathbf{P}\left(\lim_{n\to\infty} \sup A_n\right) = \mathbf{P}\left(\inf_{n \geqq 1} \sup_{m \geqq n} A_m\right)$$

$$= \mathbf{P}\left(\lim_{n\to\infty} \downarrow \bigcup_{m=n}^{\infty} A_m\right) = \lim_{n\to\infty} \downarrow \mathbf{P}\left(\bigcup_{m=n}^{\infty} A_m\right) \leqq \lim_{n\to\infty} \sum_{m=n}^{\infty} \mathbf{P}(A_m) = 0.$$

От друга страна, $\mathbf{P}\left(\lim_{n\to\infty} \sup A_n\right) = 1 - \lim_{n\to\infty} \mathbf{P}\left(\bigcap_{m=n}^{\infty} \overline{A}_m\right)$.

Когато $\{A_n\}$ са независими и е в сила (2), получаваме

$$\mathbf{P}\left(\bigcap_{m=n}^{\infty} \overline{A}_m\right) = \prod_{m=n}^{\infty} \mathbf{P}(\overline{A}_m) = \prod_{m=n}^{\infty} (1 - \mathbf{P}(A_m))$$

$$\leqq \prod_{m=n}^{\infty} e^{-\mathbf{P}(A_m)} = \exp\left\{-\sum_{m=n}^{\infty} \mathbf{P}(A_m)\right\} = 0.$$

Тук сме използвали известното неравенство $1 - x \leqq e^{-x}$.

Следствие 1. *Ако събитията $\{A_n\}$ са независими, то стойността на $\mathbf{P}(\lim_{n\to\infty} \sup A_n)$ е 0 или 1 тогава и само тогава, когато редът $\sum_{n=1}^{\infty} \mathbf{P}(A_n)$ е съответно сходящ или разходящ.*

Доказателство. Нека $\{A_n\}$ са независими събития и е изпълнено $\mathbf{P}(\lim_{n\to\infty} \sup A_n) = 0$. Ако допуснем, че $\sum_{n=1}^{\infty} \mathbf{P}(A_m) = \infty$, по лема 1 ще получим $\mathbf{P}(\lim_{n\to\infty} \sup A_n) = 1$. Полученото противоречие показва, че $\sum_{n=1}^{\infty} \mathbf{P}(A_n) < \infty$. Аналогично се разсъждава в случая $\mathbf{P}(\lim_{n\to\infty} \sup A_n) = 1$.

Теорема 1. *Нека $S_n = \xi_1 + \cdots + \xi_n$, където $\{\xi_k\}$ са независими случайни величини с крайна дисперсия $\mathbf{D}\xi_k < \infty$, $k = 1, 2, \ldots, n$. Тогава за всяко $\varepsilon > 0$ е в сила неравенството на Колмогоров*

$$\mathbf{P}\left\{\max_{1 \leqq k \leqq n} |S_k - \mathbf{E}S_k| \geqq \varepsilon\right\} \leqq \frac{\mathbf{D}S_n}{\varepsilon^2}. \tag{3}$$

Доказателство. Въвеждаме центрираните случайни величини $\eta_k = \xi_k - \mathbf{E}\xi_k$, за които очевидно $\mathbf{E}\eta_k = 0$ и $\mathbf{D}\eta_k = \mathbf{E}\eta_k^2 = \mathbf{D}\xi_k$. Да означим $\zeta_n = \eta_1 + \cdots + \eta_n$. Тогава $\mathbf{E}\zeta_n = 0$ и $\mathbf{D}\zeta_n = \mathbf{E}\zeta_n^2$, а (3) се записва в следния вид:

$$\mathbf{P}\{\max_{1 \leqq k \leqq n} |\zeta_k| \geqq \varepsilon\} \leqq \frac{\mathbf{E}\zeta_n^2}{\varepsilon^2}. \tag{4}$$

Определяме случайна величина $\nu = \min\{k : |\zeta_k| \geqq \varepsilon\}$, $k = 1, \ldots, n$, и $\nu = n + 1$, ако $\max_{1 \leqq k \leqq n} |\zeta_k| < \varepsilon$. Следователно $\zeta_n^2 = \zeta_n^2 \sum_{k=1}^{\infty} I_{\{\nu=k\}}$ и

$$\begin{aligned}
\mathbf{E}\zeta_n^2 &\geqq \sum_{k=1}^{n} \mathbf{E}(\zeta_n^2 I_{\{\nu=k\}}) \\
&= \sum_{k=1}^{n} \mathbf{E}[(\eta_1 + \cdots + \eta_k)^2 I_{\{\nu=k\}}] + \sum_{k=1}^{n} \mathbf{E}[(\eta_{k+1} + \cdots + \eta_n)^2 I_{\{\nu=k\}}] \qquad (5) \\
&+ 2\sum_{k=1}^{n} \mathbf{E}\{(\eta_1 + \cdots + \eta_k) I_{\{\nu=k\}} (\eta_{k+1} + \cdots + \eta_n)\}.
\end{aligned}$$

Случайната величина $I_{\{\nu=k\}}$ зависи само от случайните величини $\eta_1, \ldots, \eta_k$. Следователно случайната величина $(\eta_1 + \cdots + \eta_n)I_{\{\nu=k\}}$ не зависи от случайните величини $\eta_{k+1}, \ldots, \eta_n$, откъдето намираме

$$\begin{aligned}&\mathbf{E}[(\eta_1 + \cdots + \eta_k)I_{\{\nu=k\}}(\eta_{k+1} + \cdots + \eta_n)]\\ &= \mathbf{E}\{(\eta_1 + \cdots + \eta_k)I_{\{\nu=k\}}\}\,\mathbf{E}(\eta_{k+1} + \cdots + \eta_n) = 0.\end{aligned}$$

Тогава от (5) получаваме

$$\mathbf{E}\zeta_n^2 \geqq \sum_{k=1}^{n} \mathbf{E}\zeta_k^2 I_{\{\nu=k\}}. \tag{6}$$

Ако $\omega \in \{\nu = k\}$, то $|\zeta_k| \geqq \varepsilon$, $k = 1, \ldots, n$. От друга страна,

$$\mathbf{P}\{\nu \leqq n\} = \sum_{k=1}^{n} \mathbf{P}\{\nu = k\} = \mathbf{P}\{\max_{1 \leqq k \leqq n} |\zeta_k| \geqq \varepsilon\}.$$

Като вземем предвид това, от (6) получаваме

$$\mathbf{E}\zeta_n^2 \geqq \varepsilon^2 \sum_{k=1}^{n} \mathbf{E}I_{\{\nu=k\}} = \varepsilon^2 \sum_{k=1}^{n} \mathbf{P}\{\nu = k\} = \varepsilon^2 \mathbf{P}\left\{\max_{1 \leqq k \leqq n} |\zeta_k| \geqq \varepsilon\right\},$$

което доказва неравенствата (4) и (3).

Да отбележим, че $\mathbf{D}S_n = \sum_{k=1}^{n} \mathbf{D}\xi_k$ в условията на теорема 1.

Неравенството на Колмогоров обобщава неравенството на Чебишов (4.5.13), тъй като последното се получава от (3) при $n = 1$.

§ 3. Усилени закони за големите числа

В този параграф ще докажем две теореми на Колмогоров, които имат фундаментално значение за науката и практиката.

Теорема 1. *Ако $\{\xi_n\}$ са независими случайни величини, за които*

$$\sum_{n=1}^{\infty} \frac{\mathbf{D}\xi_n}{n^2} < \infty, \tag{1}$$

то е в сила УЗГЧ, *т. е.*

$$\zeta_n = \frac{(S_n - \mathbf{E}S_n)}{n} \xrightarrow{\text{п. с.}} 0 \quad \textit{при} \quad n \to \infty, \tag{2}$$

където $S_n = \xi_1 + \cdots + \xi_n$.

Доказателство. Както знаем от теорема 7.1.1, граничното съотношение (2) е еквивалентно на следното:

$$(3)\quad \mathbf{P}\left\{\sup_{k\geqq n}\left|\frac{S_k}{k}-\frac{\mathbf{E}S_k}{k}\right|\geqq\varepsilon\right\}\longrightarrow 0 \quad \text{при}\ n\to\infty\ \text{за всяко}\ \varepsilon>0.$$

Да въведем събитията

$$A_j=\left\{\max_{2^{j-1}\leqq k\leqq 2^j}\left|\frac{S_k}{k}-\frac{\mathbf{E}S_k}{k}\right|\geqq\varepsilon\right\},\quad j=1,\,2,\,\ldots$$

Тогава (3) е еквивалентно на условието

$$(4)\qquad \mathbf{P}\left(\bigcup_{j=m}^{\infty}A_j\right)\longrightarrow 0\quad\text{при}\quad m\to\infty,$$

за което, от своя страна, е достатъчно да покажем, че

$$(5)\qquad \sum_{j=1}^{\infty}\mathbf{P}(A_j)<\infty.$$

Да отбележим първо, че

$$\mathbf{P}(A_j)\leqq\mathbf{P}\left\{\max_{2^{j-1}\leqq k\leqq 2^j}|S_k-\mathbf{E}S_k|\geqq\varepsilon 2^{j-1}\right\},$$

откъдето по неравенството на Колмогоров (2.3) получаваме

$$\mathbf{P}(A_j)\leqq\frac{1}{\varepsilon^2 2^{2(j-1)}}\sum_{2^{j-1}\leqq k\leqq 2^j}\mathbf{D}\xi_k\leqq\frac{4}{\varepsilon^2 2^{2j}}\sum_{k<2^j}\mathbf{D}\xi_k.$$

Оттук следва, че

$$\sum_{j=1}^{\infty}\mathbf{P}(A_j)\leqq\frac{4}{\varepsilon^2}\sum_{j=1}^{\infty}\frac{1}{2^{2j}}\sum_{k<2^j}\mathbf{D}\xi_k=\frac{1}{\varepsilon^2}\sum_{k=1}^{\infty}\mathbf{D}\xi_k\sum_{j>\log_2 k}2^{-2j}\leqq\frac{16}{3\varepsilon^2}\sum_{k=1}^{\infty}\frac{\mathbf{D}\xi_k}{k^2}<\infty.$$

Извършената смяна в реда на сумирането е напълно законна, тъй като съответните редове са абсолютно сходящи и

$$\{1\leqq k\leqq\infty,\ \log_2 k<j\leqq\infty\}=\{1\leqq j\leqq\infty,\ 1\leqq k<2^j\}.$$

Освен това $\sum\limits_{j>\log_2 k}2^{-2j}\leqq 2^{-2\log_2 k}\sum\limits_{m=0}^{\infty}2^{-2m}=\dfrac{4}{3k^2}.$

Теоремата е доказана.

Следствие 1. *Ако случайните величини* $\{\xi_n\}$ *са независими и имат равномерно ограничени дисперсии, то е в сила* УЗГЧ.

Очевидно от $\mathbf{D}\xi_n \leqq C$, $n = 1, 2, \ldots$, следва веднага (1).

От следствие 1 очевидно следват следствия 1.1 и 1.2.

Лема 1. *Ако* ξ *е случайна величина и*

$$A_n = \{\omega : |\xi(\omega)| \geqq n\}, \quad n = 0, 1, 2, \ldots, \quad A_0 = \Omega,$$

то $\mathbf{E}|\xi| < \infty$ *тогава и само тогава, когато* $\sum\limits_{n=0}^{\infty} \mathbf{P}(A_n) < \infty$.

Доказателство. Нека $\mathbf{E}|\xi| < \infty$. Редицата $\{A_n\}$ е монотонно нерастяща, т. е. $A_n \supseteq A_{n+1}$ откъдето за всяко $n \geqq 1$ получаваме

$$\begin{aligned}\sum_{k=1}^{n} \mathbf{P}(A_n) &= \sum_{k=1}^{n}(k-1)[\mathbf{P}(A_{k-1}) - \mathbf{P}(A_k)] + n\mathbf{P}(A_n) \\ &= \sum_{k=1}^{n}(k-1) \int\limits_{A_{k-1}\setminus A_k} d\mathbf{P} + n \int\limits_{A_n} d\mathbf{P}.\end{aligned} \tag{6}$$

Тъй като $\omega \in A_{k-1} \setminus A_k$ точно тогава, когато $k - 1 \leqq |\xi(\omega)| < k$, то от (6) намираме

$$\sum_{k=1}^{n} \mathbf{P}(A_k) \leqq \sum_{k=1}^{n} \int\limits_{A_{k-1}\setminus A_k} |\xi| d\mathbf{P} + \int\limits_{A_n} |\xi| d\mathbf{P} = \int\limits_{\Omega} |\xi| d\mathbf{P} = \mathbf{E}|\xi| < \infty,$$

понеже $\sum\limits_{k=1}^{n} (A_{k-1} \setminus A_k) + A_n = \Omega$.

Нека сега $\sum\limits_{k=1}^{\infty} \mathbf{P}(A_n) < \infty$. Тогава за всяко n имаме

$$\begin{aligned}\sum_{k=1}^{n} \int\limits_{A_{k-1}\setminus A_k} |\xi| d\mathbf{P} &\leqq \sum_{k=1}^{n} k \int\limits_{A_{k-1}\setminus A_k} d\mathbf{P} = \sum_{k=1}^{n} k[\mathbf{P}(A_{k-1}) - \mathbf{P}(A_k)] \\ &= \sum_{k=0}^{n-1} \mathbf{P}(A_k) - n\mathbf{P}(A_n) \leqq \sum_{k=0}^{n} \mathbf{P}(A_k).\end{aligned}$$

Следователно

$$\mathbf{E}|\xi| = \sum_{k=1}^{\infty} \int\limits_{A_{k-1}\setminus A_k} |\xi| d\mathbf{P} = \lim_{n\to\infty} \sum_{k=1}^{n} \int\limits_{A_{k-1}\setminus A_k} |\xi| d\mathbf{P} \leqq \sum_{k=0}^{\infty} \mathbf{P}(A_k) < \infty$$

с което лемата е напълно доказана.

Лема 2. *Нека $\xi_1, \xi_2, \dots$ са независими и еднакво разпределени случайни величини с крайно математическо очакване, т. е. $\mathbf{E}|\xi_1| < \infty$. Да положим*

$$\eta_n = \begin{cases} \xi_n & \text{при } |\xi_n| \leqq n, \\ 0 & \text{при } |\xi_n| > n, \end{cases} \tag{7}$$

$n = 1, 2, \dots$ *Тогава*

$$\sum_{n=1}^{\infty} \frac{\mathbf{D}\eta_n}{n^2} < \infty, \tag{8}$$

т. е. за „отрязаните“ случайни величини $\{\eta_n\}$ е в сила УЗГЧ.

Доказателство. Да въведем събитията

$$B_n = \{|\xi_n| > n\}, \quad A_n = \{|\xi_1| > n\}.$$

Тогава случайните величини $\eta_n = \xi_n I_{\overline{B}_n}$ и $\zeta_n = \xi_1 I_{\overline{A}_n}$ са еднакво разпределени. Следователно $\mathbf{D}\eta_k \leqq \mathbf{E}\eta_k^2 = \mathbf{E}\xi_k^2 I_{\overline{B}_k} = \mathbf{E}\xi_1^2 I_{\overline{A}_k}$, откъдето за всяко $n \geqq 1$

$$\sum_{k=1}^{n} \frac{\mathbf{D}\eta_k}{k^2} \leqq \sum_{k=1}^{n} \frac{\mathbf{E}\xi_1^2 I_{\overline{A}_k}}{k^2} = \mathbf{E}\left(\xi_1^2 \sum_{k=1}^{n} \frac{I_{\overline{A}_k}}{k^2}\right). \tag{9}$$

Да положим $A_0 = \Omega$ и да отбележим, че $\{A_n\}$ монотонно намалява, т. е. $A_n \supset A_{n+1}$, $n = 0, 1, \dots$ Тогава $\Omega = \sum\limits_{k=1}^{\infty} (A_{k-1} \setminus A_k)$ и

$$\mathbf{E}\left(\xi_1^2 \sum_{k=1}^{n} \frac{I_{\overline{A}_k}}{k^2}\right) = \sum_{j=1}^{\infty} \int\limits_{A_{j-1} \setminus A_j} \left(\xi_1^2 \sum_{k=1}^{n} \frac{I_{\overline{A}_k}}{k^2}\right) d\mathbf{P}. \tag{10}$$

Да отбележим, че

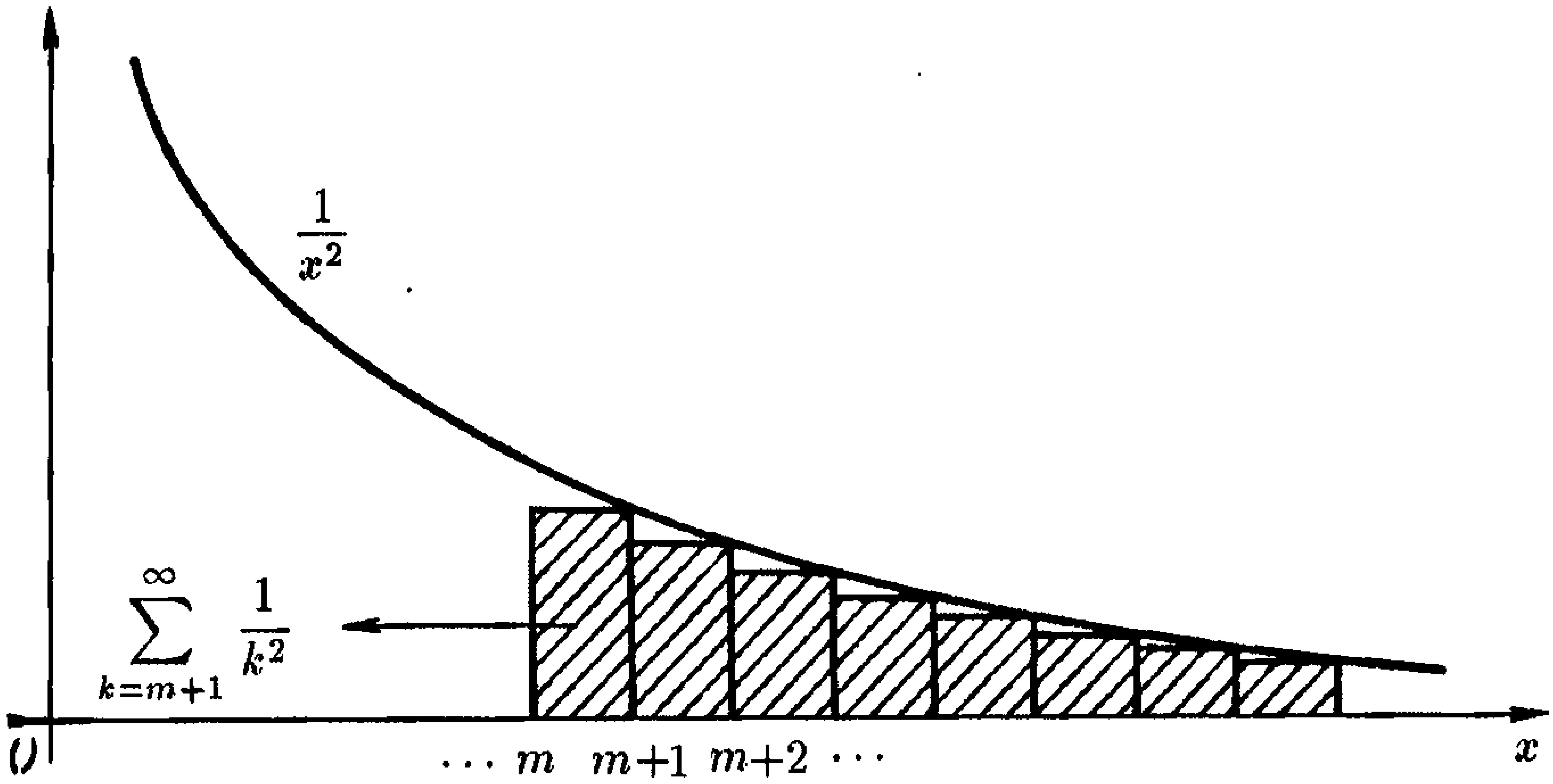
$$I_{\overline{A}_k} = \begin{cases} 0 & \text{при } |\xi_1| > k, \\ 1 & \text{при } |\xi_1| \leqq k. \end{cases}$$

От друга страна, за всяко $\omega_0 \in \Omega$ съществува $m \geqq 1$, така че

$$\omega_0 \in (A_{m-1} \setminus A_m) = \{\omega : m - 1 \leqq |\xi_1(\omega)| \leqq m\}.$$

Тогава при $n \geqq m$ следва, че $I_{\overline{A}_k}(\omega_0) = 0$ за $k = 1, 2, \dots, m-1$ и $I_{\overline{A}_k}(\omega_0) = 1$ за $k = m, m+1, \dots, n$. Следователно (вж. фиг. 15)

$$\sum_{k=1}^{n} \frac{I_{\overline{A}_k}(\omega_0)}{k^2} = \sum_{k=m}^{n} \frac{1}{k^2} \leqq \sum_{k=m}^{\infty} \frac{1}{k^2} \leqq \frac{1}{m^2} + \int\limits_{m}^{\infty} \frac{dx}{x^2} = \frac{1}{m^2} + \frac{1}{m},$$



Фиг. 15

откъдето получаваме оценката

$$\xi_1^2(\omega_0) \sum_{k=1}^{n} \frac{I_{\overline{A}_k}(\omega_0)}{k^2} \leqq \frac{\xi_1^2(\omega_0)}{m^2} + \frac{\xi_1^2(\omega_0)}{m} \leqq 1 + |\xi_1(\omega_0)|.$$

Оттук и от (9) и (10) следва, че за всяко $n \geqq 1$

$$\sum_{k=1}^{n} \frac{\mathbf{D}\eta_k}{k^2} \leqq 1 + \mathbf{E}|\xi_1| < \infty,$$

което доказва (8).

Теорема 2. *За една редица* $\{\xi_n\}$ *от независими и еднакво разпределени случайни величини необходимо и достатъчно условие за* УЗГЧ *е съществуването на крайно математическо очакване, т. е. ако* $a = \mathbf{E}\xi_1$ *е крайно число, то*

$$\frac{1}{n} \sum_{i=1}^{n} \xi_i \xrightarrow{\text{п. с.}} a \quad \text{при} \quad n \to \infty; \tag{11}$$

обратно, от (11) *следва, че случайните величини* $\{\xi_k\}$ *имат крайно математическо очакване и* $a = \mathbf{E}\xi_1$.

Доказателство. *Необходимост.* Нека е в сила (11) и да означим $S_n = \sum_{k=1}^{n} \xi_k$. Тогава

$$(12) \qquad \frac{\xi_n}{n} = \frac{S_n - S_{n-1}}{n} = \frac{S_n}{n} - \frac{n-1}{n}\frac{S_{n-1}}{n-1} \xrightarrow{\text{п. с.}} 0 \quad \text{при} \quad n \to \infty.$$

Да въведем събитията $B_n = \{|\xi_n| > n\} = \left\{\frac{|\xi_n|}{n} > 1\right\}$, $n = 1, 2, \dots$ Тогава от (12) следва, че $\mathbf{P}(\lim_{n\to\infty} \sup B_n) = 0$. Тъй като събитията $\{B_n\}$ са независими, то по следствие 1 от лемата на Борел – Кантели следва, че

$$(13) \qquad \sum_{n=1}^{\infty} \mathbf{P}(B_n) < \infty.$$

Да въведем събитията $A_n = \{|\xi_1| > n\}$, $n = 1, 2, \dots$ Тъй като $\{\xi_n\}$ са еднакво разпределени случайни величини, то $\mathbf{P}(A_n) = \mathbf{P}(B_n)$, т. е. от (13) получаваме

$$(14) \qquad \sum_{n=1}^{\infty} \mathbf{P}(A_n) < \infty.$$

Съгласно лема 1 от (14) следва, че $\mathbf{E}|\xi_1| < \infty$.

Достатъчност. Нека $\mathbf{E}|\xi_1| < \infty$. Ще покажем, че е в сила (11), където $a = \mathbf{E}\xi_1$. Полагаме $\eta_n = \xi_n I_{\overline{B}_n}$. От условието $\mathbf{E}|\xi_1| < \infty$ по лема 1 следва (14). Тъй като $\mathbf{P}(A_n) = \mathbf{P}(B_n)$, то в сила е (13). От (13) по лема 2.1 на Борел – Кантели следва, че $\mathbf{P}(\lim_{n\to\infty} \sup B_n) = 0$. Но $B_n = \{\eta_n \neq \xi_n\}$ и следователно $\mathbf{P}(\lim_{n\to\infty} \sup\{\eta_n \neq \xi_n\}) = 0$, т. е. $\mathbf{P}(\lim_{n\to\infty} \inf\{\eta_n = \xi_n\}) = 1$. С други думи, събитието $\{\eta_n \neq \xi_n\}$ се осъществява само за краен брой индекси. Следователно

$$(15) \qquad \frac{1}{n}\sum_{k=1}^{n}(\xi_k - \eta_k) \xrightarrow{\text{п. с.}} 0 \quad \text{при } n \to \infty.$$

От друга страна, съгласно лема 2, за отрязаните случайни величини $\{\eta_n\}$ е в сила УЗГЧ, т. е.

$$(16) \qquad \frac{1}{n}\sum_{k=1}^{n}(\eta_k - \mathbf{E}\eta_k) \xrightarrow{\text{п. с.}} 0 \quad \text{при} \quad n \to \infty.$$

Тъй като $\mathbf{E}\eta_n = \mathbf{E}\xi_n I_{\overline{B}_n} = \mathbf{E}\xi_n I_{\overline{A}_n} = \mathbf{E}\xi_1 I_{\overline{A}_n}$ и $\overline{A}_n \uparrow \Omega$, то

$$\lim_{n\to\infty} \mathbf{E}\eta_n = \lim_{n\to\infty} \int_{\overline{A}_n} \xi_1 d\mathbf{P} = \int_{\Omega} \xi_1 d\mathbf{P} = \mathbf{E}\xi_1 = a.$$

Следователно (вж. лема 3 по-долу)

$$\frac{1}{n}\sum_{k=1}^{n} \mathbf{E}\eta_k \longrightarrow a \quad \text{при} \quad n \to \infty. \tag{17}$$

От (16) и (17) следва, че

$$\frac{1}{n}\sum_{k=1}^{n} \eta_n \xrightarrow{\text{п. с.}} a \quad \text{при} \quad n \to \infty. \tag{18}$$

Граничните съотношения (15) и (18) доказват (11) с $a = \mathbf{E}\xi_1$, с което достатъчността е доказана.

Да се върнем на необходимостта. Показахме, че от (11) следва $\mathbf{E}|\xi_1| < \infty$. Но според току-що доказаната достатъчност константата a в (11) трябва да бъде равна на $\mathbf{E}\xi_1$.

Теоремата е доказана.

Следващият факт би трябвало да е известен от анализа, но за всеки случай привеждаме доказателството му.

Лема 3. *Нека $x_n \in \mathbf{R}^1$ и $x_n \to a$ при $n \to \infty$. Тогава $\frac{1}{n}\sum_{k=1}^{n} x_k \to a$ при $n \to \infty$.*

Доказателство. За всяко $\varepsilon > 0$ съществува $N < \infty$, така че $|x_k - a| < \frac{\varepsilon}{2}$ при $k > N$, а $|x_k - a| \leqq M$ при $k \leqq n$. Очевидно съществува n_0, така че при $n > n_0$ да имаме $\frac{MN}{n} < \frac{\varepsilon}{2}$. Тогава

$$\left|\frac{1}{n}\sum_{k=1}^{n} x_k - a\right| \leqq \frac{1}{n}\sum_{k=1}^{N} |x_k - a| + \frac{1}{n}\sum_{k=N+1}^{n} |x_k - a|$$

$$< \frac{MN}{n} + \frac{\varepsilon}{2}\,\frac{n-N}{n} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \quad \text{при} \quad n > n_1 = \max(n_0, N).$$

Очевидно следствие от теорема 2 е следният резултат, който от своя страна е обобщение на теоремата на Бернули.

Теорема 3 (теорема на Борел). *Ако* $\nu_n \in Bi(n,p)$, *то при* $n \to \infty$ *е изпълнено* $\nu_n/n \xrightarrow{\text{п. с.}} p$.

Накрая да отбележим, че доказаните в тази глава ЗГЧ имат широко приложение в математическата статистика, което ще бъде показано във втората част на книгата.

Задачи

1. Случайните величини $\{\xi_k\}$, $n = 1, 2, \ldots$, са независими и еднакво разпределени. Докажете, че с вероятност 1 ще се сбъднат само краен брой събития $A_n = \{|\xi_n| \geqq \sqrt{n}\}$ тогава и само тогава, когато $\mathbf{D}\xi_1 < \infty$.

2. Случайните величини $\xi_1, \xi_2, \ldots$ са независими и еднакво разпределени с $\mathbf{E}|\xi_1| < \infty$. Независимите от тях случайни величини $\{\theta_n\}$ са независими в съвкупност и удовлетворяват условията $|\theta_n| \leqq 1$ и $\mathbf{E}\theta_n = 0$, $n = 1, 2, \ldots$ Да се докаже, че $\frac{1}{n}\sum_{k=1}^{n}\theta_k\xi_k \xrightarrow{\text{п. с.}} 0$ при $n \to \infty$.

3. Нека $\{\xi_n\}$ са независими случайни величини и $\xi_n \in \mathcal{N}(a_n, \sqrt{n})$, където $n = 1, 2, \ldots$ Докажете:

а) че е в сила УЗГЧ;

б) СЗГЧ без условието за независимост;

Изследвайте средното аритметично при $a_n \to a$.

4. Нека $\{\eta_n\}$ са независими случайни величини и $\eta_n \in Po(\lambda_n)$. Ако $(\lambda_1 + \cdots + \lambda_n)/n \to 0$, покажете, че редицата $\{\eta_n\}$ удовлетворява УЗГЧ.

5. Нека ξ_n е редица от независими случайни величини, за които

$$\mathbf{P}\{\xi_n = \sqrt{n}\} = \mathbf{P}\{\xi_n = -\sqrt{n}\} = \frac{1}{2}, \quad S_n = \frac{(\xi_1 + \cdots + \xi_n)}{n}, \ n = 1, 2, \ldots$$

Покажете, че $\mathbf{E}S_n = 0$ за всяко n, но въпреки това СЗГЧ не е в сила, т. е. S_n не клони по вероятност към нула.

Упътване. Използвайте характеристични функции, за да покажете, че S_n не клони по разпределение към нула.

6. Нека $\{\xi_n\}$ са независими и еднакво разпределени случайни величини и $\xi_n \in Ex(a^{-n/2})$. Докажете, че при $0 < a < 1$ е в сила УЗГЧ.

7. Нека $\{\xi_n\}$ са независими и еднакво разпределени случайни величини с $\mathbf{E}\xi_1 = 0$ и $\mathbf{D}\xi_1 = 1$, а $\eta_n = \alpha + \beta^n\xi_n$, $\zeta_n = \alpha + n^\beta\xi_n$ за $\alpha \in \mathbf{R}^1$. Тогава:

а) при $-1 < \beta < 1$ за $\{\eta_n\}$ е в сила УЗГЧ;

б) при $\beta < \frac{1}{2}$ за $\{\zeta_n\}$ е в сила УЗГЧ.

8. Нека $\{\xi_n\}$ са независими случайни величини с разпределение

$$\mathbf{P}\{\xi_n = -n^\beta\} = \mathbf{P}\{\xi_n = n^\beta\} = \frac{1}{2n^\alpha}, \quad \mathbf{P}\{\xi_n = 0\} = 1 - \frac{1}{2n^\alpha},$$

където α, $\beta > 0$. Покажете, че при $\alpha > 2\beta - 1$ и $\beta > \frac{1}{2}$ е в сила УЗГЧ.

Девета глава

Централна гранична теорема

§ 1. Централна гранична теорема за независими и еднакво разпределени случайни величини

Както подчертахме в § 3.4, интегралната теорема на Моавър – Лаплас (теорема 3.4.2) е частен случай на централната гранична теорема, т. е. съгласно (3.4.16)

$$(1) \qquad \zeta_n = \frac{\nu_n - \mathbf{E}\nu_n}{\sqrt{\mathbf{D}\nu_n}} \xrightarrow{d} \zeta \in \mathcal{N}(0,1) \quad \text{при} \quad n \to \infty,$$

където $\nu_n = \xi_1 + \cdots + \xi_n$, $\xi_k \in Bi(1,p)$ и $\{\xi_k\}$ са независими.

Сега ще изследваме сходимостта по разпределение на редици от тип (1) в общия случай, когато случайните величини ξ_k имат произволно разпределение. Отначало ще изследваме случая, когато $\{\xi_k\}$ са еднакво разпределени.

Теорема 1. *Нека $\{\xi_k\}$ са независими и еднакво разпределени случайни величини с $\mathbf{E}\xi_k = a$ и $\mathbf{D}\xi_k = \sigma^2 < \infty$. Тогава за всяко $x \in \mathbf{R}^1$*

$$(2) \qquad \lim_{n\to\infty} \mathbf{P}\left\{ \frac{\sum\limits_{k=1}^{n} \xi_k - na}{\sigma\sqrt{n}} < x \right\} = \frac{1}{2\pi} \int\limits_{-\infty}^{x} e^{-t^2/2} dt.$$

Доказателство. Очевидно записването (2) е еквивалентно на (1) и може да бъде представено в следните еквивалентни форми:

$$(3) \quad \zeta_n = \frac{\sum\limits_{k=1}^{n} \xi_k - na}{\sigma\sqrt{n}} = \frac{S_n - \mathbf{E}S_n}{\sqrt{\mathbf{D}S_n}} = \frac{\frac{1}{n}\sum\limits_{k=1}^{n}\xi_k - a}{\frac{\sigma}{\sqrt{n}}} = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\frac{\xi_k - a}{\sigma} = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\eta_k,$$

където $S_n = \sum\limits_{k=1}^{n} \xi_k$, а $\eta_k = \dfrac{\xi_k - a}{\sigma}$.

Да отбележим, че случайните величини $\{\eta_k\}$ са независими и са получени от $\{\xi_k\}$ чрез „центриране“ и „нормиране“, т. е. както непосредствено се проверява $\mathbf{E}\eta_k = 0$ и $\mathbf{D}\eta_k = 1$.

И така, съгласно (3) граничното съотношение (2) се записва във вида

$$(4) \qquad \zeta_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \eta_k \xrightarrow{d} \zeta \in \mathcal{N}(0,1) \quad \text{при} \quad n \to \infty.$$

Да въведем характеристичните функции

$$\Psi(t) = \mathbf{E}e^{it\eta_k}, \qquad \Psi_n(t) = \mathbf{E}e^{it\zeta_n}.$$

Съгласно (6.3.11) при $z \to \infty$ имаме

$$(5) \qquad \Psi(z) = 1 - \frac{z^2}{2} + o(z^2).$$

От друга страна, поради мултипликативното свойство

$$(6) \qquad \Psi_n(t) = \mathbf{E}\exp\left\{i\frac{t}{\sqrt{n}} \sum_{k=1}^{n} \eta_k\right\} = \left[\Psi\left(\frac{t}{\sqrt{n}}\right)\right]^n.$$

От (5) и (6) получаваме

$$(7) \qquad \Psi_n(t) = \left[1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right]^n \longrightarrow e^{-t^2/2} \quad \text{при} \quad n \to \infty.$$

Както показахме в пример 6.3.2, изпълнено е $e^{-t^2/2} = \mathbf{E}e^{it\zeta}$, където $\zeta \in \mathcal{N}(0,1)$.

Тогава съгласно теоремата за непрекъснатост на характеристични функции (теорема 7.2.4) граничното съотношение (7) е еквивалентно на (4), т. е. на (2).

Забележка. Имайки предвид (3), съотношението (2) понякога се записва във вида $S_n \sim \mathcal{N}(an, \sigma^2 n)$ или $\dfrac{S_n}{n} \in \mathcal{N}\left(a, \dfrac{\sigma^2}{n}\right)$, като се казва, че S_n е разпределена асимптотично нормално с параметри an и $\sigma^2 n$, аналогично за $\dfrac{S_n}{n}$.

§ 2. Теореми на Линдеберг – Фелър и Ляпунов

Преди да пристъпим към доказателството на Централна Гранична Теорема (ЦГТ) за нееднакво разпределени случайни величини, ще приведем следния спомагателен резултат:

Лема 1. *При* $|\alpha| \leqq \frac{1}{2}$ *имаме*

$$|\log(1+\alpha) - \alpha| \leqq |\alpha|^2. \tag{1}$$

Доказателство. Непосредствено от развитието на логаритъма в ред при $|\alpha| \leqq \frac{1}{2}$ получаваме

$$|\log(1+\alpha) - \alpha| = \left|\sum_{k=2}^{\infty} \frac{|\alpha|^k}{k}\right| \leqq \frac{|\alpha|^2}{2} \sum_{j=0}^{\infty} |\alpha|^j = \frac{|\alpha|^2}{2} \frac{1}{1-|\alpha|} \leqq |\alpha|^2.$$

Теорема 1 (теорема на Линдеберг – Фелър). *Нека* $\xi_1, \xi_2, \ldots$ *са независими случайни величини с функции на разпределение*

$$F_k(x) = \mathbf{P}\{\xi_k < x\}, \quad a_k = \mathbf{E}\xi_k, \quad \sigma_k^2 = \mathbf{D}\xi_k < \infty, \quad k = 1, 2, \ldots$$

Да означим

$$S_n = \sum_{k=1}^{n} \xi_k, \; B_n^2 = \mathbf{D}S_n = \sum_{k=1}^{n} \sigma_k^2.$$

Ако за всяко $\tau > 0$ *е изпълнено*

$$L_n(\tau) = \frac{1}{B_n^2} \sum_{k=1}^{n} \int\limits_{\{|x-a_k|>\tau B_n\}} (x-a_k)^2 dF_k(x) \longrightarrow 0 \quad \text{при } n \to \infty, \tag{2}$$

то в сила е ЦГТ, *т. е.*

$$\zeta_n = \frac{S_n - \mathbf{E}S_n}{\sqrt{\mathbf{D}S_n}} \xrightarrow{d} \zeta \in \mathcal{N}(0,1) \quad \text{при } n \to \infty. \tag{3}$$

Доказателство. Очевидно (3) може да бъде записано в следните еквивалентни форми:

$$\zeta_n = \sum_{k=1}^{n} \frac{\xi_k - a_k}{B_n} = \sum_{k=1}^{n} \eta_{k,n}, \tag{4}$$

където независимите случайни величини $\eta_{k,n} = \dfrac{\xi_k - a_k}{B_n}$ имат функции на разпределение

$$F_{k,n}(x) = \mathbf{P}\{\eta_{k,n} < x\} = F_k(xB_n + a_k)$$

и съответни характеристики

$$\mathbf{E}\eta_{k,n} = 0, \quad \mathbf{D}\eta_{k,n} = \frac{\sigma_k^2}{B_n^2}, \quad \sum_{k=1}^{n} \mathbf{D}\eta_{k,n} = 1. \tag{5}$$

В този случай от (2) получаваме

$$\begin{aligned} L_n(\tau) &= \sum_{k=1}^{n} \int\limits_{\{|(\xi_k - a_k)/B_n| > \tau\}} \left[\frac{\xi_k - a_k}{B_n}\right]^2 d\mathbf{P} = \sum_{k=1}^{n} \int\limits_{\{|\eta_{k,n}| > \tau\}} \eta_{k,n}^2 d\mathbf{P} \\ &= \sum_{k=1}^{n} \mathbf{E}(\eta_{k,n}^2 I_{\{|\eta_{k,n}| > \tau\}}) = \sum_{k=1}^{n} \int\limits_{\{|x| > \tau\}} x^2 dF_{k,n}(x). \end{aligned} \tag{6}$$

Съгласно теоремата за непрекъснатост на характеристичните функции (теорема 7.2.4) съотношението (3) е еквивалентно на следното:

$$\Psi_n(t) = \mathbf{E}e^{it\zeta_n} \longrightarrow e^{-t^2/2} \quad \text{при} \quad n \to \infty, \tag{7}$$

тъй като съгласно (6.3.17) е изпълнено $\mathbf{E}e^{it\zeta} = e^{-t^2/2}$.

От мултипликативното свойство на характеристичните функции (вж. § 6.3) имаме

$$\Psi_n(t) = \prod_{k=1}^{n} \Psi_{k,n}(t), \tag{8}$$

където $\Psi_{k,n}(t) = \mathbf{E}e^{it\eta_{k,n}}$. Тогава от (7) следва представянето

$$\log \Psi_n(t) = \sum_{k=1}^{n} (\Psi_{k,n}(t) - 1) + U_n(t), \tag{9}$$

където

$$U_n(t) = \sum_{k=1}^{n} \log \Psi_{k,n}(t) - \sum_{k=1}^{n} (\Psi_{k,n}(t) - 1). \tag{10}$$

Като използваме лема 6.3.1, получаваме

$$|\Psi_{k,n}(t) - 1| = |\mathbf{E}(e^{it\eta_{k,n}} - 1 - it\eta_{k,n})| \leqq \mathbf{E}\left\{\frac{|t\eta_{k,n}|^2}{2}\right\} = \frac{t^2}{2}\mathbf{D}\eta_{k,n}. \tag{11}$$

От друга страна, съгласно (6), получаваме оценката

$$\max_{1 \leqq k \leqq n} \mathbf{D}\eta_{k,n} = \max_{1 \leqq k \leqq n} \mathbf{E}\eta_{k,n}^2$$

$$= \max_{1 \leqq k \leqq n} \left(\int_{\{|\eta_{k,n}| \leqq \tau\}} \eta_{k,n}^2 d\mathbf{P} + \int_{\{|\eta_{k,n}| > \tau\}} \eta_{k,n}^2 d\mathbf{P} \right)$$

$$\leqq \tau^2 + \sum_{k=1}^{n} \int_{\{|\eta_{k,n}| > \tau\}} \eta_{k,n}^2 d\mathbf{P} = \tau^2 + L_n(\tau).$$

Тъй като за всяко $\tau > 0$ е изпълнено $\lim_{n\to\infty} L_n(\tau) = 0$, от горното неравенство следва

(12) $$\max_{1 \leqq k \leqq n} \mathbf{D}\eta_{k,n} \longrightarrow 0 \quad \text{при} \quad n \to \infty.$$

Тогава от (11) и (12) получаваме

(13) $$\max_{1 \leqq k \leqq n} |\Psi_{k,n}(t) - 1| \leqq \frac{t^2}{2} \max_{1 \leqq k \leqq n} \mathbf{D}\eta_{k,n} \longrightarrow 0 \quad \text{при} \quad n \to \infty.$$

Следователно при достатъчно голямо n за всички $k \leqq n$ имаме $|\Psi_{k,n}(t) - 1| < \frac{1}{2}$, откъдето съгласно лема 1 получаваме

(14) $$|\log \Psi_{k,n}(t) - (\Psi_{k,n}(t) - 1)| \leqq |\Psi_{k,n}(t) - 1|^2.$$

Сега от (10) и (14) при достатъчно големи n намираме оценката

(15) $$|U_n(t)| \leqq \sum_{k=1}^{n} |\Psi_{k,n}(t) - 1|^2 \leqq \max_{1 \leqq k \leqq n} |\Psi_{k,n}(t) - 1| \sum_{k=1}^{n} |\Psi_{k,n}(t) - 1|.$$

Тъй като от (11) и (5) имаме

$$\sum_{k=1}^{n} |\Psi_{k,n}(t) - 1| \leqq \frac{t^2}{2} \sum_{k=1}^{n} \mathbf{D}\xi_{k,n} = \frac{t^2}{2},$$

от (15) и (13) следва, че за всяко $t \in \mathbf{R}^1$ е изпълнено

(16) $$U_n(t) \longrightarrow 0 \quad \text{при} \quad n \to \infty.$$

Сега ще изследваме поведението на главния член в разлагането (9), който ще представим в следния вид:

$$
\begin{aligned}
\sum_{k=1}^{n}(\Psi_{k,n}(t)-1) &= \sum_{k=1}^{n}\mathbf{E}(e^{it\eta_{k,n}}-1-it\eta_{k,n}) \\
&= \int_{-\infty}^{\infty}\frac{e^{itx}-1-itx}{x^2}\sum_{k=1}^{n}x^2dF_{k,n}(x) \qquad (17)\\
&= \int_{-\infty}^{\infty}\frac{e^{itx}-1-itx}{x^2}dG_n(x)=\int_{-\infty}^{\infty}h_t(x)dG_n(x).
\end{aligned}
$$

Тук за всяко n функцията

$$G_n(x)=\sum_{k=1}^{n}\int_{-\infty}^{x}y^2dF_{k,n}(y)=\sum_{k=1}^{n}\int_{\{\eta_{k,n}<x\}}\eta_{k,n}^2d\mathbf{P}$$

е функция на разпределение. Тъй като тя е ненамаляваща, е в сила $G_n(-\infty)=0$ и съгласно (5) имаме

$$G_n(+\infty)=\sum_{k=1}^{n}\int_{-\infty}^{\infty}y^2dF_{k,n}(y)=\sum_{k=1}^{n}\mathbf{D}\eta_{k,n}=1.$$

Понеже за всяко $\tau>0$

$$L_n(\tau)=1-\sum_{k=1}^{n}\int_{-\tau}^{\tau}y^2dF_{k,n}(y)=1-[G_n(\tau)-G_n(-\tau)]\longrightarrow 0 \quad \text{при } n\to\infty,$$

то за всяко $x\neq 0$

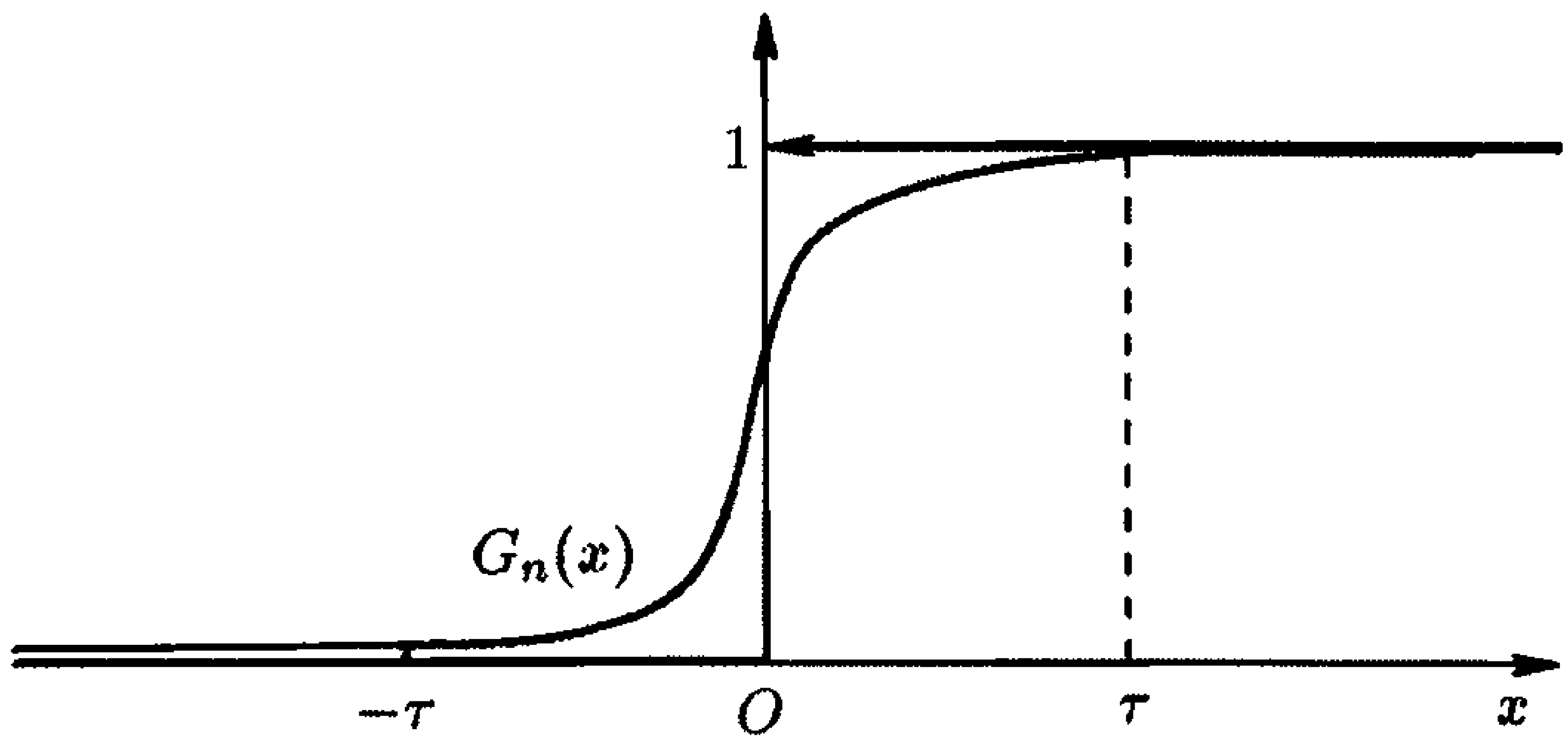
$$G_n(x)\longrightarrow E_0(x) \quad \text{при } n\to\infty, \qquad (18)$$

където $E_0(x)$ е функцията на разпределение на константата нула, т. е. $E_0(x)=0$ при $x\leqq 0$ и $E_0(x)=1$ при $x>0$ (вж. фиг. 16).

От друга страна,

$$h_t(x)=\frac{e^{itx}-1-itx}{x^2}=x^{-2}\sum_{k=2}^{\infty}\frac{(itx)^k}{k!}=-\frac{t^2}{2}+x\sum_{j=0}^{\infty}\frac{(it)^{j+3}x^j}{(j+3)!}.$$

Следователно $\lim\limits_{x\to 0}h_t(x)=h_t(0)=-\dfrac{t^2}{2}$.



Фиг. 16

Освен това от (6.3.8) получаваме

$$|h_t(x)| \leqq x^{-2}|e^{itx} - 1 - itx| \leqq x^{-2}\frac{(tx)^2}{2} = \frac{t^2}{2}.$$

Следователно за всяко $t \in \mathbf{R}^1$ функцията $h_t(x)$ е непрекъсната и ограничена по $x \in \mathbf{R}^1$. Тогава съгласно теорема 7.1.7 сходимостта по разпределение (18) е еквивалентна на слабата сходимост

$$(19) \qquad \int_{-\infty}^{\infty} h_t(x) dG_n(x) \xrightarrow{n\to\infty} \int_{-\infty}^{\infty} h_t(x) dE_0(x) = h_t(0) = -\frac{t^2}{2}.$$

Съотношенията (9), (16), (17) и (19) доказват (7), откъдето следва твърдението на теоремата.

Забележка. Условието (2) на Линдеберг на пръв поглед изглежда доста „тромаво“, макар че има „прозрачен“ вероятностен смисъл. Действително, да означим $A_{k,n} = \{|\eta_{k,n}| > \tau\}$. Тогава

$$\mathbf{P}\left\{\max_{1\leqq k\leqq n}|\eta_{k,n}| > \tau\right\} \leqq \mathbf{P}\left(\bigcup_{k=1}^{n} A_{k,n}\right) \leqq \sum_{k=1}^{n}\mathbf{P}(A_{k,n}).$$

От друга страна,

$$\mathbf{P}(A_{k,n}) = \int_{\{|x|>\tau\}} dF_{k,n}(x) < \frac{1}{\tau^2}\int_{\{|x|>\tau\}} x^2 dF_{k,n}(x).$$

Оттук съгласно (6) получаваме

$$\sum_{k=1}^{n}\mathbf{P}(A_{k,n}) < \frac{1}{\tau^2}L_n(\tau).$$

Сега за всяко $\tau > 0$ от (2) следва

$$\mathbf{P}\left\{\max_{1\leqq k\leqq n}|\eta_{k,n}| > \tau\right\} \longrightarrow 0 \quad \text{при} \quad n\to\infty, \tag{20}$$

т. е. от условието на Линдеберг следва „равномерна пренебрежимост" на събираемите $\eta_{k,n}$ в сумата $\zeta_n = \sum_{k=1}^{n} \eta_{k,n}$, където съгласно (3) $\zeta_n \xrightarrow{d} \zeta \in \mathcal{N}(0,1)$. Може да се докаже и обратното, т. е. от ЦГТ (3) и условието (20) следва условието (2) на Линдеберг.

Условието на Линдеберг е трудно проверяемо, затова на практика по-често се използват условията на Ляпунов.

Теорема 2 (теорема на Ляпунов). *Нека $\xi_1, \xi_2, \dots$ са независими случайни величини, за които*

$$F_k(x) = \mathbf{P}\{\xi_k < x\}, \quad a_k = \mathbf{E}\xi_k, \quad \sigma_k^2 = \mathbf{D}\xi_k, \quad B_n^2 = \sum_{k=1}^{n}\sigma_k^2.$$

Ако за някакво $\delta > 0$ моментът $c_k^{2+\delta} = \mathbf{E}|\xi_k - a_k|^{2+\delta}$ е краен и

$$\lim_{n\to\infty}\frac{C_n}{B_n} = 0, \tag{21}$$

където $C_n^{2+\delta} = \sum_{k=1}^{n} c_k^{2+\delta}$, то е в сила ЦГТ, *т. е.* (3).

Доказателство. Ще покажем, че от (21) следва условието (2) на Линдеберг. Действително, тъй като за всяко $\tau > 0$ върху множеството $\left\{\frac{|x-a_k|}{B_n} > \tau\right\}$ е в сила неравенството $\left(\frac{|x-a_k|}{\tau B_n}\right)^{\delta} > 1$, от (2) получаваме

$$L_n(\tau) \leqq \frac{1}{B_n^2}\sum_{k=1}^{n}\int\limits_{\{|x-a_k|>\tau B_n\}}\frac{|x-a_k|^{2+\delta}}{(\tau B_n)^{\delta}}\,dF_k(x)$$

$$\leqq \frac{1}{B_n^{2+\delta}\tau^{\delta}}\sum_{k=1}^{n}c_k^{2+\delta} = \frac{1}{\tau^{\delta}}\left(\frac{C_n}{B_n}\right)^{2+\delta} \longrightarrow 0$$

при $n\to\infty$ съгласно (21), откъдето по теорема 1 получаваме ЦГТ.

Условието на Ляпунов е най-популярно и лесно се използва при $\delta = 1$, т. е. когато третият централен момент $c^3 = \mathbf{E}|\xi_k - a_k|^3$ е краен и $C_n^3 = \sum\limits_{k=1}^{n} c_k^3$. Тогава от (21) имаме

$$\frac{C_n}{B_n} = \frac{\sqrt[3]{C_n^3}}{\sqrt{B_n^2}} \longrightarrow 0 \quad \text{при} \quad n \to \infty. \tag{22}$$

Не е трудно да се провери, че ако случайните величини $\{\xi_k\}$ са независими и еднакво разпределени с

$$F(x) = \mathbf{P}\{\xi_k < x\}, \ a = \mathbf{E}\xi_k, \ \sigma^2 = \mathbf{D}\xi_k < \infty,$$

то в сила е условието (2) на Линдеберг. Действително, в този случай $B_n^2 = n\sigma^2$ и за всяко $\tau > 0$ имаме

$$L_n(\tau) = \frac{1}{B_n^2}\sum_{k=1}^{n} \int\limits_{\{|x-a|>\tau B_n\}} (x-a)^2 dF(x) = \frac{1}{\sigma^2} \int\limits_{\{|x-a|/\tau\sigma>\sqrt{n}\}} (x-a)^2 dF(x) \longrightarrow 0$$

при $n \to \infty$, откъдето следва теорема 1.1.

Да отбележим,че ЦГТ намира широко приложение в математическата статистика, което е илюстрирано във втората част на книгата.

Задачи

1. Случайните величини $\{\xi_n\}$ и $\{\eta_n\}$ са независими и $\xi_n, \eta_n \in Po(n\lambda)$. Докажете, че

$$\lim_{n\to\infty} \mathbf{P}\left\{\frac{\xi_n - \eta_n}{\sqrt{n}} < x\right\} = \frac{1}{2\pi} \int\limits_{-\infty}^{x/\sqrt{2\lambda}} e^{-u^2/2} du.$$

Упътване. Използвайте ЦГТ и факта, че $\xi_n = X_1 + \cdots + X_n$ където $\{X_k\}$ са независими и $X_k \in Po(\lambda)$.

2. Големината на стъпката на пешеходец е равномерно разпределена в интервала от 70 до 80 cm, а отделните стъпки са независими една от друга. Като използвате нормалното приближение (т. е. ЦГТ), покажете, че вероятността за 10 000 стъпки пешеходецът да измине разстояние, не по-малко от 5,1 km и не повече от 7,49 km, е 0,99947.

3. Случайните величини ξ_1, ξ_2, ... са независими помежду си и имат разпределения

$$\mathbf{P}\{\xi_n = n^\alpha\} = \mathbf{P}\{\xi_n = -n^\alpha\} = \frac{1}{2n^\beta}, \quad \mathbf{P}\{\xi_n = 0\} = 1 - \frac{1}{n^\beta}.$$

Покажете, че при $0 \leqq \beta < 1$ и $2\alpha > \beta - 1$ са в сила условията на теорема Ляпунов.

4. Случайните величини $\xi_1, \xi_2, \ldots$ са независими и равномерно ра делени в интервала $(0, 1)$. Покажете, че

$$\mathbf{P}\left\{\prod_{k=1}^{100} \xi_k < 10 \cdot 2^{-100}\right\} \approx 0,99952.$$

Упътване. Използвайте нормално приближение за $\sum_{k=1}^{100} \log \xi_k$.

5. Случайните величини $\{\xi_k\}$ са независими и еднакво разпределе плътност $f(x) = \frac{3x^2}{2}$, $x \in (-1, 1)$. Покажете, че

$$\lim_{n\to\infty} \mathbf{P}\left\{\sum_{k=1}^{n} \xi_k < \sqrt{n}\right\} = \Phi\left(\frac{\sqrt{6}}{60}\right),$$

където $\Phi(x) = \frac{1}{2\pi} \int\limits_{-\infty}^{x} e^{-u^2/2} du$ е стандартната нормална функция на раз деление.

6. С помощта на ЦГТ докажете, че $\lim\limits_{n\to\infty} e^{-n} \sum\limits_{k=0}^{n} \frac{n^k}{k!} = \frac{1}{2}$.

Упътване. Ако $S_n \in Po(n)$, то

$$\mathbf{P}\{S_n < n\} = e^{-n} \sum_{k=1}^{n} \frac{n^k}{k!}.$$

7. Нека $\{\xi_k\}$ са независими случайни величини,

$$\mathbf{P}\{\xi_k = n^\delta\} = \frac{1}{2} = \mathbf{P}\{\xi_n = -n^\delta\}.$$

Докажете, че при $\delta > -\frac{1}{3}$ за редицата $\{\xi_n\}$ е в сила теоремата на Ляпунс

Втора част

МАТЕМАТИЧЕСКА СТАТИСТИКА

Десета глава

Задачи на математическата статистика. Оценяване на параметри

§ 1. Уводни бележки

В теорията на вероятностите се въвеждат и изследват в теоретичен план редица важни понятия за числови характеристики на случайните събития и случайните величини (вероятност, функция на разпределение, математическо очакване, дисперсия, квантили и др.). Обаче в много редки случаи на практика може с увереност да се твърди, че дадено случайно събитие има определена вероятност (известно число) или че конкретна случайна величина има дадена функция на разпределение (напълно известна функция). Обикновено всички тези характеристики на конкретни случайни събития и величини се оценяват въз основа на наблюдения, на получени по един или друг начин експериментални данни.

За да добием представа за предмета на математическата статистика и за връзката ѝ с теорията на вероятностите, ще приведем две типични за тези дисциплини задачи.

1. Отделно и независимо един от друг са взети за контролен преглед n детайла. Известно е, че вероятността отделно взет детайл да е дефектен е равна на определено число p. Каква е вероятността измежду n-те детайла да се окажат точно k дефектни?

2. Независимо един от друг и по случаен начин са взети за контролен преглед n детайла. При прегледа се оказва, че сред тях има точно k дефектни.

а) Как да определим неизвестната вероятност p отделно взет детайл да се окаже дефектен?

б) Ако p е оценка за тази вероятност, как да определим точността на тази оценка?

в) Как да проверим хипотезата, че бракът е под 1%?

г) Как да различаваме две хипотези, при които бракът е съответно 2 или 5%?

Извличането на качествена и количествена информация въз основа на специално провеждани за целта наблюдения или експерименти е отдавна призната необходимост в обществената практика. Това е първата причина да се създаде, оформи и развива като наука съвременната статистика. Под статистика днес се разбира един обширен клон от науката и практиката, включващ три основни типа дейности:

- събиране на статистически сведения;
- изследване на събраните данни;
- разработка на общи методи и правила за събиране на статистически данни, за тяхното изследване, за получаване на верни изводи от резултатите и за вземане на научно обосновани решения.

Предмет на математическата статистика са третият тип дейности, с които са свързани многобройни теоретични и практически задачи. Като решава тези задачи, статистиката едновременно посочва и какъв трябва да бъде експериментът (наблюденията), за да бъдат правилни съжденията, получени въз основа на събраните по такъв начин данни. Това води дотам, че самата статистика често дава препоръки по провеждането на експеримента, като се превръща и в наука за планирането на експеримента.

Статистическата теория в основите си се опира на теорията на вероятностите. Често се наблюдава и обратната връзка — при изграждане на конкретни вероятностни модели се използват статистически данни. Теорията на вероятностите подсказва как да се подходи към интерпретацията на статистическите данни, какви методи да се използват в работата със статистически материал, как да се възприемат и тълкуват резултатите. Тя трябва да се схваща като изкуство и наука за събиране, анализ и получаване на изводи от количествени (числови), а понякога и от чисто качествени (нечислови) емпирични данни. Специалистът по математическа статистика е длъжен да познава теорията на вероятностите толкова добре, колкото своята област. За потребителя на статистическите методи е достатъчно да познава понякога само „рецептите“ при работа със статистически материал. За пълноценно използване на математическия апарат са необходими познания за същината на статистическите модели, за условията им за прилагане, за формата на поднасяните резултати, за правилата за вземане на решения.

Специалистът по математическа статистика изследва и теоретично свойствата на различни типове вероятностни модели, като предполага известни някои от определящите вероятностни закономерности в моделите. За да могат да се използват подобни модели в практиката, е необходимо да им се „вдъхне живот“ — те да се конкретизират така, че да има точно и обективно съответствие между използваните в модела закономерности и описваните в тях случайни обекти в действителното явление. Именно тази конкретизация на закономерностите и установяването на съответствието им с действителността представляват основните задачи на статистиката.

§ 2. Генерална съвкупност и извадка

Обикновено на статистическите данни се гледа като на еднородни представители на една или няколко наблюдавани характеристики в голяма съвкупност от еднородни индивиди. Когато е уточнена изследваната характеристика на индивидите в общата съвкупност, можем да отъждествим тази съвкупност със съвкупността от възможните стойности на измерваната характеристика у отделните индивиди. Обикновено тя се нарича *генерална съвкупност*. Когато данни за характеристиката не се събират от всички индивиди, а само от тяхна част, говорим за *извадка*. Отделните стойности на характеристиката в реда на получените данни наричаме *наблюдения*. Именно наблюденията за нас са представителите на (извадката от) генералната съвкупност. Броят на наблюденията се нарича *обем на извадката*.

Към използване на извадки се прибягва, когато интересуващата ни генерална съвкупност е твърде многобройна или обектите ѝ са трудно достъпни, или има други причини, непозволяващи да се изучат всички обекти.

Например съвкупността на обектите може да бъде цялото население на България на възраст 16 години. Измерваният признак нека е ръстът на съответните индивиди. Може да отъждествим съвкупността на индивидите със съвкупността от възможните стойности на ръста. Вместо да се измерват всички деца на 16-годишна възраст, може да бъдат проведени измервания в няколко селища и получените резултати ще представят извадката от генералната съвкупност. По нея по-нататък може да се правят изводи за ръста на всички деца в страната на посочената възраст.

Няколко думи за характера на извадката. Тя трябва да бъде непреднамерена и да съдържа в себе си приблизително всички особености на генералната съвкупност, за да представи най-добре нейната цялост, т.е. да бъде *репрезентативна*. При горния пример преднамереност може да се получи, ако избраните класове за измерване ръста на децата са само от момчета (или само от момичета) или само от спортни училища, или от училища за бавно развиващи се деца. Извадката няма да представя правдоподобно съвкупността на децата на 16-годишна възраст. Освен това извадката трябва да има достатъчно голям обем. Така по измерения ръст на няколко деца едва ли може да се съди за ръста на всички деца. Въобще извадката трябва да бъде микромодел на генералната съвкупност. Това се постига чрез т. нар. *чисто случаен избор*. Тук е по-просто да покажем как не трябва да се прави извадка.

Нарушаването на принципа за случаен избор понякога води до сериозни грешки. Известно със своя неуспех е запитването, проведено от американското списание „Литературен преглед" относно изхода на предстоящите президентски избори през 1936 г. Кандидати били Ф. Д. Рузвелт и А. М. Ландън. Като генерална съвкупност редакцията използвала телефонните указатели. Като избрала по случаен начин 4 милиона адреса, тя изпратила по цялата страна картички с въпроси за отношението към кандидатите за президенти. Списанието изразходвало големи суми за разпращането и обработката на картичките и обявило, че за президент ще бъде избран А. М. Ландън, и то с голяма преднина в гласовете. Резултатът от изборите обаче се оказал противоположен на тази прогноза.

Тук са извършени наведнъж две грешки. Първо, телефонните указатели не давали репрезентативна извадка от населението на страната, защото абонатите тогава били главно от по-заможни семейства. Второ, отговори са получени не от всички, а от хора, достатъчно уверени в своето мнение и привикнали да отговарят на писма — от представителите на деловия свят, които поддържали Ландън. Ако редакцията по-критично беше пристъпила към своята работа, тя би открила и без това известното на всички.

Явление, подобно на току-що описаното, когато извадката не представлява цялата генерална съвкупност, а само някакъв неин слой, се нарича „отместване на извадката“. Отместването е един от основните източници на грешки при използването на метода на извадките.

Същевременно тогава социолозите Дж. Галъп и Е. Роупър правилно предсказали победата на Ф. Рузвелт, обосновавайки я само на 4 хиляди анкети. Причината за този успех била не само правилното съставяне на извадката. Социолозите отчели, че обществото се разпада на социални групи, които са сравнително еднородни в отношението си към кандидатите за президенти. Затова извадката от слоя може да бъде малобройна, но със същия резултат на точност. С помощта на резултатите от изследването по слоеве може да се характеризира обществото като цяло. Днес подобна методика е общоприета.

Генералната съвкупност също дава своето отражение върху извадката. Когато броят на индивидите в съвкупността е безкрайно голям, обемът на извадката няма да окаже влияние върху броя на неизследваните индивиди. При построяване на процедурите по статистическото изследване се има предвид, че се изследва неограничена съвкупност от индивиди, и се говори за статистика на безкрайните индивиди. При непреднамерен случаен подбор на индивидите е удобно да се приеме, че извадката представлява съвкупност от независими реализации на случайната величина ξ, която представя измерваната характеристика в индивидите на генералната съвкупност. В такива случаи последователните наблюдения ще могат да бъдат представени с получената редица от измервания $\xi_1, \xi_2, \ldots, \xi_n$. Числото n е обемът на извадката. Удобно е при теоретични разглеждания (най-вече при оценки за точността и сигурността на получаваните статистически решения) елементите на извадката да се третират като случайни величини, докато в изчислителните процедури на тях се гледа като на числа, с които се извършват пресмятанията.

Векторът $\vec{\xi} = (\xi_1, \xi_2, \ldots, \xi_n)$, подобно на разглежданията в теорията на вероятностите (вж. I, § 2.5), се третира като единично наблюдение в сложен опит, състоящ се от последователното провеждане на n-те наблюдения. Съвкупността $\mathcal{X}$ от възможни стойности на вектора $\vec{\xi}$ се нарича *извадково пространство.* Неговата плътност на разпределение е функция с дефиниционно множество $\mathcal{X}$ и играе изключително важна роля в статистиката.

Нарича се *функция на правдоподобие* и обикновено се бележи с $L(\vec{x}) = L(x_1, \ldots, x_n)$.

Най-често непреднамерената репрезентативна извадка от безкрайна съвкупност се третира като редица от *независими еднакво разпределени* случайни величини $\xi_1, \xi_2, \ldots, \xi_n$, всички с разпределението на изследваната характеристика ξ (нарича се още проста извадка).

Ако ξ има вероятностна плътност на разпределение $f(x)$ или дискретно разпределение $\{p_x\}$, функцията на правдоподобие ще се представя с израза

$$(1) \qquad L(\vec{x}) = \prod_{i=1}^{n} f(x_i) \quad \text{или} \quad L(\vec{x}) = \prod_{i=1}^{n} p_{x_i}.$$

Когато генералната съвкупност е съществено крайна, т.е. общият брой N на индивидите в нея не може да се приеме за безкраен, всеки нов индивид, изследван при формирането на извадката, ще се избира от оставащата част на съвкупността. Този начин на избиране прави членовете на извадката зависими и налага разработването на теорията на статистиката на крайните съвкупности. Хубаво изложение на най-новите постижения на тази теория може да се намери в гл. 12 на книгата на Ю. К. Беляев „Статистический контроль качества“, М., Наука, 1975.

За да отбележим ролята на статистическото описване на генералната съвкупност, ще приведем един пример, посочен от бележития английски учен сър Роналд Фишер — един от създателите на съвременната математическа статистика.

„... Йоханес Шмит от Карлсберговата лаборатория в Копенхаген бе не само ихтиолог, но и неуморен биостатистик. Той развиваше идеята, че рибите от един вид се делят на относително изолирани съобщества. Между тези групи той намираше статистически различия по броя на гръбначните прешлени или по лъчевите кости на плавниците. Причината за различията е, че тези съобщества не се смесват при размножаване. Всяка група хвърля хайвера си на свое място. Често такива различия са забелязвани между рибни стада от един вид, обитаващи един фиорд.

Обаче за змиорките Йоханес Шмит не можа да открие никакви статистически различия между индивидите, уловени даже в твърде отдалечени едно от друго места — различни части на Европейския континент, Азорските острови, Нил или Исландия. Между другото броят на гръбначните прешлени при змиорката е твърде изменчив. Йоханес Шмит реши, че змиорките от всички различни речни системи образуват едно съобщество. Следователно те трябва да имат общо място за размножаване.

След известно време това предположение се потвърди по време на експедицията на изследователския кораб „Дана“, един от главните успехи на която бе улавянето на личинки от змиорка в ограничен район на Западния Атлантик.“

§ 3. Задачи на математическата статистика

Ще направим преглед на някои от задачите на статистиката, свързвайки ги с изучаването на безкрайните съвкупности.

1. Задачи, свързани с оценка на вероятността. Събитието A има вероятност p, чиято стойност не познаваме. По резултатите от опити трябва да оценим числото p, тъй като априорните съждения не са достатъчни. Генералната съвкупност е съвкупността на всички опити, провеждани при едни и същи условия, при които може да се осъществи събитието A. Извадката — това са резултатите от проведените опити, които на практика са винаги краен брой, понякога доста голям.

2. Задачи, свързани с оценка на неизвестни функции на разпределение. Тези задачи включват оценка на неизвестна вероятност. Ако дефинираме случайна величина ξ, равна на 1, когато се осъществи случайното събитие A, и равна на 0, когато не се осъществи A, определянето на $\mathbf{P}(A)$ е частен случай от задачата за определяне разпределението на ξ. В много случаи не знаем нищо за интересуващата ни функция на разпределение $F(x)$ освен, разбира се, дефиницията ѝ като вероятност $\mathbf{P}\{\xi < x\} = F(x)$. Така се наблюдава известна прилика между задачата за определяне на $F(x)$ при фиксирано x и задачата за определяне неизвестната вероятност на случайно събитие (което в случая е неравенството $\{\xi < x\}$).

Понякога може да е познат аналитичният вид $F(x;\theta)$ на функцията на разпределение, но да е неизвестна стойността на неизвестния параметър θ, който може да бъде вектор $\vec{\theta} = (\theta_1, \ldots, \theta_n)$ с неизвестни координати. Тук задачата за оценка на $F(x)$ се свежда до вида:

3. Задачи за оценка на неизвестни параметри на наблюдаваната случайна величина ξ. Тези задачи включват оценката както на параметри, влизащи в явния вид на функцията на разпределение (например параметрите a и σ^2 в нормалното разпределение, λ — в експоненциалното и поасоновото, p — в биномното и геометричното и пр.), така и на други параметри, свойствени за самата величина (например медианата и други квантили на разпределението, средното $\mathbf{E}\xi$, дисперсията $\mathbf{D}\xi$ или някои моменти от вида $\mathbf{E}^r$ и др.). Оценките могат да бъдат точкови (за всеки неизвестен параметър се предлага число, получено от резултатите от наблюдението) или интервални (вместо стойност на параметъра се предлага интервал, в който с известна вероятност се твърди, че се намира истинската му стойност).

4. Задачи за проверка на хипотези. Това е широк клас от задачи, съпътстващ почти всяко статистическо изследване. По някакви съображения може да се изгради хипотезата, че търсената функция на разпределение е $F(x)$. Как ще се провери дали тази хипотеза е вярна? Подобно е положението винаги след оценка на неизвестните параметри в известния клас на разпределенията на наблюдаваната величина. След заместване на параметрите с оценките твърдението, че така получената $F(x)$ е истинското разпределение, е една хипотеза, получена въз основа на предшестващите статистически изводи. Тази хипотеза трябва да се провери за съответствие с резултатите от наблюденията. Пак към този клас задачи се отнасят задачите за проверка дали наблюдаваните две случайни величини имат еднакви средни стойности (или дисперсии, или функции на разпределение), дали разпределението на изследваната величина е от предполагаем клас (нормално, експоненциално или с монотонна интензивност на отказите), дали зависимостта между две случайни величини (или две групи от величини) се изразява точно със закона, който нашите разсъждения или изчисления ни подсказват, и много други.

Задачите от този тип за проверка на хипотези се отнасят към различни раздели на математическата статистика — ***параметрична и непараметрична статистика, многомерен анализ*** и др.

5. Задачи за установяване на статистически зависимости между събития и величини. Това е също обширен клас задачи, включващ проверките за независимост на наблюденията в извадката, установяване степента на зависимост между две случайни събития или величини, установяване на математическия вид на количествената зависимост между две съвкупности от случайни величини. При това може да е известна функционалната форма, но да са неизвестни някои коефициенти в нея, или пък формата въобще да не е известна и да се търси заедно с параметрите, с чиято помощ се записва, и пр. Когато даден основен показател зависи от голям брой случайни величини, е интересно те да се класифицират по степента на въздействието им върху този показател и да се знае кои от величините влияят благоприятно върху него и кои не. Задачите от този клас са действително много важни и разнообразни. Те довеждат до силно развитие на такива дялове на математическата статистика като корелационния, регресионния, дисперсионния, факторния, дискриминантния, компонентния и други анализи, които се обособиха вече като самостоятелни направления със свои проблеми и методи за тяхното решаване.

§ 4. Обща характеристика на точковите оценки

А. Обект на изследване в математическата статистика найчесто са някои обобщаващи характеристики на генералната съвкупност. В статистическите задачи става дума не за намиране точните стойности на параметрите, а за оценяването им, тъй като експериментът не може да даде абсолютно точен резултат за наблюдаваната величина. Резултатите от опитите (от наблюденията на индивидите в извадката) са случайни, затова и оценките, които се получават с тяхна помощ, имат този случаен характер. С помощта на статистическите правила се получават приближени стойности на неизвестните параметри в следния обобщен смисъл: постоянните величини, характерни за изследваната съвкупност, се заместват със случайни величини, построени съгласно определени правила с помощта на резултатите от наблюденията.

Математическата постановка на задачата изглежда така: Разглежда се определена генерална съвкупност Ω. Нека $\vec{\theta}=(\theta_1,\ldots,\theta_k)$ е неизвестен постоянен вектор, характеризиращ определен признак в Ω.

В съвкупността Ω дефинираме случайна величина ξ, която по определен начин зависи от θ, т.е. в своите проявления у отделни индивиди на величината ξ съдържа някаква информация за параметъра θ. Предполага се, че е получена извадка от n елемента на Ω, в които са наблюдавани (измерени) стойностите $\xi_1,\xi_2,\ldots,\xi_n$ на ξ. *Оценка* или *статистика* се нарича всяка функция $t(\xi_1,\ldots,\xi_n)$ на получените наблюдения, която приемаме за стойност на дадена функция $r(\theta)$ от неизвестните параметри $\theta_1, \ldots, \theta_k$. Функцията $r(\theta)$ също е параметър на генералната съвкупност. Например оценяваните параметри могат да бъдат средният ръст и средното тегло на едногодишните индивиди от нововнедрена порода животни или отношението между ръста и теглото им, делът на разходите за културни мероприятия в бюджета на домакинствата в България (изобщо или само на домакинствата от определена категория), процентът на брака в производството на едно предприятие и др.

Съществуват много възможности за построяване на статистически оценки на каквито и да са конкретни параметри или на техни функции. Преди да се спрем на някаква оценка, обикновено разглеждаме какви изисквания тя трябва да удовлетворява. Очевидно е, че оценките трябва да бъдат по възможност прости, точни и при увеличаване броя на наблюденията точността им да расте. Затова трябва да се започне с установяване на някои общи принципи, към които се придържаме при прегледа на цялото многообразие от оценки, предлагани в практиката.

Зависимостта от параметъра θ на разпределението на ξ ще бележим с $F(x \mid \theta)$, на функцията на правдоподобие — с $L(\vec{x} \mid \theta)$, а на математическите очаквания на каквито и да са функции от вектора на наблюденията $\vec{\xi}$ — с $\mathbf{E}_\theta$.

Б. Нека $t(\xi_1, \ldots, \xi_n)$ (накратко ще я бележим $t_n(\xi)$) е конструираната оценка за функцията $r(\theta)$ от неизвестния параметър $\theta = (\theta_1, \ldots, \theta_k)$. Оценката се нарича неизместена, ако

$$(1)\quad \mathbf{E}_\theta t_n(\xi) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} t(x_1, \ldots, x_n) L(x_1, \ldots, x_n \mid \theta) dx_1 \ldots dx_n = r(\theta).$$

Когато оценката на функцията $r(\theta)$ не удовлетворява равенството (1), разликата $\mathbf{E}_\theta t_n(\xi) - r(\theta)$ се нарича систематична грешка на оценката $t_n(\xi)$.

Лема 1. *Средното аритметично*

$$\bar{\xi}_k = \frac{1}{n}(\xi_1 + \xi_2 + \cdots + \xi_n)$$

е неизместена оценка на математическото очакване $\mathbf{E}\xi$ *на наблюдаваната случайна величина* ξ.

Доказателство. Въз основа на линейното свойство на математическото очакване $\mathbf{E}(a\xi + b\eta) = a\mathbf{E}\xi + b\mathbf{E}\eta$ имаме

$$\mathbf{E}\bar{\xi}_n = \frac{1}{n}(\mathbf{E}\xi_1 + \cdots + \mathbf{E}\xi_n) = \mathbf{E}\xi,$$

тъй като $\mathbf{E}\xi_i = \mathbf{E}\xi$.

Би могло да се помисли, че и оценките

$$\hat{\nu}_r = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - \bar{\xi}_n)^r$$

на централния момент от ред r, $\mathbf{E}(\xi - \mathbf{E}\xi)^r$, също са неизместени. Оказва се, че това не е така. При независими наблюдения $\xi_1, \ldots, \xi_n$ неизместени са оценките

$$\hat{\mu}_r = \frac{n^{r-1}}{(n-1)(n-2)\cdots(n-r+1)} \cdot \frac{1}{n}\sum_{i=1}^{n}(\xi_i - \bar{\xi}_n)^r. \tag{2}$$

(Докажете!) Специално при независими наблюдения неизместена оценка за дисперсията $\mathbf{D}\xi$ на наблюдаваната случайна величина ξ не е величината

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - \bar{\xi}_n)^2,$$

а величината

$$\hat{s}_n^2 = \frac{1}{n-1} \cdot \sum_{i=1}^{n}(\xi_i - \bar{\xi}_n)^2.$$

Наистина лесно се пресмята, че

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\xi_i^2 - \bar{\xi}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\xi_i^2 - \frac{1}{n^2}\left(\sum_{i=1}^{n}\xi_i^2 - \sum_{i\neq j}\xi_i\xi_j\right)$$

$$= \frac{n-1}{n^2}\sum_{i=1}^{n}\xi_i^2 - \frac{1}{n^2}\sum_{i\neq j}\xi_i\xi_j.$$

Оттук поради независимостта на ξ_i и ξ_j намираме

$$\mathbf{E}\hat{\sigma}_n^2 = \frac{n-1}{n}\mathbf{E}\xi^2 - \frac{1}{n^2}\sum_{i\neq j}\mathbf{E}\xi_i\mathbf{E}\xi_j = \frac{n-1}{n}\left[\mathbf{E}\xi^2 - (\mathbf{E}\xi)^2\right].$$

Оценката $\hat{\sigma}_n^2$ е изместена, докато

$$\hat{s}_n^2 = \frac{n}{n-1}\hat{\sigma}_n^2 \quad \text{и} \quad \mathbf{E}\hat{s}_n^2 = \frac{n}{n-1}\mathbf{E}\hat{\sigma}_n^2 = \mathbf{D}\xi. \tag{3}$$

Като се използва, че $\mathbf{D}\hat{s}_n^2 = \mathbf{E}(\hat{s}_n^2)^2 - [\mathbf{E}\hat{s}_n^2]^2$, след прости пресмятания може да се получи

$$\mathbf{D}\hat{s}_n^2 = \frac{1}{n}\left[\mathbf{E}(\xi - \mathbf{E}\xi)^4 - \frac{n-3}{n-1}(\mathbf{D}\xi)^2\right] \quad \text{при} \quad \mathbf{E}\xi^4 < \infty.$$

Известно е още, че не може да се получат неизместени оценки за централните моменти от ред r при $r > n$.

В. Ясно е, че изискването дадена оценка да е неизместена, е естествено. В такива случаи поне ще се избягнат систематични грешки. Понякога обаче е трудно да се открие на само как да се отстрани изместването, но дори самото наличие или липса на изместване на предложената оценка. За щастие в повечето случаи сравнително просто се намират *асимптотично неизместени оценки*. Това са такива (възможно изместени) оценки, чието математическо очакване клони към истинската стойност на оценявания параметър при неограничено увеличаване на броя n на наблюденията.

Така, ако оценяваме дисперсията на случайната величина ξ, оценките $\hat{\sigma}_n^2$ и $\hat{s}_n^2$ са асимптотично неизместени и при достатъчно големи n тези оценки малко се различават. Това се вижда добре от връзката (3) и от дефинициите на $\hat{\sigma}_n^2$ и $\hat{s}_n^2$.

Дефиниция. *Казваме, че $t_n(\xi)$ е състоятелна оценка (с.о.) за $r(\theta)$, когато при $n \to \infty$ клони по вероятност към оценяваната функция $r(\theta)$, т.е. когато е изпълнено*

$$\mathbf{P}\{|\, t_n(\xi) - r(\theta)\,| < \varepsilon\} \xrightarrow[n\to\infty]{} 1, \tag{4}$$

каквото и да е числото $\varepsilon > 0$.

Състоятелността установява неограничено повишаване на точността на оценката при увеличаване броя на наблюденията, без да показва колко бързо расте тази точност.

Дефиниция. *Оценката $t_n(\xi)$ се нарича силно състоятелна, ако $t_n(\xi)$ клони към $r(\theta)$ при $n \to \infty$ с вероятност 1, т.е. когато за всяко $\varepsilon > 0$ е изпълнено асимптотичното съотношение*

$$\mathbf{P}\{\sup_{k \geqq n} |\, t_k(\xi) - r(\theta)\,| < \varepsilon\} \xrightarrow[n\to\infty]{} 1.$$

Различието между двата вида състоятелност е същото, както различието между слабите и усилените закони за големите числа. Състоятелността на оценките най-често се установява с използване на законите за големите числа. От теоремата на Колмогоров например следва, че оценката $\bar{\xi}_n$ е силно състоятелна за $\mathbf{E}\xi$. От представянето на оценката

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\xi_i^2 - \bar{\xi}_n^2$$

следва (по теоремата на Слуцки, § 7.1), че ако съществува дисперсията $\mathbf{D}\xi$, то $\hat{\sigma}_n^2$ е силно състоятелна оценка за $\mathbf{D}\xi$.

Пример. Известно е, че квантилите x_p (по-точно p-квантилите) на разпределението на случайната величина ξ се дефинират като най-големите решения на уравненията (вж. § 4.5)

$$\mathbf{P}\{\xi < x_p\} \leqq p, \quad p \in (0,1).$$

Тук величините p се задават като числа между 0 и 1, а x_p са неизвестни. Така p-квантилът x_p е точка върху числовата права, която я дели на две полуправи. Вероятността случайната величина ξ да приеме стойност наляво от x_p е равна на p, а $1-p$ е вероятността ξ да приеме стойност, не по-малка от x_p. Квантилът $x_{1/2}$ се нарича медиана (Me) на разпределението.

Емпиричните оценки на квантилите са такива числа $\hat{x}_p$ от вариационния ред

$$(\xi^*): \ \xi_{(1)} \leqq \xi_{(2)} \leqq \cdots \leqq \xi_n$$

на наблюденията $\xi_1, \xi_2, \ldots, \xi_n$, които делят редицата (ξ^*) на две части, така че наляво от $\hat{x}_p$ да се намират част p от наблюденията. С други думи, броят на наблюденията, по-малки от емпиричния квантил $\hat{x}_p$, се отнася към общия брой на наблюденията, по-големи от $\hat{x}_p$, почти както $p:(1-p)$. Налага се да се каже „почти", тъй като отношенията $k:(n-k)$ могат да приемат само крайно много стойности.

Емпиричният p-квантил се определя с равенството

$$\hat{x}_p = \xi_{([np])}.$$

По закона за големите числа (теоремата на Бернули) емпиричните квантили клонят по вероятност към теоретичните, т.е. $\hat{x}_p$ е състоятелна оценка за x_p. Специално $\xi_{([\frac{n}{2}])}$ е състоятелна оценка за $\mathrm{Me}(\xi)$ (медианата на разпределението на ξ).

Г. Изискванията за неизместеност и състоятелност не определят оценката $t_n(\xi)$ еднозначно. Например от теорията е известно, че за случайна величина ξ, разпределена по закона на Поасон с параметър λ, е изпълнено $\mathbf{E}\xi = \mathbf{D}\xi = \lambda$. Ние разполагаме с две неизместени оценки $\bar{\xi}_n$ и $\hat{s}_n^2$ за λ, които съответстват на оценките за средното и за дисперсията на ξ. Коя от двете оценки да предпочетем?

Желателно е оценките да бъдат по възможност близки до оценявания параметър. Оценката обаче е случайна величина и поради това може повече или по-малко да се отклонява от параметъра. Необходимо е големите отклонения да се срещат по-рядко, а мярка за отклоненията на случайната величина (оценката) от нейното средно (параметъра при неизместени оценки) е дисперсията. С оглед на това най-желателни са неизместените оценки с минимална дисперсия, които се наричат *ефективни оценки*.

Изобщо познаването на дисперсията на оценката $t_n(\xi)$ за параметъра $r(\theta)$ е полезно и по други съображения. Ако $t_n(\xi)$ е неизместена оценка за $r(\theta)$, по неравенството на Чебишов можем да получим една долна граница за вероятностите на събитията $|t_n(\xi) - r(\theta)| \leqq \varepsilon$, където $\varepsilon > 0$ е дадено число. Имаме

$$\mathbf{P}\left\{|t_n(\xi) - r(\theta)| \leqq \varepsilon\right\} \geqq 1 - \frac{\mathbf{D}t_n(\xi)}{\varepsilon^2}.$$

Затова изискването оценката да има минимум дисперсия е съвсем естествено. Ще отбележим обаче, че само в редки случаи оценката е едновременно неизместена, състоятелна и ефективна. Получената по-горе оценка $\bar{\xi}_n$ за $\mathbf{E}\xi$ има дисперсия

$$\mathbf{D}\bar{\xi}_n = \mathbf{D}\left(\frac{1}{n}\sum_{i=1}^{n}\xi_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbf{D}\xi_i = \frac{1}{n}\mathbf{D}\xi.$$

Чрез неравенството на Чебишов можем да правим оценки за точността на приближението на $\bar{\xi}_n$ до истинското средно $\mathbf{E}\xi$:

$$\mathbf{P}\left\{\left|\bar{\xi}_n - \mathbf{E}\xi\right| \leqq \varepsilon\right\} = 1 - \mathbf{P}\left\{\left|\bar{\xi}_n - \mathbf{E}\xi\right| \geqq \varepsilon\right\} \geqq 1 - \frac{\mathbf{D}\bar{\xi}_n}{\varepsilon^2} = 1 - \frac{\mathbf{D}\xi}{n\varepsilon^2}.$$

Оттук се вижда още веднъж, че колкото броят на наблюденията е по-голям, толкова по-вероятно е оценката да е близка до оценяваната стойност. Последното неравенство често се използва при планирането на броя на наблюденията с цел да се получат оценки с желана точност. Абсолютна точност никога не се гарантира поради случайния характер на оценките, но може да се гарантира достатъчно висока вероятност, че оценката е на разстояние ε от оценяваната характеристика.

Д. Вероятността

$$\alpha_n = \mathbf{P}\left\{|t_n(\xi) - r(\theta)| < \varepsilon\right\}$$

се нарича *надеждност на оценката с точност* ε. И така, всяка оценка се характеризира със своята точност и надеждност — две важни величини за т. нар. теория на планирането на експеримента. Определянето на надеждността α_n при дадена точност ε е трудна задача, защото трябва да се знае разпределението на оценката. Много често обаче това разпределение е асимптотично близко до нормалното или до някои известни други разпределения и задачата може да се реши поне приблизително. В общия случай неравенствата като чебишовото помагат при оценяване надеждността и точността на статистическите оценки.

Е. Пример. Едно предприятие произвежда серийни изделия при постоянна технология. Вероятността p отделно взето изделие да няма дефекти не е известна. Желателно е да се определи p с грешка, ненадминаваща 0,005. Какъв трябва да е броят n на проверените изделия, който би гарантирал, че честотата $\frac{\nu_n}{n}$ на изделията без дефекти ще се различава от p не повече от 0,005, т.е. че $\left|\frac{\nu_n}{n} - p\right| < 0,005$.

Няма абсолютна гаранция, каквото и да е n, че непременно $\left|\frac{\nu_n}{n} - p\right| < 0,005$. Затова определяме достатъчно висока вероятност — например $\gamma = 0,95$, и искаме неравенството $\left|\frac{\nu_n}{n} - p\right| < 0,005$ да е изпълнено с вероятност 0,95 или повече, т.е.

$$\alpha_n = \mathbf{P}\left\{\left|\frac{\nu_n}{n} - p\right| < 0,005\right\} \geqq 0,95.$$

Ще забележим, че

$$\mathbf{E}\left(\frac{\nu_n}{n}\right) = \frac{1}{n}\mathbf{E}\nu_n = \frac{1}{n}.np = p \quad \text{и} \quad \mathbf{D}\left(\frac{\nu_n}{n}\right) = \frac{p(1-p)}{n}.$$

Затова можем да използваме неравенството на Чебишов, според което

$$\alpha_n = \mathbf{P}\left\{\left|\frac{\nu_n}{n} - p\right| < 0,005\right\} \geqq 1 - \frac{p(1-p)}{n.(0,005)^2} \geqq 1 - \frac{1/4}{n.25.10^{-6}}.$$

Неравенството $\alpha_n \geqq 0,95$ ще е изпълнено, ако изберем такова n, че $1 - [4n.25.10^{-6}]^{-1} \geqq 0,95$. Минималното такова n е 200 000.

Полученото число е доста завишено. Използването на централната гранична теорема (т.е. на по-точни знания за разпределението на величината $\frac{\nu_n}{n}$) дава много по-малко число — $n \approx 10\,000$.

§ 5. Извадкови статистики от нормално разпределени случайни величини

Когато наблюдаваната случайна величина ξ има нормално разпределение, извадковите статистики, получени от наблюденията, притежават редица важни свойства, които се използват в практическите изследвания.

Теорема 1. *Ако $\xi_1, \xi_2, \ldots, \xi_n$ е проста извадка от нормално разпределена случайна величина $\xi \in \mathcal{N}(a, \sigma^2)$, то разпределението на сумата от квадратите $\frac{1}{\sigma^2}\sum_{i=1}^{n}(\xi_i - a)^2$ е χ^2-разпределение с n степени на свобода.*

Доказателство. Ако положим $\eta_i = \frac{\xi_i - a}{\sigma}$, то $\eta_i \in \mathcal{N}(0,1)$ и посочената сума е $\chi^2_{(n)} = \sum_{i=1}^{n}\eta_i^2$, където η_i са независими. По § 5.4 твърдението е очевидно.

Теорема 2. *Ако $\xi_1, \xi_2, \ldots, \xi_n$ е проста извадка от нормално разпределена случайна величина $\xi \in \mathcal{N}(a, \sigma^2)$, извадковото средно $\bar{\xi}_n$ и извадковата дисперсия $\hat{s}_n^2$ са статистически независими. Случайната величина $\sqrt{n}\frac{\bar{\xi}_n - a}{\sigma}$ е $\mathcal{N}(0,1)$-разпределена, а величината $(n-1)\frac{\hat{s}_n^2}{\sigma^2}$ има χ^2-разпределение с $n-1$ степени на свобода.*

Доказателство. Като сума от независими, нормално разпределени $\mathcal{N}(a, \sigma^2)$ случайни величини, величината $\bar{\xi}_n$ има разпределение $\mathcal{N}\left(a, \frac{\sigma^2}{n}\right)$. Това е равнозначно на твърдението, че преобразуваната (стандартизирана) величина $\sqrt{n}\frac{\bar{\xi}_n - a}{\sigma}$ има разпределение $\mathcal{N}(0,1)$.

Величината $\hat{s}_n^2$ няма да се промени, ако заменим всички наблюдения ξ_i с центрираните отклонения $\xi_i' = \xi_i - a$. Ето защо без ограничение на общността можем да считаме, че $a = 0$, т.е. $\xi_i \in \mathcal{N}(0, \sigma^2)$. Ще преминем от величините ξ_i към нови случайни величини η_i, въведени по формулите

$$\eta_i = \sum_{j=1}^{n} a_{ij}\xi_j; \quad i = 1, \ldots, n.$$

Тук коефициентите a_{ij} образуват квадратна матрица **A** от ред n. Нека всички елементи от последния ред на матрицата **A** са равни на $\frac{1}{\sqrt{n}}$, така че

$$(1) \qquad \eta_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i = \sqrt{n}\,\bar{\xi}_n \quad \text{и} \quad \sum_{j=1}^{n} a_{nj}^2 = 1.$$

Нека по-нататък матрицата **A** е такава, че са изпълнени условията за ортогоналност на редовете ѝ, т.е.

$$\sum_{j=1}^{n} a_{kj}a_{ij} = \delta_{kj},$$

където δ_{kj} е кронекеровият символ. Това винаги може да се направи. По такъв начин сме извършили ортогонална трансформация в n-мерното извадково пространство, което означава въртене около началото на координатната система. При това дължините на векторите не се променят, нито ъглите между тях, т.е. изпълнено е $\sum\limits_{i=1}^{n} \xi_i^2 = \sum\limits_{i=1}^{n} \eta_i^2$. Всички разпределения на величините η_i, $i = 1, \ldots, n$, обаче са нормални като линейни комбинации от независими нормално разпределени случайни величини. При това

$$\mathrm{E}\eta_i = 0; \qquad \mathrm{D}\eta_i = \mathrm{E}\eta_i^2 = \sum_{j=1}^{n} a_{ij}^2 \mathrm{E}\xi_j^2 = \sigma^2 \sum_{j=1}^{n} a_{ij}^2 = \sigma^2.$$

От условието за ортогоналност при $k \neq i$ имаме

$$\mathrm{E}\eta_k\eta_i = \sum_{j=1}^{n} a_{kj}a_{ij}\mathrm{E}\xi_j^2 = \sigma^2 \sum_{j=1}^{n} a_{kj}a_{ij} = 0.$$

Това означава, че нормално разпределените случайни величини $\eta_1, \dots, \eta_n$ са некорелирани, а следователно (вж. § 5.2) са и независими.

За величината $\hat{s}_n^2$ имаме

$$(n-1)\hat{s}_n^2 = \sum_{i=1}^{n}(\xi_i - \bar{\xi}_n)^2 = \sum_{i=1}^{n}\xi_i^2 - n\bar{\xi}_n^2 = \sum_{i=1}^{n}\eta_i^2 - \eta_n^2 = \sum_{i=1}^{n-1}\eta_i^2; \quad n\bar{\xi}_n^2 = \eta_n^2.$$

Оттук тривиално представяме

$$\frac{n-1}{\sigma^2}\hat{s}_n^2 = \sum_{i=1}^{n-1}\left(\frac{\eta_i}{\sigma}\right)^2 \tag{2}$$

като сума от квадрати на $\mathcal{N}(0,1)$-разпределени случайни величини. С това теоремата е доказана, тъй като от (1) и (2) се вижда, че сме представили $\bar{\xi}_n$ и $\hat{s}_n^2$ като функции на две групи от независими случайни величини — съответно η_n и $\eta_1, \dots, \eta_{n-1}$.

Теорема 3. *Ако $\xi_1, \xi_2, \dots, \xi_n$ е извадка от $\mathcal{N}(a, \sigma^2)$-разпределена случайна величина ξ, то величината $t = \sqrt{n}\dfrac{\bar{\xi}_n - a}{\hat{s}_n}$ има разпределение на Стюдънт с $n-1$ степени на свобода.*

Доказателство. Нека положим $X = \sqrt{n}\dfrac{\xi_n - a}{\sigma}$, $Y = (n-1)\dfrac{\hat{s}_n^2}{\sigma^2}$. Тогава ще е вярно представянето

$$t = \frac{X}{\sqrt{Y/(n-1)}},$$

където X е $\mathcal{N}(0,1)$-разпределена случайна величина, а Y има $\chi^2(n-1)$-разпределение. Това е достатъчно (вж. § 5.4), за да твърдим, че t има съответното разпределение на Стюдънт, посочено в теоремата.

§ 6. Оценки с минимална дисперсия

Нека $\xi_1, \xi_2, \dots, \xi_n$ са наблюдения над случайна величина ξ, която зависи от неизвестен параметър (възможно многомерен) θ, и нека $t_n(\xi) = t(\xi_1, \xi_2, \dots, \xi_n)$ е неизместена оценка за някаква функция $r(\theta)$. В съответствие с въведената в предишния параграф терминология това означава, че $\mathbf{E}_\theta t_n(\xi) = r(\theta)$. Тази оценка се нарича *неизместена оценка с минимална дисперсия*, когато за всяка друга неизместена оценка $\tilde{t}_n(\xi)$ е изпълнено

$$\mathbf{D}_\theta \tilde{t}_n(\xi) \geqq \mathbf{D}_\theta t_n(\xi).$$

Теорема 1. *Ако съществува неизместена оценка с минимална дисперсия, тя е единствена.*

Доказателство. Да допуснем, че съществуват две различни неизместени оценки с минимална дисперсия $t_n^{(1)}$ и $t_n^{(2)}$, т.е.

$$\mathbf{E}_\theta t_n^{(1)} = \mathbf{E}_\theta t_n^{(2)} = r(\theta);$$

$$\mathbf{D}_\theta t_n^{(1)} = \mathbf{D}_\theta t_n^{(2)} = d^2 \text{ – минимални.}$$

Да образуваме новата оценка $t_n^{(3)} = \dfrac{1}{2}\left(t_n^{(1)} + t_n^{(2)}\right)$. Очевидно тя е неизместена, защото

$$\mathbf{E}t_n^{(3)} = \frac{1}{2}\mathbf{E}\left(t_n^{(1)} + t_n^{(2)}\right) = \frac{1}{2}\left(\mathbf{E}t_n^{(1)} + \mathbf{E}t_n^{(2)}\right) = r(\theta).$$

Дисперсията на $t_n^{(3)}$ е

$$\mathbf{D}t_n^{(3)} = \frac{1}{4}\mathbf{D}\left(t_n^{(1)} + t_n^{(2)}\right) = \frac{1}{4}\left[\mathbf{D}t_n^{(1)} + \mathbf{D}t_n^{(2)} + 2\operatorname{cov}\left(t_n^{(1)}, t_n^{(2)}\right)\right].$$

По неравенството на Коши – Шварц имаме

(1) $$\left|\operatorname{cov}\left(t_n^{(1)}, t_n^{(2)}\right)\right| = \left|\mathbf{E}\left[\left(t_n^{(1)} - r(\theta)\right)\left(t_n^{(2)} - r(\theta)\right)\right]\right|$$

$$\leqq \sqrt{\mathbf{E}\left(t_n^{(1)} - r(\theta)\right)^2}\sqrt{\mathbf{E}\left(t_n^{(2)} - r(\theta)\right)^2} = d^2.$$

Следователно

$$\mathbf{D}t_n^{(3)} \leqq \frac{1}{4}(d^2 + d^2 + 2d^2) = d^2,$$

т.е. и $t_n^{(3)}$ е неизместена оценка с минимална дисперсия, като дисперсията ѝ е равна на d^2. Това означава, че в (1) навсякъде имаме равенства. А в неравенството на Коши – Шварц равенството е възможно само когато двете функции са пропорционални, т.е. когато $t_n^{(2)} - r(\theta) = k\left(t_n^{(1)} - r(\theta)\right)$. Това равенство, използвано в (1), ще ни доведе до израза

$$\operatorname{cov}\left(t_n^{(1)}, t_n^{(2)}\right) = k\mathbf{E}\left(t_n^{(1)} - r(\theta)\right)^2 = kd^2.$$

Тъй като $\mathbf{E}\left(t_n^{(1)} - r(\theta)\right)^2 = d^2$, от горното равенство следва, че $k=1$. Тогава $t_n^{(2)} - r(\theta) = t_n^{(1)} - r(\theta)$, т.е. $t_n^{(2)} \equiv t_n^{(1)}$. Теоремата е доказана.

И така, неизместената оценка с минимална дисперсия е единствена. Това я прави предпочитана пред всички други оценки за неизвестната функция $r(\theta)$. Как да я намираме, когато съществува? Отговорът се получава като страничен резултат от доказателството на едно неравенство за дисперсиите на всички неизместени оценки, известно под името „неравенство на Рао – Крамер".

Ще предполагаме, че са изпълнени следните изисквания:

1) Неизвестният параметър θ е едномерен и явен параметър в разпределението на наблюдаваната величина ξ, т.е. θ е параметър на функцията на правдоподобие $L(x_1, \ldots, x_n \,|\, \theta) = L(x \,|\, \theta)$ на съвместната плътност на разпределение на наблюденията $\xi_1, \ldots, \xi_n$;

2) $t_n(\xi) = t(\xi_1, \ldots, \xi_n)$ е неизместена оценка за функцията $r(\theta)$;

3) $\ln L(x \,|\, \theta)$ е определен в една и съща област от стойности на $x = (x_1, \ldots, x_n)$ независимо от стойността на $\theta \in \Theta$ (Θ е областта от допустими стойности на θ);

4) $r(\theta)$ и $L(x \,|\, \theta)$ са диференцируеми функции на θ в областта Θ и допускат диференциране под знака на интеграла;

5) Съществува $\mathbf{E}_\theta \left(\dfrac{\partial}{\partial\theta} \ln L(x \,|\, \theta)\right)^2$.

Тогава е вярно следното твърдение:

Теорема 2 (неравенство на Рао – Крамер). *Ако $t_n(\xi)$ е неизместена оценка за $r(\theta)$, то при условията 1) – 5) имаме*

$$\mathbf{D}_\theta t_n(\xi) \geqq \frac{[r'(\theta)]^2}{\mathbf{E}_\theta \left(\dfrac{\partial}{\partial\theta} \ln L(x \,|\, \theta)\right)^2}. \tag{2}$$

Равенството се достига само когато е допустимо представянето

$$\frac{\partial}{\partial\theta} \ln L(x \,|\, \theta) = k(\theta)[t_n(x) - r(\theta)]. \tag{3}$$

В такъв случай дисперсията на оценката $t_n(\xi)$ от (3) е

$$\mathbf{D}_\theta t_n(\xi) = \left|\frac{r'(\theta)}{k(\theta)}\right|. \tag{4}$$

Доказателство. Започваме с диференциране по θ на равенствата (интегралът е n-кратен)

$$\int L(x \mid \theta)dx = 1; \qquad \int t_n(x)L(x \mid \theta)dx = r(\theta),$$

от които получаваме съответно

$$\int \frac{\partial}{\partial\theta}L(x \mid \theta)dx = 0; \qquad \int t_n(x)\frac{\partial}{\partial\theta}L(x \mid \theta)dx = r'(\theta).$$

Тогава

$$\begin{aligned} r'(\theta) &= \int [t_n(x) - r(\theta)]\frac{\partial}{\partial\theta}L(x \mid \theta)dx \\ &= \int [t_n(x) - r(\theta)]\left[\frac{\partial}{\partial\theta}\ln L(x \mid \theta)\right]L(x \mid \theta)dx \end{aligned} \tag{5}$$

(в последното равенство сме използвали $\dfrac{\partial L}{\partial\theta} = L\dfrac{\partial \ln L}{\partial\theta}$).

Ако положим

$$\psi(x) = [t_n(x) - r(\theta)]\sqrt{L(x \mid \theta)}, \qquad \varphi(x) = \left[\frac{\partial}{\partial\theta}\ln L(x \mid \theta)\right]\sqrt{L(x \mid \theta)}$$

и повдигнем равенствата (5) в квадрат, ще получим

$$[r'(\theta)]^2 = \left(\int \varphi(x)\psi(x)dx\right)^2.$$

Оттук по неравенството на Коши намираме

$$\begin{aligned} [r'(\theta)]^2 &\leqq \int \varphi^2(x)dx \int \psi^2(x)dx \\ &= \int [t_n(x) - r(\theta)]^2 L(x \mid \theta)dx \int \left(\frac{\partial}{\partial\theta}\ln L(x \mid \theta)\right)^2 L(x \mid \theta)dx \\ &= \mathbf{D}_\theta t_n(\xi)\mathbf{E}_\theta\left(\frac{\partial}{\partial\theta}\ln L(x \mid \theta)\right)^2. \end{aligned} \tag{6}$$

С това (2) е доказано.

Да предположим, че в (2) имаме равенство. Тогава в неравенството на Коши би трябвало да има равенство, което е възможно само ако съществува константа $k(\theta)$, независеща от x, такава че

$$\varphi(x) = k(\theta)\psi(x),$$

т.е. само когато е изпълнено (3).

Нека е изпълнено (3). Като заместим $\frac{\partial}{\partial\theta}\ln L(x\,|\,\theta)$ в (6), ще получим

$$[r'(\theta)]^2 = k^2(\theta)\,(\mathrm{D}_\theta t_n(\xi))^2 .$$

Това равенство пък е еквивалентно на (4). Теоремата е доказана.

Следствие 1. *Ако $\xi_1,\xi_2,\ldots,\xi_n$ е проста случайна извадка, за всяка неизместена оценка $t_n(\xi)$ за $r(\theta)$ е изпълнено*

$$\mathrm{D}_\theta t_n(\xi) \geqq \frac{(r'(\theta))^2}{n\mathrm{E}_\theta\left(\frac{\partial}{\partial\theta}\ln f(\xi\,|\,\theta)\right)^2}.$$

Равенство има само когато е изпълнено

$$\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln f(x_i\,|\,\theta) = k(\theta)[t_n(x) - r(\theta)],$$

където $k(\theta) = k(\theta, n)$ може да зависи от θ и n, но не зависи от x.

Доказателство. При проста случайна извадка функцията на правдоподобие е $L(x\,|\,\theta) = \prod_{i=1}^{n} f(x_i\,|\,\theta)$.

Поради независимостта на ξ_i имаме

$$\mathrm{E}_\theta\left(\frac{\partial}{\partial\theta}\ln L(\xi\,|\,\theta)\right)^2 = \mathrm{E}_\theta\left[\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln f(\xi_i\,|\,\theta)\right]^2$$

$$= \sum_{i=1}^{n}\mathrm{E}_\theta\left[\frac{\partial}{\partial\theta}\ln f(\xi_i\,|\,\theta)\right]^2 + 2\sum_{i<j}\mathrm{E}_\theta\left[\frac{\partial}{\partial\theta}\ln f(\xi_i\,|\,\theta)\frac{\partial}{\partial\theta}\ln f(\xi_j\,|\,\theta)\right]$$

$$= n\mathrm{E}_\theta\left[\frac{\partial}{\partial\theta}\ln f(\xi\,|\,\theta)\right]^2 .$$

Последното равенство следва от

$$\mathrm{E}_\theta\left[\frac{\partial}{\partial\theta}\ln f(\xi_j\,|\,\theta)\right] = \int\frac{\partial}{\partial\theta}\ln f(x\,|\,\theta).f(x\,|\,\theta)dx = \int\frac{\partial}{\partial\theta}f(x\,|\,\theta)dx$$

$$= \frac{\partial}{\partial\theta}\left[\int f(x\,|\,\theta)dx\right] = \frac{\partial}{\partial\theta}[1] = 0.$$

Следствие 2. *Ако $t_n(\xi)$ е неизместена оценка за едномерния параметър θ, конструирана върху проста случайна извадка $\xi_1, \xi_2, \dots, \xi_n$, то*

$$\mathbf{D}_\theta t_n(\xi) \geqq \left\{ n\mathbf{E}\left[\frac{\partial}{\partial\theta}\ln f(\xi\,|\,\theta)\right]^2\right\}^{-1}.$$

Доказателство. Следва от следствие 1 при $r(\theta) = \theta$.

Дефиниция. *Неизместени оценки, за които е в сила* (3) *при условията* 1) – 5), *се наричат ефективни.*

Примери. Нека $\xi_1, \xi_2, \dots, \xi_n$ са n независими наблюдения над случайната величина ξ, чието разпределение зависи от неизвестен параметър θ. Да се определи ефективна оценка за някоя функция на параметъра θ и да се намери дисперсията на тази оценка, ако:

1) ξ е геометрично разпределена случайна величина, за която $\mathbf{P}\{\xi = x\} = \theta(1-\theta)^x$, $x = 0, 1, 2, \dots$;

2) ξ е нормално разпределена случайна величина и

$$f(x\,|\,\theta) = \frac{1}{\theta\sqrt{2\pi}}\exp\left(\frac{-x^2}{2\theta^2}\right).$$

За 1) имаме

$$L(x\,|\,\theta) = \prod_{i=1}^{n}\theta(1-\theta)^{x_i} = \theta^n(1-\theta)^{x_1+\dots+x_n}.$$

Допустимо е представянето

$$\frac{\partial}{\partial\theta}\ln L(x\,|\,\theta) = \frac{n}{\theta} - \frac{x_1+\dots+x_n}{1-\theta} = \left(\frac{x_1+\dots+x_n}{n} - \frac{1-\theta}{\theta}\right)\cdot\frac{-n}{1-\theta}.$$

Следователно $\bar{\xi}_n = \dfrac{1}{n}(\xi_1+\dots+\xi_n)$ е ефективна оценка за функцията $r(\theta) = \dfrac{1-\theta}{\theta}$. Дисперсията на тази оценка според (4) е

$$\mathbf{D}\bar{\xi}_n = \frac{1-\theta}{\theta}\cdot\frac{d}{d\theta}\left(\frac{1-\theta}{\theta}\right) = \frac{1-\theta}{n\theta^2}.$$

2) Функцията на правдоподобие в случая е

$$L(x\,|\,\theta) = \prod_{i=1}^{n}\frac{1}{\theta\sqrt{2\pi}}\exp\left(\frac{-x_i^2}{2\theta^2}\right) = \left(\frac{1}{\theta\sqrt{2\pi}}\right)^n\exp\left(-\frac{1}{2\theta^2}\sum_{i=1}^{n}x_i^2\right).$$

Допустимо е представянето

$$\frac{\partial}{\partial\theta}\ln L(x\,|\,\theta) = \frac{-n}{\theta} + \frac{1}{\theta^3}\sum_{i=1}^{n} x_i^2 = \frac{-n}{\theta^3}\left(\theta^2 - \frac{1}{n}\sum_{i=1}^{n} x_i^2\right).$$

Следователно $t_n(\xi) = \frac{1}{n}\sum_{i=1}^{n}\xi_i^2$ е неизместена оценка с минимална дисперсия за θ^2, като дисперсията на тази оценка е $\mathbf{D}t_n(\xi) = \frac{2\theta^4}{n}$.

§ 7. Достатъчни статистики. Теорема за факторизацията

Ще предполагаме, че разпределението $F(x,\theta)$ на наблюдаваната случайна величина зависи явно от неизвестния параметър θ, т.е. че $F \in \mathcal{F}$ — дадено семейство от функции на разпределение, зависещо от параметъра $\theta \in \Theta$. Нека $\vec{\xi}$ е векторът на наблюденията.

Дефиниция 1. *Статистиката $t_n(\vec{\xi})$ се нарича достатъчна за семейството $\mathcal{F}$, ако условната вероятност*

$$\mathbf{P}_\theta\left\{\vec{\xi} \in A\,|\,t_n(\vec{\xi}) = t\right\} = \mathbf{P}\left\{\vec{\xi} \in A\right\} \tag{1}$$

не зависи от параметъра θ, т.е. тя е една и съща за всички функции на разпределение от $\mathcal{F}$.

Пример 1. Нека $\mathcal{F}$ е семейството от поасонови разпределения с параметър $\theta > 0$ и нека $\xi_1, \ldots, \xi_n$ е проста извадка. Тогава $t_n(\vec{\xi}) = \sum_{i=1}^{n}\xi_i$ е достатъчна статистика за $\mathcal{F}$.

Наистина функцията на правдоподобие $L(\vec{\xi}\,|\,\theta)$ в случая е

$$L(\vec{x}\,|\,\theta) = e^{-n\theta}\theta^{\sum\limits_{i=1}^{n} x_i}\left[\prod_{i=1}^{n} x_i!\right]^{-1}. \tag{2}$$

Ако $\sum_{i=1}^{n} x_i = s$, това значи, че $\xi_1 + \cdots + \xi_n = s$, вероятността за което е $(n\theta)^s e^{-n\theta}/s!$. Лесно се проверява, че условната плътност

$$(3) \qquad L\left(\vec{x},\theta \,\Big|\, \sum_{i=1}^{n}\xi_i = s\right) = \begin{cases} \dfrac{\left(\sum\limits_{i=1}^{n} x_i\right)!}{\prod\limits_{i=1}^{n} x_i!}\cdot\dfrac{1}{n^{\Sigma x_i}} & \text{при } \sum\limits_{i=1}^{n} x_i = s, \\ 0 & \text{при } \sum\limits_{i=1}^{n} x_i \neq s \end{cases}$$

не зависи от θ. Ще забележим, че (3) е полиномното разпределение за s проведени опита с еднакви вероятности $\dfrac{1}{n}$ на всеки от n-те изхода $\xi_i = x_i$.

Дефиниция 2. *Статистиката $t_n(\vec{\xi})$, $\vec{\xi} = (\xi_1, \dots, \xi_n)$, се нарича достатъчна за семейството $\mathcal{F} = \{F(x,\theta),\ \theta \in \Theta\}$, ако*

$$(4) \qquad L(\vec{x}\,|\,\theta) = h(t_n,\theta)g(\vec{x}),$$

където $g(\vec{x})$ не зависи от параметъра θ.

Дефиниция 2 показва, че функцията на правдоподобие може да се раздели на два множителя. Единият от тях, в който има зависимост от параметъра на семейството, зависи от извадката само чрез израз, определящ достатъчната статистика. Другият множител, зависещ от извадката, не съдържа параметъра θ.

Теорема 1 (за факторизацията). *Дефинициите 1 и 2 на достатъчна статистика са еквивалентни.*

Доказателство. Нека наблюдаваната случайна величина ξ има дискретно разпределение и $t_n(\vec{\xi})$ е достатъчна статистика в смисъла на (4). Тогава имаме

$$(5) \qquad \begin{aligned} \mathbf{P}\left\{\vec{\xi} = \vec{k} | t_n(\vec{\xi}) = t\right\} &= \frac{L(\vec{k}\,|\,\theta)}{\mathbf{P}_\theta\left\{t_n(\vec{\xi}) = t\right\}} = \frac{h(t_n,\theta)g(\vec{k})}{\sum\limits_{\vec{k}:t_n(\vec{k})=t} L(\vec{k}\,|\,\theta)} \\ &= \frac{h(t,\theta)g(\vec{k})}{h(t,\theta)\sum\limits_{\vec{k}:t_n(\vec{k})=t} g(\vec{k})} = \frac{g(\vec{k})}{\sum\limits_{\vec{k}:t_n(\vec{k})=t} g(\vec{k})} \end{aligned}$$

и последният израз не зависи от θ. Обратно, ако е вярно (1), тогава $\mathbf{P}\left\{\vec{\xi} = \vec{k} | t_n(\vec{\xi}) = t\right\} = h(\vec{k},t)$ няма да зависи от θ. От първото равенство на (5) виждаме, че $L(\vec{k}\,|\,\theta) = h(\vec{k},t)\mathbf{P}\left\{t_n(\vec{\xi}) = t\right\}$, което е тъждествено с представяне от вида (4). Теоремата е доказана. В непрекъснатия случай доказателството е аналогично.

Теорема 1 показва, че (4) е необходимо и достатъчно условие за съществуване на достатъчна статистика t_n. Същевременно тя дава конструктивен метод за намиране на достатъчни статистики. Според Р. Фишер статистиките, съдържащи цялата информация, дадена от експеримента, се наричат достатъчни статистики. Ако са ни известни достатъчните статистики, отделните резултати не са необходими, защото не носят повече информация. Тривиална достатъчна статистика е редицата $\vec{\xi} = (\xi_1, \ldots, \xi_n)$ от извадката. Не всяка достатъчна статистика е практически необходима. Полезни са статистиките с възможно по-малка размерност. Достатъчна статистика с минимална размерност се нарича *минимална достатъчна статистика.*

Пример 2. Ако ξ е експоненциално разпределена с параметър θ, то $t_n(\vec{\xi}) = \xi_1 + \cdots + \xi_n$ при проста извадка е достатъчна статистика, при това минимална.

Наистина имаме

$$L(\vec{x}\,|\,\theta) = \prod_{i=1}^{n} \theta e^{-\theta x_i} = \theta^n e^{-\theta(x_1+\cdots+x_n)} = \theta^n e^{-\theta t_n(\vec{x})},$$

което има вида (4) с $g(\vec{x}) = 1$.

От доказателството на теоремата може да се види, че ако статистиката $t_n(\vec{\xi})$ е достатъчна, то за всяка функция $G(t)$, задаваща взаимно еднозначно изображение на $\mathbf{R}^1$ в $\mathbf{R}^1$, статистиката $G(t_n(\vec{\xi}))$ също е достатъчна. Освен това всяка ефективна оценка същевременно е и достатъчна статистика.

Важен факт съдържа следващото твърдение.

Теорема 2. *Ако $T(\vec{\xi})$ е неизместена оценка с минимална дисперсия за $r(\theta)$, а $t_n(\xi)$ — достатъчна статистика, то съществува функция $G(t)$, такава че $T(\vec{\xi}) = G\left(t_n(\vec{\xi})\right)$.*

Доказателство. Условното математическо очакване $\mathbf{E}(\eta\,|\,\xi) = h(\xi)$ винаги е измерима функция $h(\cdot)$ на величината, намираща се в условието. Затова е изпълнено

$$\mathbf{E}_\theta\left[T(\vec{\xi})|t_n(\vec{\xi})\right] = G\left(t_n(\vec{\xi})\right).$$

Тъй като $t_n(\vec{\xi})$ е достатъчна статистика, то функцията $T_1(\vec{\xi}) = G(t_n(\vec{\xi}))$ не зависи от θ. Обаче от свойството $\mathbf{E}(\mathbf{E}(\eta\,|\,\xi)) = \mathbf{E}\eta$ получаваме, че

$$\mathbf{E}_\theta\left[T_1(\vec{\xi})\right] = \mathbf{E}_\theta\left[T(\vec{\xi})\right] = r(\theta).$$

Следователно $T_1(\vec{\xi})$ е неизместена оценка за $r(\theta)$. Ще покажем, че $\mathbf{D}_\theta T_1(\vec{\xi}) \leqq \mathbf{D}_\theta T(\vec{\xi})$, т.е че $T_1(\vec{\xi})$ е неизместена оценка с минимална дисперсия за $r(\theta)$. Поради единствеността на неизместената оценка с минимална дисперсия (вж. 10.6) ще следва, че $T_1(\vec{\xi}) = T(\vec{\xi})$.

Наистина, ако приложим неравенството на Йенсен

$$\mathbf{E}\xi^2 \geqq (\mathbf{E}\xi)^2$$

към условните математически очаквания, написани по-долу, ще получим

$$\begin{aligned}
\mathbf{D}_\theta T(\vec{\xi}) &= \mathbf{E}_\theta\left[T(\vec{\xi}) - r(\theta)\right]^2 = \mathbf{E}_\theta\left\{\mathbf{E}_\theta\left(\left[T(\vec{\xi}) - r(\theta)\right]^2 \,|\, t_n(\vec{\xi})\right)\right\} \\
&\geqq \mathbf{E}_\theta\left\{\mathbf{E}_\theta\left[\left(T(\vec{\xi}) - r(\theta)\right) |t_n(\vec{\xi})\right]\right\}^2 \\
&= \mathbf{E}_\theta\left[T_1(\vec{\xi}) - r(\theta)\right]^2 = \mathbf{D}_\theta T_1(\vec{\xi}).
\end{aligned}$$

С това теоремата е доказана.

§ 8. Методи за получаване на точкови оценки

А. Метод на максималното правдоподобие. При определянето на някои общи параметри на генералната съвкупност (отразяващи се в разпределението на наблюдаваната случайна величина) — средно, моменти, квантили, корелации, форми на зависимост и пр., значителна роля се пада на догадката и аналогията с известното от теорията на вероятностите. Съществуват обаче общи методи, които по еднакъв начин могат да се използват във всички случаи за търсене на оценки на неизвестни параметри на генералната съвкупност, когато тези параметри влияят по явен начин във функцията на разпределение на наблюдаваната величина ξ. Тогава тези параметри (нека ги означим с $\theta_1, \ldots, \theta_k$) ще фигурират и в съвместната функция на разпределение на $\xi_1, \ldots, \xi_n$ (наблюденията), т.е. във функцията на правдоподобие $L(\vec{x}\,|\,\theta) = L(x_1, \ldots, x_n; \theta_1, \ldots, \theta_k)$.

Методът на максималното правдоподобие за намиране на оценки $\hat{\theta}_i = \hat{\theta}_i(\xi_1, \ldots, \xi_n)$, $i = 1, \ldots, k$, по резултатите от наблюденията е бил известен още на немския математик К. Ф. Гаус през миналия век. Името му е дадено от английския статистик Р. Фишер, който го е изследвал системно през двадесетте години на нашия век. Идеята се гради върху зависимостта на функцията на правдоподобие от неизвестния параметър θ. В n-мерното извадково пространство за всяка фиксирана точка $\vec{x} = (x_1, \ldots, x_n)$ плътността $L(\vec{x} \mid \theta)$ може да бъде по-голяма или по-малка в зависимост от избора на θ. Методът на максималното правдоподобие препоръчва да се избере θ така, че плътността $L(\vec{x} \mid \theta)$ да приеме най-голямата си възможна стойност за дадено наблюдение $\vec{x}$. Така избраното θ зависи от $\vec{x}$ и ще го означаваме $\hat{\theta} = \hat{\theta}(\vec{x})$.

И така, същността на метода на максималното правдоподобие се изразява със следното правило: Ако във функцията на правдоподобие положим $\theta_i = \hat{\theta}_i(x_1, \ldots, x_n)$ за $i = 1, \ldots, k$, тя трябва да достига своя максимум, т.е. да е изпълнено условието

$$\hat{\theta}(\vec{x}) = \arg\max_{\theta \in \Theta} L(\vec{x} \mid \theta). \tag{1}$$

Фактически функциите $\hat{\theta}_i(\vec{x})$ се намират, като се използва дефиниционното условие (1). Тъй като максимумът на функциите $L(\vec{x} \mid \theta)$ и $\ln L(\vec{x} \mid \theta)$ се достига при едни и същи стойности на $\theta_1, \ldots, \theta_k$, търсените стойности са измежду решенията на системата уравнения

$$\frac{\partial}{\partial \theta_i} \ln L(\vec{x} \mid \theta) = 0, \quad i = 1, \ldots, k, \tag{2}$$

наречени уравнения на правдоподобието.

Вземането на $\ln L(\vec{x} \mid \theta)$ вместо директното ползване на $L(\vec{x} \mid \theta)$ е продиктувано от съображения за простота в случаите на независими наблюдения. Тогава уравненията на максималното правдоподобие са

$$\sum_{j=1}^{n} \frac{\partial f(x_j, \theta)}{\partial \theta_i} / f(x_j, \theta) = 0, \quad i = 1, \ldots, k.$$

Оценките, получени по метода на максималното правдоподобие, ще наричаме максимално правдоподобни оценки.

Пример 1. Нека $\xi_1, \dots, \xi_n$ са независими наблюдения над случайна величина ξ, разпределена по нормален закон $\mathcal{N}(a, \sigma^2)$ с неизвестен параметър $\theta = (a, \sigma^2)$. Да намерим максимално правдоподобните оценки за θ. В дадения случай функцията на правдоподобие е

$$L(\vec{x} \mid \theta) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - a)^2\right].$$

Двете уравнения на правдоподобието дават

$$\frac{\partial}{\partial a}\ln L = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - a) = 0;$$

$$\frac{\partial}{\partial \sigma}\ln L = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - a)^2 = 0,$$

откъдето

$$\hat{a} = \frac{1}{n}\sum_{i=1}^{n}x_i; \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{a})^2.$$

Следователно оценките на максимално правдоподобие за параметрите a и σ^2 са

$$\hat{a} = \bar{\xi}_n; \qquad \hat{\sigma} = \hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - \bar{\xi}_n)^2.$$

Значението на метода на максималното правдоподобие става ясно от следните теореми:

Теорема 1. *Ако за числовия параметър θ съществува ефективна оценка θ^*, уравнението на максималното правдоподобие има единствено решение θ^*.*

Доказателство. Нека съществува ефективна оценка $\theta^*(\vec{\xi})$ за числовия параметър θ. По теорема 6.2 това означава, че е изпълнено

$$\frac{\partial}{\partial\theta}\ln L(\vec{\xi} \mid \theta) = k(\theta)\left[\theta - \theta^*(\vec{\xi})\right].$$

Оттук и от условията (2), определящи максимално правдоподобните оценки $\hat{\theta}_i$, следва, че $\hat{\theta}(x_1, \dots, x_n) = \theta^*(x_1, \dots, x_n)$.

При прости извадки и някои твърде общи условия с аналитичен характер решенията на уравненията на правдоподобие са състоятелни оценки, т.е. $\hat{\theta}_n \xrightarrow{p} \theta_0$ при $n \to \infty$, където θ_0 е истинската стойност на оценявания параметър.

Теорема 2. *При прости извадки, когато са изпълнени условията за съществуване и асимптотична състоятелност на максимално правдоподобните оценки, плътността $f(x,\theta)$ има производни от втори ред по аргументите си $\theta_1,\dots,\theta_k$ и съществуват смесените моменти*

$$b_{ij}(\theta)=\mathbf{E}_\theta\left(\frac{\partial \ln f(\xi,\theta)}{\partial \theta_i}\cdot\frac{\partial \ln f(\xi,\theta)}{\partial \theta_j}\right)=\int_0^\infty \frac{\partial \ln f(x,\theta)}{\partial \theta_i}\cdot\frac{\partial \ln f(x,\theta)}{\partial \theta_j}f(x,\theta)dx,$$

то максимално правдоподобната оценка $\hat{\theta} = (\hat{\theta}_1,\dots,\hat{\theta}_k)$ е асимптотично нормална (многомерна) случайна величина. Векторът на средните на това асимптотично разпределение $\theta^0 = (\theta_1^0,\dots,\theta_k^0)$ е съставен от истинските стойности на оценяваните параметри, а ковариационната матрица е $\mathbf{A}=\frac{1}{n}\mathbf{B}_0^{-1}$, като $\mathbf{B}_0=\mathbf{B}(\theta^0)=\{b_{ij}(\theta^0)\}$.

Смисълът на тази теорема [9, с. 69] е, че при $n \to \infty$ разпределението на случайния вектор $\sqrt{n}(\hat{\theta}_n-\theta^0)$ клони към нормалното разпределение $\mathcal{N}(0,\mathbf{B}_0^{-1})$.

Доказателството на теорема 2 ще приведем в случая на едномерен неизвестен параметър θ и независими наблюдения $\xi_1,\dots,\xi_n$.

При по-голям брой наблюдения оценката $\hat{\theta}_n$ с висока вероятност е близка до оценяваната неизвестна стойност θ^0 на параметъра θ. Развитието на функцията $\frac{\partial}{\partial\theta}\ln L(\vec{x},\theta)$ в околност на θ^0 в ред на Тейлор до втория член има вида

$$\text{(3)}\quad \frac{\partial}{\partial\theta}\ln L(\vec{x},\theta) = \frac{\partial}{\partial\theta}\ln L(\vec{x},\theta^0)+(\theta-\theta^0)\frac{\partial^2}{\partial\theta^2}\ln L(\vec{x},\theta^0)+(\theta-\theta^0)o(1),$$

където $o(1) \xrightarrow{p} 0$ при $\theta \to \theta^0$. Ако в (3) положим $\vec{x}=\vec{\xi}$ и $\theta=\hat{\theta}_n$, където $\hat{\theta}_n$ е максимално правдоподобната оценка за θ_0, и развием членовете, които ще се получат вдясно на (3), ще стигнем до

$$\text{(4)}\quad 0=\sum_{i=1}^n \frac{\partial}{\partial\theta}\ln f(\xi_i,\theta^0)+(\hat{\theta}_k-\theta^0)\sum_{i=1}^n\frac{\partial^2}{\partial\theta^2}\ln f(\xi_i,\theta^0)+(\theta_k-\theta^0)o(1),$$

Равенството (4) е вярно за големи n ($n \to \infty$). Като вземем предвид, че $\frac{\partial}{\partial\theta}\ln f(\xi_i, \theta^0)$ са независими еднакво разпределени случайни величини със средна стойност

$$(5) \qquad \mathbf{E}\left[\frac{\partial}{\partial\theta}\ln f(\xi_i,\theta^0)\right] = \frac{\partial}{\partial\theta}\int f(x,\theta^0)dx = 0$$

и дисперсия $\mathbf{B}_0 = \mathbf{E}\left[\frac{\partial}{\partial\theta}\ln f(\xi_i,\theta^0)\right]^2$, ще получим, че (според централната гранична теорема)

$$(6) \qquad \frac{1}{\sqrt{n\mathbf{B}_0}}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln f(\xi_i,\theta^0) \xrightarrow[n\to\infty]{d} \eta \in \mathcal{N}(0,1).$$

От друга страна, не е трудно да се види след диференциране на равенството (5) по θ^0 и прехвърляне на един член вдясно, че

$$\begin{aligned}
\mathbf{E}\frac{\partial^2}{\partial\theta^2}\ln f(\xi_i,\theta^0) &= \int \frac{\partial^2}{\partial\theta^2}\ln f(x,\theta^0).f(x,\theta^0)dx \\
&= -\int \frac{\partial}{\partial\theta}\ln f(x,\theta^0).\frac{\partial}{\partial\theta}f(x,\theta^0)dx \\
&= -\int \left[\frac{\partial}{\partial\theta}\ln f(x,\theta^0)\right]^2 f(x,\theta^0)dx = \mathbf{B}_0.
\end{aligned}$$

Следователно по закона за големите числа за втората група събираеми вдясно на (4) е изпълнено

$$(7) \qquad \frac{1}{n\mathbf{B}_0}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\ln f(\xi_i,\theta^0) \xrightarrow[n\to\infty]{p} 1.$$

Равенството (4) можем да преобразуваме във вида

$$(8) \qquad \frac{1}{\sqrt{n\mathbf{B}_0}}\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln f(\xi_i,\theta^0)$$

$$= \sqrt{n\mathbf{B}_0}(\hat{\theta}_n - \theta^0).\frac{1}{n\mathbf{B}_0}\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\ln f(\xi_i,\theta^0) + \frac{1}{\sqrt{n\mathbf{B}_0}}(\hat{\theta}_n - \theta^0)o(1).$$

Като вземем предвид (6), (7) и факта, че

$$\frac{1}{\sqrt{n\mathbf{B}_0}}(\hat{\theta}_n - \theta^0)o(1) \xrightarrow[n\to\infty]{p} 0,$$

ще получим $\sqrt{n \mathrm{B}_0}(\hat{\theta}_n - \theta^0) \xrightarrow{d} \mathcal{N}(0,1)$, т.е. твърдението на теорема 2 след граничен преход $n \to \infty$ в (8).

Максимално правдоподобните оценки (зависещи от броя на наблюденията n) освен това са и асимптотично ефективни, т.е. имат дисперсии, които при $n \to \infty$ клонят към минимума на дисперсиите на всички възможни неизместени оценки за интересуващите ни параметри.

Б. Една евристична мотивировка на метода на максималното правдоподобие. Изложените съображения за избор на най-вероятната възможност, съответстваща на резултата от наблюденията, не обясняват защо този метод води към добри оценки за неизвестните параметри $\theta = (\theta_1, \ldots, \theta_k)$. Въз основа на закона за големите числа донякъде ще обясним целесъобразността на този принцип.

Нека наблюденията са независими. Тогава (1) означава, че търсим $\max\limits_{\theta}[f(\xi_1, \theta) \ldots f(\xi_n, \theta)]$. Без да се изменя положението на максимума, можем да преминем от дадена функция към нейния логаритъм, т.е. към задачата за $\max\limits_{\theta} \ln \prod\limits_{i=1}^{n} f(\xi_i, \theta)$. Множителят $\dfrac{1}{n}$ също не променя решението и можем да преминем към задачата за търсене на

$$\max_{\theta}\left[\frac{1}{n}\sum_{i=1}^{n} \ln f(\xi_i, \theta)\right].$$

По закона за големите числа при големи n този максимум ще бъде близък до $\mathbf{E}_\theta \ln f(\xi, \theta)$. Затова максимално правдоподобната оценка $\hat{\theta}_k = \hat{\theta}_k(\xi_1, \ldots, \xi_k)$ ще е близка до такава стойност на параметъра θ, която максимизира $\mathbf{E}_\theta \ln f(\xi, \theta)$ като функция на θ. Остава да посочим коя е тази стойност.

Нека θ^0 е истинската стойност на параметъра θ, която съответства на наблюдаваната случайна величина ξ и която се опитваме да оценим. Тогава

$$I(\theta) = \mathbf{E}_{\theta^0} \ln f(\xi, \theta) = \int f(\xi, \theta^0) \ln f(x, \theta) dx. \tag{9}$$

Остава да изследваме при кои θ величината $I(\theta)$ достига максималната си стойност.

За целта има смисъл да направим малко отклонение, тъй като задачата води към някои неравенства от теорията на информацията. Най-простият вариант на тези неравенства изглежда така:

Нека $\{p_k\}$ и $\{q_k\}$ са две дискретни разпределения (редици от положителни числа, за които $\sum\limits_{k=0}^{\infty} p_k = \sum\limits_{k=0}^{\infty} q_k = 1$). Тогава винаги е изпълнено неравенството

$$\sum_{k=0}^{\infty} p_k \ln q_k \leqq \sum_{k=0}^{\infty} p_k \ln p_k.$$

Ако гледаме на числата p_k като на постоянни, а на q_k — като на променливи, последното неравенство може да се изкаже така: Изразът $I\{q_k\} = \sum\limits_{k=0}^{\infty} p_k \ln q_k$ като функция на променливите $\{q_k\}$ достига своята максимална стойност в случая, когато $\{q_k\} = \{p_k\}$.

Доказателството на това твърдение може да се извърши по метода на неопределените множители на Лагранж. Образуваме функцията на Лагранж

$$\mathcal{L}(\{q_k\}, \lambda) = \sum_{k=0}^{\infty} p_k \ln q_k + \lambda \left(\sum_{k=0}^{\infty} q_k - 1 \right),$$

която според теоремата на Лагранж за екстремум на функция при наличие на ограничения има екстремум в същите точки, в които има екстремум $I\{q_k\}$ за редица променливи $\{q_k\}$, удовлетворяваща ограничението $\sum\limits_{k=0}^{\infty} q_k = 1$. Но екстремумите на $\mathcal{L}$ са измежду решенията на системата уравнения

$$\left| \begin{array}{l} \dfrac{\partial}{\partial q_k}\mathcal{L}(\{q_k\}, \lambda) = p_k.\dfrac{1}{q_k} + \lambda = 0, \quad k = 0, 1, 2, \ldots, \\ \dfrac{\partial}{\partial \lambda}\mathcal{L}(\{q_k\}, \lambda) = \sum\limits_{k=0}^{\infty} q_k - 1 = 0. \end{array} \right.$$

Първата група уравнения дава $q_k = -\dfrac{p_k}{\lambda}$, а от последното уравнение $\lambda = -\sum\limits_{k=0}^{\infty} p_k = -1$. Следователно екстремумът на $\mathcal{L}$ се достига при $q_k = p_k$, $k = 0, 1, 2, \ldots$ Твърдението е доказано.

Интегралният аналог на разглежданото неравенство се формулира по следния начин: За произволни вероятностни плътности $p(x)$ с еднаква дефиниционна област е изпълнено

$$\int p(x) \ln q(x) dx \leqq \int p(x) \ln p(x) dx. \tag{10}$$

Ако се върнем сега към максимално правдоподобните оценки и си спомним, че търсим такава стойност на θ, при която се достига максимумът на $I(\theta)$ от (9), от приведения резултат в (10) следва, че трябва да вземем $f(x,\theta) = f(x,\theta^0)$, т.е. да вземем $\theta = \theta^0$. И тогава максимално правдоподобната оценка $\hat{\theta}_n$ при голям брой наблюдения се оказва близка до стойността на параметъра.

Р. Фишер доказва в своите изследвания, че в определен смисъл максимално правдоподобните оценки по най-добър начин използват информацията за параметрите, която носят наблюденията $\xi_1,\ldots,\xi_n$. Тези изследвания направиха метода на максималното правдоподобие твърде популярен. Идеята за най-добра обработка на експерименталните данни е привлекателна и се реализира достатъчно пълно в теорията на статистическите решения, построена от А. Валд през 40-те години на нашия век. Ще я изясним по-късно. Тук ще споменем още веднъж, че идеята за най-добро използване на резултатите от наблюденията може да доведе и до неправилни заключения. Работата е там, че всички предположения за принадлежност на неизвестната плътност на разпределение на наблюдаваната величина към определено параметрично семейство (нормално, експоненциално, гама и кое да е друго) не винаги са изпълнени без някакви допълнителни изисквания.

В. Метод на моментите. Един от първите общи методи за оценка на неизвестните параметри в разпределението на изследваната случайна величина е методът на моментите. Той се прилага широко в статистическата практика, тъй като обикновено не води до сложни изчислителни процедури. Идеята на метода е следната:

Нека законът на разпределение на величината ξ съдържа подлежащи на оценка параметри $\theta_1,\ldots,\theta_k$, т.е. $F_\xi(x) = F(x;\theta_1,\ldots,\theta_k)$. Тогава моментите на разпределението на ξ — например първите k начални момента (предполага се, че те съществуват)

$$m_r = \int_{-\infty}^{\infty} x^r dF(x;\theta_1,\ldots,\theta_k),$$

ще са функции от параметрите $\theta_1,\ldots,\theta_k$, т.е.

$$m_r = m_r(\theta_1,\ldots,\theta_k), \qquad r = 1,2,\ldots,k. \tag{11}$$

Разглеждаме написаните равенства като система уравнения относно неизвестните $\theta_1,\ldots,\theta_k$ и предполагаме, че тази система може да се реши. Достатъчно е да вземем k момента. Нека са получени решенията $\theta_j = \theta_j(m_1,\ldots,m_k)$, $j = 1,\ldots,k$, на системата (11), които представят неизвестните параметри като функции на моментите $m_1, \dots, m_k$. Предполагаме, че тези функции са непрекъснати. Като заместим теоретичните моменти m_r с емпиричните

(12) $$\hat{\nu}_r^{(n)} = \frac{1}{n}\sum_{i=1}^{n} \xi_i^r, \qquad r = 1, \dots, k,$$

ще получим оценките

(13) $$\hat{\theta}_j = \theta_j(\hat{\nu}_1^{(n)}, \dots, \hat{\nu}_k^{(n)}).$$

Теорема 1. *Оценките* (13) *са състоятелни оценки за неизвестния параметър* θ.

Доказателството е просто следствие от закона за големите числа, според който $\hat{\nu}_r^{(n)} \xrightarrow[n\to\infty]{p} m_r$, и от това, че $\theta_j = \theta_j(m_1, \dots, m_k)$ са непрекъснати. Тогава наистина

$$\hat{\theta}_j = \theta_j(\hat{\nu}_1^{(n)}, \dots, \hat{\nu}_k^{(n)}) \xrightarrow{p} \theta_j(m_1, \dots, m_k) = \theta_j \text{ при } n \to \infty.$$

Последното твърдение следва от теоремата на Слуцки (вж. 7.1).

Пример 2. Нека A е случайно събитие с неизвестна вероятност $\theta = \mathbf{P}(A)$. Дефинираме случайната величина ξ по следния начин: $\xi = 1$, ако се осъществи A в единичен експеримент, и $\xi = 0$, ако не се осъществи. Средното на ξ е

$$m_1 = \mathbf{E}\xi = 0.\mathbf{P}(\xi = 0) + 1.\mathbf{P}(\xi = 1) = \theta.$$

Функцията $\theta = m_1$ е непрекъсната по m_1. Нека са направени n независими опита, в които величината ξ е приела стойности $\xi_1, \dots, \xi_n$. Оценката за нейния първи момент е

$$\hat{\nu}_1^{(n)} = \frac{1}{n}(\xi_1 + \dots + \xi_n),$$

откъдето оценката за θ е

(14) $$\hat{\theta}_n = \hat{\nu}_1^{(n)} = \frac{1}{n}(\xi_1 + \dots + \xi_n).$$

Теорията на оценките по метода на моментите дава само, че $\hat{\theta}_n$ е състоятелна оценка, т.е. $\hat{\theta}_n \xrightarrow{p} \theta$ при $n \to \infty$. Но можем да я изучим подробно, като забележим, че сумата $\xi_1 + \dots + \xi_n$ в (14) е биномно разпределена случайна величина с параметри n и θ. Лесно се вижда, че

$$\mathbf{E}\hat{\theta}_n = \theta, \qquad \mathbf{D}\hat{\theta}_n = \frac{\theta(1-\theta)}{n}.$$

Съответно за надеждността на оценката с точност ε получаваме долна граница

$$\mathbf{P}\left\{|\hat{\theta}_n - \theta| < \varepsilon\right\} \geqq 1 - \frac{\mathbf{D}\hat{\theta}_n}{\varepsilon^2} = 1 - \frac{\theta(1-\theta)}{n\varepsilon^2}.$$

Но θ е число между 0 и 1 и най-голямата стойност, която може да приеме функцията $\theta(1-\theta)$ в интервала $[0,1]$, е $1/4$. С помощта на тази константа можем да получим една абсолютна долна граница за надеждността на оценките на неизвестната вероятност θ със зададена точност ε. Имаме

$$\mathbf{P}\left\{|\hat{\theta}_n - \theta| < \varepsilon\right\} \geqq 1 - \left(4n\varepsilon^2\right)^{-1}.$$

Оттук винаги можем да посочим какъв минимален брой наблюдения n е необходим, за да гарантира дадена точност ε на емпирично определената вероятност $\hat{\theta}_n$ за сбъдване на дадено случайно събитие при каквато и да е предварително зададена надеждност α_0. За n е достатъчно да изберем най-малкото цяло число, за което е изпълнено неравенството $1 - \left(4n\varepsilon^2\right)^{-1} \geqq \alpha_0$.

Пример 3. Като използваме метода на моментите, да намерим оценки на параметрите α и λ на гама-разпределението, чиято плътност при $x \geqq 0$ се задава с равенството

$$f(x;\alpha,\lambda) = \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}.$$

Лесно се вижда, че

$$m_1 = \frac{\alpha}{\lambda}; \qquad m_2 = \frac{\alpha(1+\alpha)}{\lambda^2} = \frac{\alpha^2}{\lambda^2} + \frac{\alpha}{\lambda}.\frac{1}{\lambda}.$$

Оттук намираме решенията

$$\lambda = \frac{m_1}{m_2 - m_1^2}, \qquad \alpha = \frac{m_1^2}{m_2 - m_1^2}.$$

Получаваме състоятелните оценки

$$\hat{\lambda}_n = \frac{\hat{\nu}_1^{(n)}}{\hat{\nu}_2^{(n)} - \left(\hat{\nu}_1^{(n)}\right)^2}, \qquad \hat{\alpha}_n = \frac{\left(\hat{\nu}_1^{(n)}\right)^2}{\hat{\nu}_2^{(n)} - \left(\hat{\nu}_1^{(n)}\right)^2},$$

където емпиричните моменти $\hat{\nu}_1^{(n)}$ и $\hat{\nu}_2^{(n)}$ се определят по (2).

Методът на моментите се прилага успешно, когато уравненията (11) са прости и могат лесно да бъдат решени. Той не е приложим за разпределения, при които теоретичните моменти от съответния порядък не съществуват. Например за разпределението на Коши с плътност $f(x,\theta) = \theta/[\pi(\theta^2 + x^2)]$, $x \in (-\infty, \infty)$, не съществуват никакви теоретични моменти от ред, по-голям или равен на 1. За отбелязване е, че при краен брой наблюдения емпиричните моменти $\hat{\nu}_r^{(n)}$ от произволен ред са винаги крайни.

§ 9. Интервални оценки

А. Едномерен случай. Идея на подхода. Получаването или построяването на оценка на неизвестна величина по резултати от наблюдения или други експериментални данни представлява намиране на приблизителна стойност на тази неизвестна величина. Без посочване точността на приближението приблизителната стойност би била безинтересна. В практиката усилията се съсредоточават около гарантиране на определена точност на пресмятанията с възможно по-висока вероятност. От тази гледна точка разгледахме въпроса достатъчно подробно в т. Д. от § 4. Тук ще видим, че понятието точност може да се даде под формата на *доверителен интервал* или *доверително множество.*

В задачите с нестохастичен характер, където се изследват неизвестни константи, се счита за разумно да се намерят горни и долни граници за тези неизвестни. Ползата от границите е толкова по-голяма, колкото по-близки са те една до друга. По аналогичен начин в статистиката се въвеждат понятията горен и долен доверителен интервал. Удобно е, както и досега, да считаме неизвестната величина θ за параметър в разпределението на наблюдаваната случайна величина. Първо ще предположим, че θ е едномерен параметър.

Нека по наблюденията $\xi_1, \dots, \xi_n$ е получена такава статистика $T^\gamma(\vec{\xi})$, че да може да се установи верността на равенството

$$\mathbf{P}_\theta \left\{ T^\gamma(\vec{\xi}) \geqq \theta \right\} = \gamma, \tag{1}$$

където $\gamma > 0$ е достатъчно висока вероятност, различна от 1. Статистиката T^γ в този случай се нарича *горна доверителна граница* за θ с коефициент на доверие γ (горна γ-доверителна граница). Обикновено се изисква условието: Ако за две вероятности

$\gamma_1 < \gamma_2$ (гледайки на (1) като на уравнение, свързващо $T^\gamma, n, \theta, \gamma$) при еднакви θ са определени статистиките $T^{\gamma_1}(\vec{\xi})$ и $T^{\gamma_2}(\vec{\xi})$, то $T^{\gamma_1}(\vec{\xi}) \leqq T^{\gamma_2}(\vec{\xi})$, т.е. по-високият коефициент на доверие води към принудително разширяване на областта, определена от доверителната граница. Задачи от подобен род (с едностранни доверителни граници) могат да възникнат в практиката и са интересни — например, когато се оценява нов летателен апарат, процентът на брака в произведената продукция, допустимите норми на замърсеност на околната среда, праговете на токсичност на ново лекарство, вероятността за предизвикване на усложнения в организма на детето при ваксиниране и пр.

Пример 1. За да определим горна граница на степента на радиоактивност λ на дадено радиоактивно вещество, го наблюдаваме с Гайгер-Мюлеровия брояч до момента на регистрация на n-тата поредна частица. Нека ξ_i са интервалите между $(i-1)$-вата и i-тата регистрация. От общата теория на радиоактивното разпадане е известно, че ξ_i са независими, експоненциално разпределени случайни величини с параметър λ, т.е. с плътност $f(x, \lambda) = \lambda e^{-\lambda x}$, $x \geqq 0$. По наблюденията $\xi_1, \ldots, \xi_n$ ще определим горна γ-доверителна граница за λ.

Пълното време на наблюденията $T = \xi_1 + \cdots + \xi_n$ е статистика, която (покажете за упражнение!) има разпределение на Ерланг с функция на плътност

$$f_T(x) = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} \quad \text{за } x \geqq 0.$$

Величината $2\lambda T$ има χ^2-разпределение с $2n$ степени на свобода, като

$$f_{\chi^2(2n)}(x) = \frac{x^{n-1} e^{-x/2}}{2^n (n-1)!}, \quad x \geqq 0.$$

Нека е дадена доверителната вероятност γ и нека константата C_γ е определена като решение на уравнението

$$\mathbf{P}\left\{\chi^2(2n) < C_\gamma\right\} = \gamma, \quad \text{т.е.} \int\limits_0^{C_\gamma} \frac{x^{n-1} e^{-x/2}}{2^n (n-1)!} dx = \gamma.$$

Тогава неравенството $2\lambda T \leqq C_\gamma$ дава оценката $\lambda \leqq \dfrac{C_\gamma}{2T}$ за истинската степен на радиоактивност λ на изследваното вещество. Зависимостта на оценката от наблюденията е $T = \xi_1 + \cdots + \xi_n$.

По аналогичен начин се дефинира и *долна γ-доверителна граница* $T_\gamma(\vec{\xi})$ с равенството

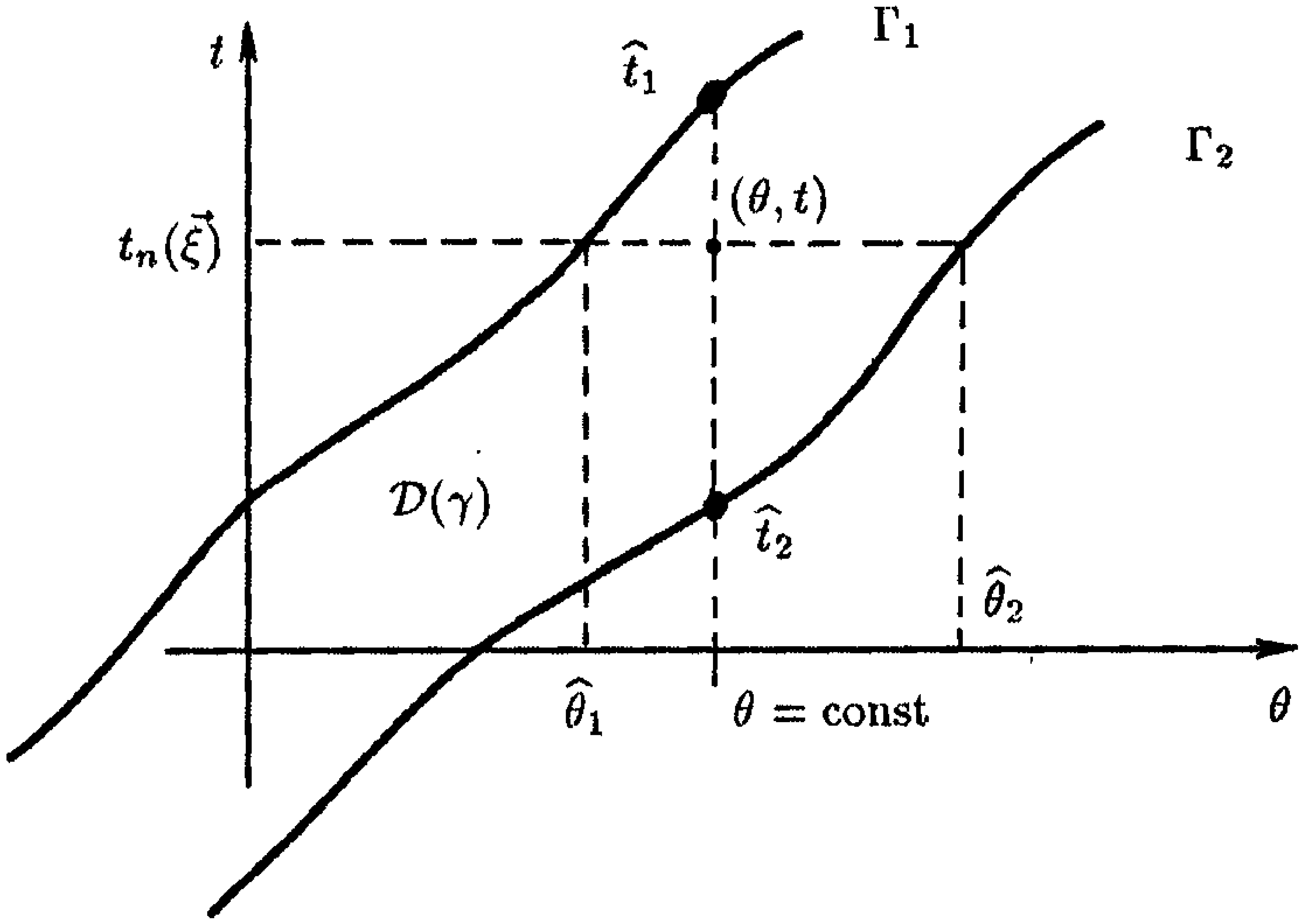
$$\mathbf{P}_\theta\left\{T_\gamma(\vec{\xi}) \leqq \theta\right\} = \gamma. \tag{2}$$

Когато вместо равенства в (1) и (2) се поиска съответните вероятности отляво да не са по-малки от γ, съответните γ-доверителни граници се наричат *консервативни*. Добро изложение на теорията на едностранните доверителни граници може да се намери в книгите на Кокс и Хинкли [11, гл. 7, § 2] и на Леман [16, с. 93 и сл.]

Ако предположим, че $T_\gamma(\vec{\xi}) \leqq T^\gamma(\vec{\xi})$, ще получим двустранна оценка за θ, или т.нар. *интервална оценка*. Интервалните оценки са особено полезен и често използван метод за теорията на статистическите изводи (вж. гл. 14). Интервалната оценка на параметъра θ означава, че е посочен интервал $[\hat{\theta}_1, \hat{\theta}_2]$ от числовата ос, за който може да се твърди с достатъчно висока вероятност, че истинската стойност на параметъра се намира вътре в този интервал. Тук границите $\hat{\theta}_1$ и $\hat{\theta}_2$ на интервала са функции на извадката $\xi_1, \ldots, \xi_n$ и следователно са случайни величини. Вместо пресмятането на едно число $\hat{\theta}$ като оценка на параметъра θ сега е определена група от числа — числата в интервала $[\hat{\theta}_1, \hat{\theta}_2]$, и едно от тези числа може би е истинската стойност на θ. Вероятността γ, с която се гарантира, че θ се намира вътре в $[\hat{\theta}_1, \hat{\theta}_2]$, се нарича *доверителна вероятност* за доверителния интервал $[\hat{\theta}_1, \hat{\theta}_2]$.

Б. Построяване на доверителни интервали. При дадена доверителна вероятност механизъм за построяване на доверителни интервали е предложен от Нейман и Уилкс [15, с. 554] и е известен като метод на доверителните интервали. Идеята е следната:

На параметъра θ се гледа като на променлива величина с допустима област на изменението Θ. Считаме, че θ е параметър на разпределението на наблюдаваната случайна величина ξ, т.е. $f_\xi(x) = f(x, \theta)$. По наблюденията е построена оценка (статистика) $t_n(\vec{\xi})$, чиято плътност на разпределение $g(t, \theta)$ е известна. При всяко $\theta \in \Theta$ функцията $g(t, \theta)$ е вероятностна плътност на разпределение, която може да се интерпретира като разпределение на единична маса по вертикалната права $\theta = \text{const}$ в равнината на променливите (θ, t) (фиг. 17).



Фиг. 17

Дадена е вероятността $\gamma \in (0,1)$. За всяко фиксирано θ съществуват такива числа $\hat{t}_1(\theta,\gamma)$ и $\hat{t}_2(\theta,\gamma)$, че събитието $\{\hat{t}_2 \leqq t_n(\vec{\xi}) \leqq \hat{t}_1\}$ да има вероятност γ, т.е. да е изпълнено

(3) $$\mathbf{P}\left\{\hat{t}_2 \leqq t_n(\vec{\xi}) \leqq \hat{t}_1\right\} = \int\limits_{\hat{t}_2}^{\hat{t}_1} g(t,\theta)dt = \gamma.$$

Очевидно $\hat{t}_1$ и $\hat{t}_2$ съществуват винаги и могат да се изберат нееднозначно. Например, ако $\varepsilon_1 > 0$ и $\varepsilon_2 > 0$ са произволни числа, за които $\varepsilon_1 + \varepsilon_2 = 1 - \gamma$, и определим $\hat{t}_1(\varepsilon_1)$ и $\hat{t}_2(\varepsilon_2)$ така, че $\varepsilon_1 = \int\limits_{\hat{t}_1}^{\infty} g(t,\theta)dt$, $\varepsilon_2 = \int\limits_{-\infty}^{\hat{t}_2} g(t,\theta)dt$, тези числа ще удовлетворяват (3). Обикновено се предпочита онзи избор, при който $\hat{t}_1$ и $\hat{t}_2$ са възможно най-близко помежду си или пък е изпълнено $g(\hat{t}_1,\theta) = g(\hat{t}_2,\theta)$ (от съображения за симетрия).

Числата $\hat{t}_1$ и $\hat{t}_2$ зависят от θ. Ако се остави θ да се мени в допустимата област Θ, точките с координати $(\theta,\hat{t}_1)$ и $(\theta,\hat{t}_2)$ ще опишат в равнината (θ,t) две криви Γ_1 и Γ_2. Да предположим, че всяка от кривите Γ_1 и Γ_2 има само по една пресечна точка с хоризонтална права $t = \text{const}$. Да означим тези точки с $\hat{\theta}_1(t)$ и $\hat{\theta}_2(t)$, т.е. $(\hat{\theta}_1, t) \in \Gamma_1$, $(\hat{\theta}_2, t) \in \Gamma_2$. Очевидно $\hat{\theta}_i = \theta_i(t, \gamma)$.

Да означим с $\mathcal{D}(\gamma)$ областта, заключена между двете криви Γ_1 и Γ_2, т.е. $\mathcal{D}(\gamma) = \{(\theta, t),\ \theta \in \Theta,\ t \in [\hat{t}_2(\theta), \hat{t}_1(\theta)]\}$.

Нека са извършени n-те наблюдения и е конструирана статистиката $t_n(\xi)$. Да разгледаме случайните събития

$$A = \left\{ \left(\theta_0, t_n(\vec{\xi})\right) \in \mathcal{D}(\gamma) \right\};$$

$$B = \left\{ \hat{t}_2(\theta_0, \gamma) \leqq t_n(\vec{\xi}) \leqq \hat{t}_1(\theta_0, \gamma) \right\};$$

$$C = \left\{ \hat{\theta}_1 \left(t_n(\vec{\xi})\right) \leqq \theta_0 \leqq \hat{\theta}_2 \left(t_n(\vec{\xi})\right) \right\}.$$

При фиксирано θ_0 (онази стойност на параметъра, която е съответна на наблюдаваната случайна величина ξ) всяко от събитията A, B и C е изпълнено за някое подмножество $S_{A,B,C} = \{\vec{x}\}_{A,B,C} \subset \mathcal{X}$ на извадковото пространство. Обаче трите събития A, B и C са еквивалентни, защото изразяват един и същ факт — принадлежност на точките $(\theta_0, t_n(\vec{x}))$ към областта $\mathcal{D}(\gamma)$. Следователно каквото и да е θ_0, случайните величини $\hat{\theta}_1$ и $\hat{\theta}_2$ са избрани така, че е изпълнено $\mathbf{P}\left\{\hat{\theta}_1 \leqq \theta_0 \leqq \hat{\theta}_2\right\} = \gamma$.

Събитието B означава, че оценката (статистиката) $t_n(\vec{\xi})$ ще приеме стойност между двете постоянни граници $\hat{t}_1$ и $\hat{t}_2$, а събитието C — че случайната величина $\hat{\theta}_1$ има по-малка стойност от истинския и неизвестен параметър θ_0, а $\hat{\theta}_2$ — по-голяма стойност. Казва се, че случайният интервал покрива фиксираната точка θ_0, каквато и да е стойността на θ_0, с вероятност, равна на γ.

Много е важно умението да се разчитат резултатите от теорията на доверителните интервали. Тук помагат познанията от теория на вероятностите. Можем да дадем следната честотна интерпретация:

Да свържем експериментите и резултатите от тях с една хипотетична схема от опити на Бернули. След многократни наблюдения и конструиране на доверителния интервал $\left[\hat{\theta}_1, \hat{\theta}_2\right]$ с доверителна вероятност γ ще считаме, че е регистриран успех ($\mathcal{У}$), ако се е сбъднало събитието $\left\{\theta_0 \in \left[\hat{\theta}_1, \hat{\theta}_2\right]\right\}$. В противен случай е регистриран неуспех (Н). Вероятността за успех в такъв единичен опит на Бернули е $\mathbf{P}(У) = \gamma$. Интерпретацията, за която става дума, е, че в достатъчно дълга редица N от по n наблюдения над величината ξ конструираният във всеки отделен случай по горния начин γ-доверителен интервал ще покрива θ_0 в $\gamma N\%$ от случаите. Твърдението, че θ_0 е заключено между числата $\hat{\theta}_1$ и $\hat{\theta}_2$, ще бъде невярно само в $(1-\gamma)N\%$ от случаите.

Доверителните интервали, които можем да получим от експерименталните данни, просто правят една равносметка за онова, което данните говорят за неизвестните параметри в дадения случай.

Тълкуването на доверителната вероятност е следното: γ е вероятността истинската стойност на неизвестния параметър θ_0, съответна на наблюдаваната величина ξ, да е в границите $\hat{\theta}_1, \hat{\theta}_2$ (които могат да се менят с извадката $\xi_1, \ldots, \xi_n$, тъй като стойностите им зависят от статистиката $t_n(\vec{\xi})$, фиг. 17). Неправилно е да се изказва твърдение от вида: Вероятността θ_0 да е заключена между фиксираните граници $\hat{\theta}_1$ и $\hat{\theta}_2$ е равна на γ. Величината θ_0 не е случайна величина (вж. пример 1).

В. Проблеми при използването на доверителните вероятности. Част от тези проблеми са свързани с факта, че доверителните граници при една и съща доверителна вероятност не са определени еднозначно. Препоръчват се правила, които дават възможно най-къси интервали. Обикновено това се постига, като се използва за $t_n(\vec{\xi})$ статистика с минимална дисперсия. От близкото разположение на $\hat{t}_1$ и $\hat{t}_2$ следва, че $\hat{\theta}_1$ и $\hat{\theta}_2$ също ще са близки. Друга група трудности се появяват при дискретно разпределени случайни величини (и статистики) поради невъзможност за точен избор на числата $\hat{t}_1$ и $\hat{t}_2$, удовлетворяващи равенството $\mathbf{P}_\theta\left\{\hat{t}_2 \leqq t_n(\vec{\xi}) \leqq \hat{t}_1\right\} = \gamma$ при дадена γ. В такива случаи се препоръчва да се премине към консервативни граници, т.е. най-късия възможен интервал $[\hat{t}_2, \hat{t}_1]$, удовлетворяващ условието $\mathbf{P}_\theta\left\{\hat{t}_2 \leqq t_n(\vec{\xi}) \leqq \hat{t}_1\right\} \geqq \gamma$ при дадена γ. Сега рискът за грешка, че $[\hat{\theta}_1, \hat{\theta}_2]$ няма да съдържа θ_0, не надминава $1-\gamma$.

Пример 2. Наблюдава се случайна величина ξ, която има неизвестно средно $\theta = \mathbf{E}\xi$ и известна дисперсия $\sigma^2 = \mathbf{D}\xi$. Постройте γ-доверителен интервал за $\mathbf{E}\xi$ по данни $\xi_1, \ldots, \xi_n$ от голяма по обем проста случайна извадка.

Знаем, че извадковото средно $\bar{\xi}_n$ в посочените условия има средно θ и дисперсия σ^2/n. Когато n е достатъчно голямо, $\bar{\xi}_n$ (по централна гранична теорема от теорията на вероятностите) е приблизително нормално разпределена със същите средно и дисперсия. Следователно за величината $\sqrt{n}(\bar{\xi}_n - \theta)/\sigma$ е изпълнено приблизителното равенство

$$\mathbf{P}\left\{\frac{\sqrt{n}(\bar{\xi}_n - \theta)}{\sigma} < x\right\} \approx \Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} \exp\left(-\frac{u^2}{2}\right) du.$$

Нека е зададена γ. Да определим $t_1(\gamma)$ и $t_2(\gamma)$ така, че

$$\gamma = \mathbf{P}\left\{t_2(\gamma) \leqq \frac{\sqrt{n}(\bar{\xi}_n - \theta)}{\sigma} \leqq t_1(\gamma)\right\} = \Phi(t_1(\gamma)) - \Phi(t_2(\gamma)).$$

Интервалът $[t_1, t_2]$ е най-къс, когато $t_2 = -t_1$ (фиг. 17) — това следва от симетричността на $\Phi(x)$. Тогава за решението $t_1(\gamma)$, като се използва, че $\Phi(-t) = 1 - \Phi(t)$, се получава уравнението (пропускаме индекса за краткост)

$$\gamma = \Phi(t) - \Phi(-t) = 2\Phi(t) - 1,$$

т.е. $t(\gamma)$ е решение на уравнението

$$\Phi(t) = \frac{1+\gamma}{2} \tag{4}$$

и то се намира от таблиците за стандартната нормална функция на разпределение $\Phi(x)$ (вж. табл. 1). Например имаме $t(\gamma) = 1,64$ при $\gamma = 0,9$, $t(\gamma) = 2,58$ при $\gamma = 0,99$. Но неравенствата

$$-t(\gamma) \leqq \frac{\sqrt{n}(\bar{\xi}_n - \theta)}{\sigma} \leqq t(\gamma)$$

са еквивалентни (след умножение със знаменателя и прехвърляне на $\bar{\xi}_n$ вляво и вдясно) на неравенствата

$$\bar{\xi}_n - \frac{t(\gamma)\sigma}{\sqrt{n}} \leqq \theta \leqq \bar{\xi}_n + \frac{t(\gamma)\sigma}{\sqrt{n}}. \tag{5}$$

Следователно $\hat{\theta}_1 = \bar{\xi}_n - t(\gamma)\sigma/\sqrt{n}$ и $\hat{\theta}_2 = \bar{\xi}_n + t(\gamma)\sigma/\sqrt{n}$ са граници на γ-доверителен интервал за неизвестното средно $\theta = \mathbf{E}\xi$.

Пример 3. При изследване на популация били измерени 900 индивида, за които предишни по-точни измервания показали, че дисперсията на интересуващата ги величина е $\sigma^2 = 225$. Емпиричното средно на извадката се оказало $\bar{\xi}_n = 103,72$. Да се построи 99%-ов доверителен интервал за истинската стойност на параметъра на извадката.

При $\gamma = 0,99$ решението на (4) е $t = 2,58$. При дадените условия може да се счита, че n е достатъчно голямо. Доверителният интервал се определя по (5) . Неговите граници са

$$\hat{\theta}_1 = 103,72 - \sqrt{900}.\frac{2,58}{\sqrt{225}} = 102,43;$$

$$\hat{\theta}_2 = 103,72 + 1,29 = 105,01.$$

З а б е л е ж к а. Нямаме право да записваме

$$\mathbf{P}\left\{102,43 \leqq \mathbf{E}\xi \leqq 105,01\right\} = 0,99.$$

В това записване означението за вероятност няма смисъл, тъй като няма случайна величина в определянето на събитието. Тълкуването на доверителния интервал $[102,43; 105,01]$ е точно даденото в края на т. Б. Не бива да се казва „вероятността $[102,43; 105,01]$ да покрива $\mathbf{E}\xi$ е равна на 0,99“. Стойността $\mathbf{E}\xi$ или е между числата 102,43 и 105,01 или не е. Тъй като стойността на $\mathbf{E}\xi$ не е известна, не знаем ще се намира ли тя в този интервал или не, но вярна е само една от двете възможности. Ако изследователите в разглеждания случай вземат много извадки от същата популация и за всяка от тях построяват подобни доверителни интервали, то в 99% от получените интервали би се намирала и истинската стойност на $\mathbf{E}\xi$.

Всеки изследовател практически използва само няколко извадки — често даже само една, и не му е неизвестно дали построеният от него доверителен интервал ще покрива или не търсения параметър. Ако доверителната вероятност обаче е достатъчно висока, изследователят може с голяма увереност да счита, че този току-що получен от него интервал съдържа стойността на параметъра. Това е основателна причина за избор на големи стойности за доверителните вероятности (0,9; 0,95; 0,99 и т.н.), обаче един разсъдлив изследовател винаги ще допуска и незначителна вероятност (0,1; 0,05 и др.), че неговото твърдение може да бъде невярно, ако иска да получава смислени резултати.

§ 10. Примери за доверителни интервали

Техниката на построяване на доверителен интервал за едномерен параметър има проста структура. Грубо тя може да се изрази така: Построява се функция от наблюденията $t(\xi_1,\dots,\xi_n;\theta)$ и от неизвестния параметър, която да има поне при $n\to\infty$ известно разпределение, вече независещо от θ, т.е. $F_t(x)=\mathbf{P}\{t(\vec{\xi},\theta)<x\}$ да не зависи от θ. При дадена γ се определят границите $t_1(\gamma)$ и $t_2(\gamma)$ по такъв начин, че $\mathbf{P}\{t_2(\gamma)\leqq t(\vec{\xi},\theta)\leqq t_1(\gamma)\}=F_t(t_1)-F_t(t_2)=\gamma$. Обикновено неравенствата

$$t_2(\gamma)\leqq t(\xi_1,\dots,\xi_n;\theta)\leqq t_1(\gamma) \tag{1}$$

могат да се решават спрямо неизвестния параметър θ и тези решения водят до еквивалентните на (1) неравенства

$$\hat{\theta}_1(\vec{\xi})\leqq\theta\leqq\hat{\theta}_2(\vec{\xi}). \tag{2}$$

Тъй като неравенствата (1) имат вероятност γ, то и еквивалентните им неравенства (2) са изпълнени със същата вероятност, т.е. $[\hat{\theta}_1,\hat{\theta}_2]$ е един γ-доверителен интервал за θ.

По-късно теорията на проверките на статистическите хипотези ще ни даде още един метод за построяване на доверителни интервали. Следващите примери са илюстрация на посочената по-рано обща техника. При това познанията от теория на вероятностите за разпределения на функции от случайни величини и граничните теореми са ключ към успешното решаване на този кръг задачи.

Пример 1. Случайната величина ξ е разпределена по експоненциален закон $f_\xi(x)=\theta e^{-\theta x}$, $x>0$. По дадена проста случайна извадка $\xi_1,\dots,\xi_n$ да се построи γ-доверителен интервал за θ:

а) при малки n; б) при големи n.

а) Случайната величина $2n\theta\bar{\xi}_n=2\theta(\xi_1+\dots+\xi_n)$ има разпределение на Ерланг с n фази, което е χ^2-разпределение с $2n$ степени на свобода. Плътността му е (вж. пример 1 от § 9)

$$f_{\chi^2(2n)}(x)=\frac{x^{n-1}e^{-x/2}}{2^n(n-1)!}\quad \text{за } x>0.$$

Следователно за дадена доверителна вероятност γ и числа $\chi_{1,\gamma}$ и $\chi_{2,\gamma}$, определени с условията

$$\mathbf{P}\{\chi_{1,\gamma}\geqq\chi^2(2n)\}=1-\frac{\gamma}{2};\quad \mathbf{P}\{\chi^2(2n)\geqq\chi_{2,\gamma}\}=1-\frac{\gamma}{2},$$

ще е изпълнено

$$\mathbf{P}\{\chi_{1,\gamma} \leqq \chi^2(2n) \leqq \chi_{2,\gamma}\} = \gamma. \tag{3}$$

Вътрешните неравенства в (3) са еквивалентни на неравенствата

$$\chi_{1,\gamma} \geqq 2n\theta\bar{\xi}_n \geqq \chi_{2,\gamma},$$

от които следва, че γ-доверителният интервал $[\hat{\theta}_1, \hat{\theta}_2]$ за параметъра θ е определен с неравенствата

$$\hat{\theta}_1 = \frac{\chi_{1,\gamma}}{2n\bar{\xi}_n} \leqq \theta \leqq \frac{\chi_{2,\gamma}}{2n\bar{\xi}_n} = \hat{\theta}_2.$$

б) При достатъчно големи n използваме, че $\chi^2(2n)$ е приблизително нормално разпределена със средно $2n$ и дисперсия $4n$, т.е. при $n \to \infty$

$$\frac{\chi^2(2n) - 2n}{\sqrt{4n}} \xrightarrow{d} \mathcal{N}(0,1).$$

Следователно, ако подберем $t_{1,\gamma}$ и $t_{2,\gamma}$ по такъв начин, че

$$\gamma = \mathbf{P}\left\{t_{1,\gamma} \leqq \frac{2n\theta\bar{\xi}_n - 2n}{\sqrt{4n}} \leqq t_{2,\gamma}\right\} \approx \Phi(t_{2,\gamma}) - \Phi(t_{1,\gamma})$$

(тук $\Phi(t)$ е стандартната нормална функция на разпределение), то вътрешните неравенства ще са еквивалентни на следните неравенства за θ:

$$\frac{\sqrt{n} + t_{1,\gamma}}{\bar{\xi}_n\sqrt{n}} \leqq \theta \leqq \frac{\sqrt{n} + t_{2,\gamma}}{\bar{\xi}_n\sqrt{n}}.$$

При $t_{1,\gamma} = -t_{2,\gamma} = -t_{(1+\gamma)/2}$, където $t_{(1+\gamma)/2}$ е решение на уравнението $\Phi(t) = (1+\gamma)/2$, се получава най-късият интервал $[t_{1,\gamma}; t_{2,\gamma}]$. Следователно γ-доверителният интервал $[\hat{\theta}_1, \hat{\theta}_2]$ за θ при големи n се получава при

$$\hat{\theta}_1 = \frac{\sqrt{n} - t_{(1+\gamma)/2}}{\bar{\xi}_n\sqrt{n}}; \quad \hat{\theta}_2 = \frac{\sqrt{n} + t_{(1+\gamma)/2}}{\bar{\xi}_n\sqrt{n}}.$$

Пример 2. Нека ξ е с поасоново разпределение

$$\mathbf{P}\{\xi = x\} = \theta^x e^{-\theta}/x!, \quad x = 0, 1, 2, \ldots,$$

и $\xi_1, \ldots, \xi_n$ е проста случайна извадка с разпределението на ξ. Да се построи γ-доверителен интервал за θ при големи n.

При големи n според централната гранична теорема величината $\frac{(\xi_1+\cdots+\xi_n)-n\theta}{\sqrt{n\theta}}$ има асимптотично нормално разпределение $\mathcal{N}(0,1)$. Ако $t_{(1+\gamma)/2}$ е избрана така, че

$$\Phi(t_{(1+\gamma)/2}) - \Phi(-t_{(1+\gamma)/2}) \approx \mathbf{P}\left\{-t \leqq \frac{(\xi_1+\cdots+\xi_n)-n\theta}{\sqrt{n\theta}} \leqq t\right\} = \gamma,$$

неравенствата в скобите са еквивалентни на $\hat{\theta}_1 \leqq \theta \leqq \hat{\theta}_2$, където

$$\hat{\theta}_1 = \bar{\xi}_n + \frac{t^2}{2n} - \sqrt{\frac{\bar{\xi}_n t}{n} + \frac{t^4}{4n^2}}; \qquad \hat{\theta}_2 = \bar{\xi}_n + \frac{t^2}{2n} + \sqrt{\frac{\bar{\xi}_n t}{n} + \frac{t^4}{4n^2}},$$

с $t = t_{(1+\gamma)/2}$, $\Phi(t) = (1+\gamma)/2$. Следователно $[\hat{\theta}_1, \hat{\theta}_2]$ е един γ-доверителен интервал за неизвестния параметър θ.

Пример 3. Случайната величина ξ е нормално разпределена $\mathcal{N}(\theta,\sigma^2)$, където неизвестно е само математическото очакване θ, а дисперсията σ^2 на ξ е известна. При дадени n независими наблюдения $\xi_1,\ldots,\xi_n$ да се построи γ-доверителен интервал за θ.

Знае се (вж. зад. 6.5), че сумата на независими нормално разпределени случайни величини е нормална със средно — сумата от средните, и дисперсия — сумата от дисперсиите. Затова

$$t_n(\vec{\xi}) = \frac{n\bar{\xi}_n - n\theta}{\sigma\sqrt{n}} \in \mathcal{N}(0,1).$$

Ако определим числото $t_{(1+\gamma)/2}$ по такъв начин, че

$$2\Phi(t) - 1 = \mathbf{P}\left\{-t \leqq t_n(\vec{\xi}) \leqq t\right\} = \gamma,$$

то γ-доверителният интервал за θ ще бъде

$$\bar{\xi}_n - t_{(1+\gamma)/2}\frac{\sigma}{\sqrt{n}} \leqq \theta \leqq \bar{\xi}_n + t_{(1+\gamma)/2}\frac{\sigma}{\sqrt{n}}.$$

Пример 4. Случайната величина ξ е нормално разпределена $\mathcal{N}(a,\theta^2)$, където $a = \mathbf{E}\xi$ е известна величина, а $\theta^2 = \mathbf{D}\xi$ — неизвестна. Постройте γ-доверителен интервал за θ^2 по дадени n независими наблюдения $\xi_1,\ldots,\xi_n$ над ξ.

Известно е, че нормираните наблюдения $\tilde{\xi}_i = (\xi_i - a)/\theta$ са стандартно нормално разпределени $\mathcal{N}(0,1)$. Величината

$$\chi^2(n) = \sum_{i=1}^{n} \tilde{\xi}_i^2 = \frac{\sum\limits_{i=1}^{n}(\xi_i - a)^2}{\theta^2}$$

има χ^2-разпределение с n степени на свобода. Ако определим $\chi^2_{1-\gamma/2}$ и $\chi^2_{\gamma/2}$ с равенствата

$$\text{(4)} \qquad \mathbf{P}\left\{\chi^2(n) \leqq \chi^2_{1-\gamma/2}\right\} = 1-\gamma/2; \quad \mathbf{P}\left\{\chi^2(n) \geqq \chi^2_{\gamma/2}\right\} = 1-\gamma/2,$$

то ще е изпълнено

$$\mathbf{P}\left\{\chi^2_{1-\gamma/2} \leqq \chi^2(n) \leqq \chi^2_{\gamma/2}\right\} = \gamma.$$

Вътрешните неравенства в последния израз са еквивалентни на следните неравенства за θ^2:

$$\frac{\sum\limits_{i=1}^{n}(\xi_i - a)^2}{\chi^2_{\gamma/2}} \leqq \theta^2 \leqq \frac{\sum\limits_{i=1}^{n}(\xi_i - a)^2}{\chi^2_{1-\gamma/2}}.$$

Следователно тези неравенства определят един γ-доверителен интервал за θ^2.

Пример 5. Случайната величина ξ е нормално разпределена $\mathcal{N}(\theta_1, \theta_2^2)$ с неизвестни средна стойност θ_1 и дисперсия θ_2^2. По дадени независими наблюдения $\xi_1, \ldots, \xi_n$ над ξ постройте γ-доверителни интервали за θ_1 и θ_2^2.

а) Ще използваме факта, че когато ξ е нормално разпределена, извадковото средно $\bar{\xi}_n$ и извадковата дисперсия

$$\hat{s}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\xi_i - \bar{\xi}_n)^2$$

са независими случайни величини. При това $\dfrac{\sqrt{n}(\bar{\xi}_n - \theta_1)}{\theta_2}$ е нормална $\mathcal{N}(0,1)$, а $\dfrac{n\hat{s}_n^2}{(n-1)\theta_2^2}$ е разпределена $\chi^2(n-1)$ (вж. § 5). Отношението $t(n-1) = \dfrac{\sqrt{n-1}(\bar{\xi}_n - \theta_1)}{\hat{s}_n}$ има разпределение на Стюдънт с $n-1$ степени на свобода. Неговата плътност е

$$f_{t(n-1)}(x) = \Gamma\left(\frac{n}{2}\right)\left(1 + \frac{x}{n-1}\right)^{-\frac{n}{2}} \Big/ \left[\Gamma\left(\frac{n-1}{2}\right)\sqrt{(n-1)\pi}\right].$$

Ако определим числата $t_{1,\gamma}$ и $t_{2,\gamma}$ по такъв начин, че

$$\mathbf{P}\left\{t_{1,\gamma} \leqq t(n-1) \leqq t_{2,\gamma}\right\} = \gamma,$$

то неравенствата в скобите ще са еквивалентни на следните неравенства за θ_1:

$$\bar{\xi}_n - t_{2,\gamma}\frac{\hat{s}_n}{\sqrt{n-1}} \leqq \theta_1 \leqq \bar{\xi}_n - t_{1,\gamma}\frac{\hat{s}_n}{\sqrt{n-1}}.$$

Най-късият доверителен интервал се получава при $t_{1,\gamma} = -t_{2,\gamma}$ и в практиката се използва именно този случай.

б) За оценка на θ_2^2 ще използваме вече споменатия факт, че $\frac{n\hat{s}_n^2}{(n-1)\theta_2^2}$ има χ^2-разпределение с $n-1$ степени на свобода. Ако определим $\chi^2_{1-\gamma/2}$ и $\chi^2_{\gamma/2}$ както в (4), но за случайната величина $\chi^2(n-1)$, вътрешните неравенства в $\mathbf{P}\left\{\chi^2_{1-\gamma/2} \leqq \chi^2(n-1) \leqq \chi^2_{\gamma/2}\right\} = \gamma$ ще са еквивалентни на следните неравенства за θ_2^2:

$$\frac{n\hat{s}_n^2}{\chi^2_{\gamma/2}} \leqq \theta_2^2 \leqq \frac{n\hat{s}_n^2}{\chi^2_{1-\gamma/2}}.$$

Следователно γ-доверителният интервал $[\hat{\theta}_2', \hat{\theta}_2'']$ за θ_2^2 се определя с равенствата

$$\hat{\theta}_2' = \frac{n\hat{s}_n^2}{\chi^2_{\gamma/2}}; \qquad \hat{\theta}_2'' = \frac{n\hat{s}_n^2}{\chi^2_{1-\gamma/2}}.$$

Пример 6. Нека A е случайно събитие с неизвестна вероятност. Извършени са n независими повторни опита при еднакви условия, при които A се е сбъднало ν_n пъти. Определете асимптотичен γ-доверителен интервал за $p = \mathbf{P}(A)$ при $n \to \infty$.

Величината $\hat{p} = \frac{\nu_n}{n}$ е неизместена оценка за p, която по теоремата на Лаплас при $n \to \infty$ е асимптотично нормална $\mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$. Следователно изпълнено е асимптотичното (по разпределение) равенство

$$\frac{\frac{\nu_n}{n} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim \mathcal{N}(0,1).$$

Ако определим числото $t_{(1+\gamma)/2}$ от уравнението

$$2\Phi(t) - 1 \approx \mathbf{P}\left\{ -t \leqq \frac{\frac{\nu_n}{n} - p}{\sqrt{\frac{p(1-p)}{n}}} \leqq t \right\} = \gamma,$$

неравенствата в големите скоби са еквивалентни на неравенствата

$$p - t_{(1+\gamma)/2}\sqrt{\frac{p(1-p)}{n}} \leqq \hat{p} \leqq p + t_{(1+\gamma)/2}\sqrt{\frac{p(1-p)}{n}}. \tag{5}$$

Решени относно p, тези неравенства дават

$$\frac{n}{n + t_{(1+\gamma)/2}}\left[\hat{p} + \frac{t^2_{(1+\gamma)/2}}{2n} - t_{(1+\gamma)/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{t_{(1+\gamma)/2}}{4n}}\right] \tag{6}$$

$$\leqq p \leqq \frac{n}{n + t_{(1+\gamma)/2}}\left[\hat{p} + \frac{t^2_{(1+\gamma)/2}}{2n} + t_{(1+\gamma)/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{t_{(1+\gamma)/2}}{4n}}\right].$$

В съответствие с теорията от § 9 областта $\mathcal{D}$ тук е ограничена от кривите $\hat{p} = p \pm t\sqrt{\frac{p(1-p)}{n}}$. Това са двете половини на елипса, минаваща през точките $(0,0)$ и $(1,1)$, с голяма полуос по диагонала на единичния квадрат в първи квадрант (фиг. 18). Фактът, че точката $(p, \hat{p})$ лежи вътре в елипсата, е изразен с неравенството (5), като е казано, че $\hat{p}$ е в границите $p \pm t\sqrt{\frac{p(1-p)}{n}}$, $t = t_{(1+\gamma)/2}$.

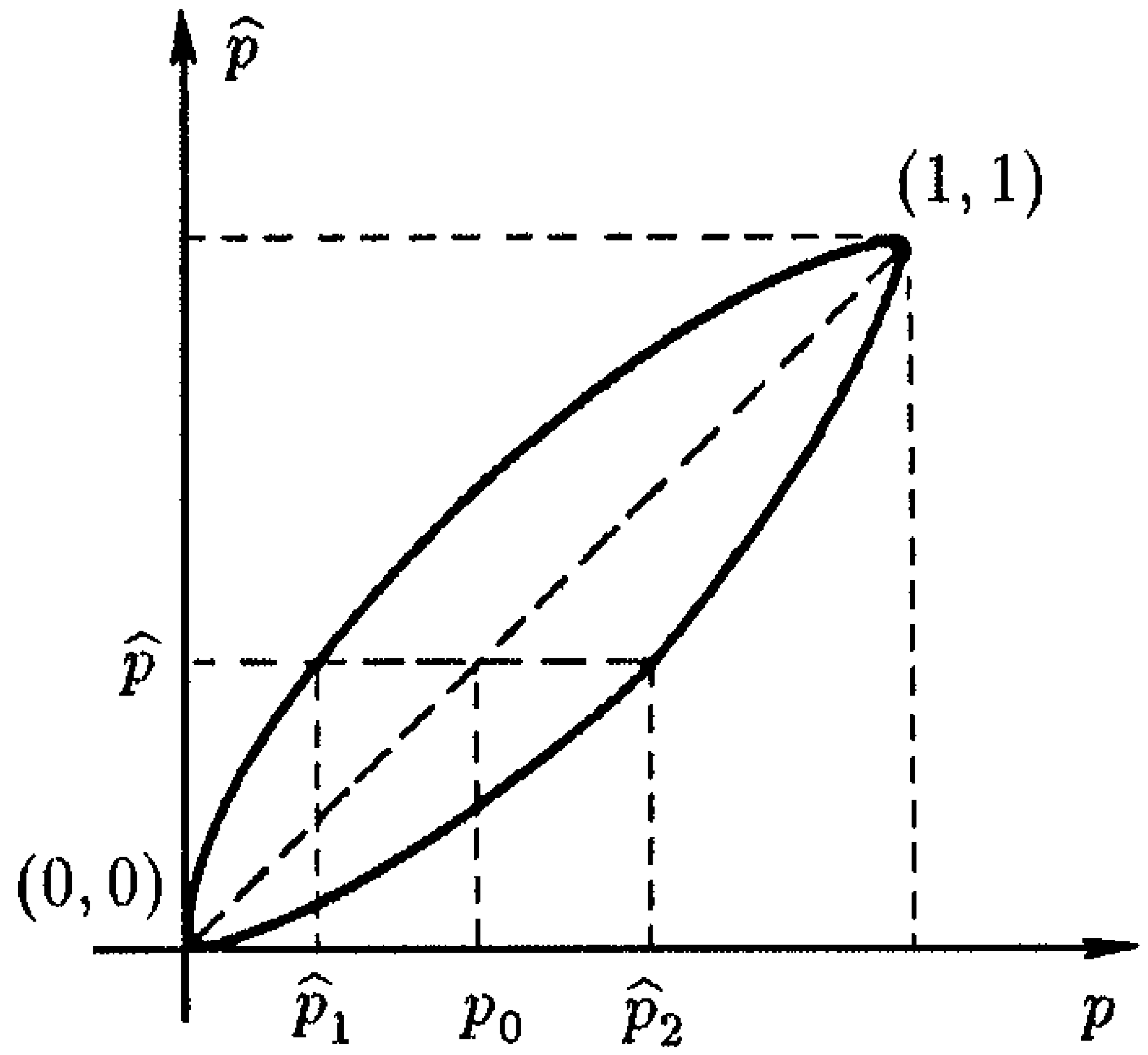
Еквивалентното твърдение (6) пък показва, че p_0 се намира между случайните граници $[\hat{p}_1, \hat{p}_2]$, определени от местоположението на случайната права $\hat{p}$.

Задачи

1. При проверка на 1000 изделия 46 от тях се оказали дефектни:

а) определете вероятността, че оценката $\hat{p} = 0,046$ за неизвестната вероятност p отделно взето изделие да се окаже дефектно се отличава от p не повече от 0,005;

б) определете доверителен интервал за неизвестната вероятност p при доверителна вероятност, не по-малка от 0,98 (използвайте пример 10.4);



Фиг. 18

в) какъв минимум изделия трябва да бъдат проверени, та вероятността оценката $\hat{p}$ да се различава от p не повече от 0,001 да не е по-малка от 0,95.

Отг. а) 0,52; в) $n \geqq 152000$.

2. Нека $\xi_1, \ldots, \xi_n$ е проста случайна извадка от експоненциално разпределена случайна величина ξ с плътност на разпределение $f(x) = \lambda^{-1} \exp(-x/\lambda)$. Покажете, че ако

$$\bar{\xi}_n = \frac{\xi_1 + \cdots + \xi_n}{n}; \qquad \hat{\nu}_2^{(n)} = \frac{\xi_1^2 + \cdots + \xi_n^2}{n}$$

и $\xi_{(1)} \leqq \cdots \leqq \xi_{(n)}$ е вариационният ред на наблюденията $\xi_1, \ldots, \xi_n$, то $\bar{\xi}_n$, $\sqrt{2/\hat{\nu}_2^{(n)}}$ и $[\ln 2]/\xi_{([n/2])}$ са състоятелни оценки за неизвестния параметър λ.

Упътване. Използвайте, че $\mathbf{E}\xi = \lambda; \mathbf{E}\xi^2 = 2/\lambda$, медианата на експоненциалното разпределение е равна на $[\ln 2]/\lambda$.

3. Нека ξ е равномерно разпределена в интервала $[\theta_1, \theta_1 + \theta_2]$. Докажете, че ако $\xi_1, \ldots, \xi_n$ са независими наблюдения над ξ, то

$$\hat{\theta}_1 = \hat{\nu}_1^{(n)} - \sqrt{3\left[\hat{\nu}_2^{(n)} - \left(\hat{\nu}_1^{(n)}\right)^2\right]} \quad \text{и} \quad \hat{\theta}_2 = 2\sqrt{3\left[\hat{\nu}_2^{(n)} - \left(\hat{\nu}_1^{(n)}\right)^2\right]}$$

са състоятелни оценки за θ_1 и θ_2.

4. Нека $\xi_1, \ldots, \xi_n$ са независими наблюдения над случайната величина ξ. Покажете:

а) ако $\mathbf{P}\{\xi = k\} = pq^k$, $k = 0, 1, 2, \ldots$; $p \in (0, 1)$, $q = 1 - p$, то $2(\bar{\xi}_n + 1)$ е ефективна оценка за $\tau(p) = 2/p$ с дисперсия $4/(np^2 q)$;

б) ако $\mathbf{P}\{\xi = k\} = \lambda^k e^{-\lambda}/k!$, $k = 0, 1, 2, \ldots$; $\lambda > 0$, то $\hat{\nu}_3^{(n)} - 3\hat{\nu}_2^{(n)} + 2\hat{\nu}_1^{(n)}$ е неизместена и състоятелна оценка за λ^3;

в) ако $f(x) = x^{\alpha-1}e^{-x}/\Gamma(\alpha)$, $x \geqq 0$, то $t_n(\vec{\xi}) = \sum_{i=1}^{n}(\xi_i - \bar{\xi}_n)/(n-1)$ е неизместена оценка за α.

5. Случайната величина ξ приема две стойности 0 и 1 съответно с вероятности

$$f(1,0) = \mathbf{P}_\theta\{\xi = 1\} = \begin{cases} \theta, & \text{ако } \theta \text{ е рационално;} \\ 1-\theta, & \text{ако } \theta \text{ е ирационално} \end{cases}$$

и

$$f(0,\theta) = \mathbf{P}_\theta\{\xi = 0\} = 1 - \mathbf{P}_\theta\{\xi = 1\} \quad \text{за} \quad \theta \in (0,1).$$

Нека $\xi_1, \ldots, \xi_n$ са независими наблюдения над θ и нека $S_n = \xi_1 + \cdots + \xi_n$. Покажете, че максимално правдоподобните оценки за θ са:

$$\begin{cases} \hat{\theta}_1 = S_n/n, & \text{когато } \theta \text{ е рационално;} \\ \hat{\theta}_2 = 1 - S_n/n, & \text{когато } \theta \text{ е ирационално,} \end{cases}$$

обаче те не са състоятелни оценки за θ. (Това е пример за максимално правдоподобни оценки, които не са състоятелни. Причината е, че плътността на разпределение $f(x,\theta)$ не е непрекъсната и диференцируема по θ.)

6. Покажете, че ако $\mathbf{E}\xi^{2r} < \infty$, математическото очакване и дисперсията на началния емпиричен момент $\hat{\nu}_r^{(n)}$ са съответно $\mathbf{E}\xi^2$ и $\left[\mathbf{E}\xi^{2r} - (\mathbf{E}\xi^r)^2\right]/n$.

7. Нека $\xi_1, \ldots, \xi_n$ са независими наблюдения над случайни величини с едно и също средно $\mathbf{E}\xi_i = a$ и различни, но известни дисперсии $\sigma_i^2 = \mathbf{D}\xi_i$. Покажете, че неизместена оценка с минимална дисперсия на неизвестния параметър a от вида $\hat{a} = \sum_{i=1}^{n} c_i\xi_i$ се достига, когато $c_i = \sigma_i^{-2}/(\sigma_1^{-2} + \cdots + \sigma_n^{-2})$.

8. Нека ν_n е броят на успехите в n опита на Бернули с вероятност p за успех в единичен опит. Направено е едно наблюдение над ν_n. Покажете, че ν_n/n е максимално правдоподобна оценка за p, която е неизместена и състоятелна.

9. Нека $\xi_{(1)} \leqq \ldots \leqq \xi_{(n)}$ е вариационният ред, построен върху проста случайна извадка $\xi_1, \ldots, \xi_n$ от стойности на случайната величина ξ. Покажете, че:

а) ако ξ е равномерно разпределена в интервала $[\theta_1, \theta_2]$, то

а$_1$) $\hat{\theta}_1 = \xi_{(1)}$ и $\hat{\theta}_2 = \xi_{(n)}$ са изместени оценки за θ_1 и θ_2 съответно със средни $\theta_1 + (\theta_2 - \theta_1)/(n+1)$, $\theta_2 - (\theta_2 - \theta_1)/(n+1)$ и дисперсии $\mathbf{D}\hat{\theta}_1 = \mathbf{D}\hat{\theta}_2 = n(\theta_2 - \theta_1)^2/[(n+1)^2(n+2)]$;

а$_2$) $\mathbf{E}\left[\dfrac{\hat{\theta}_1 + \hat{\theta}_2}{2}\right] = \dfrac{\theta_1 + \theta_2}{2}, \quad \mathbf{D}\left[\dfrac{\hat{\theta}_1 + \hat{\theta}_2}{2}\right] = \dfrac{\theta_2 - \theta_1}{2(n+1)^2(n+2)}$;

б) ако ξ е експоненциално разпределена с плътност $f(x) = \lambda e^{-\lambda x}$, $\lambda > 0$, $x \geq 0$, то

б$_1$) $\mathbf{E}\xi_{(1)} = (\lambda n)^{-1}$; $\qquad \mathbf{D}\xi_{(1)} = (\lambda n)^{-2}$;

$$\mathbf{E}\xi_{(n)} = \frac{1}{\lambda}\sum_{i=1}^{n}\frac{1}{i}; \qquad \mathbf{D}\xi_{(n)} = \frac{1}{\lambda^2}\sum_{i=1}^{n}\frac{1}{i^2};$$

б$_2$) $$\mathbf{E}\left[\left(\xi_{(1)} + \xi_{(n)}\right)/2\right] = \frac{1}{2\lambda}\left(\sum_{i=1}^{n-1}\frac{1}{i} + \frac{2}{n}\right);$$

$$\mathbf{D}\left[\left(\xi_{(1)} + \xi_{(n)}\right)/2\right] = \frac{1}{4\lambda^2}\left(\sum_{i=1}^{n-1}\frac{1}{i^2} + \frac{4}{n^2}\right).$$

10. Покажете, че ако $\xi_1, \ldots, \xi_n$ са независими наблюдения над случайната величина ξ, разпределена по съответния параметричен закон, то максимално правдоподобните оценки на посочените параметри са следните:

а) за геометричното разпределение $\mathbf{P}\{\xi = k\} = p(1-p)^k$, $k = 0, 1, 2, \ldots$:

по едно наблюдение — $\hat{p} = 1/(1+\xi)$;

по n наблюдения — $\hat{p} = 1/(1+\bar{\xi}_n)$;

б) за отрицателното биномно разпределение $\mathbf{P}\{\xi = k\} = \binom{-r}{k}p^r(1-p)^k$, $k = 0, 1, 2, \ldots$:

по едно наблюдение — $\hat{p} = r/(r+\xi)$ (изместена);

в) за разпределението на Паскал $\mathbf{P}\{\xi = k\} = \binom{-r}{k-r}p^r(1-p)^k$, $k = 0, 1, 2, \ldots$:

по едно наблюдение — $\hat{p} = r/\xi$ (изместена);

г) за поасоновото разпределение $\mathbf{P}\{\xi = k\} = \lambda^k e^{-\lambda}/k!$, $k = 0, 1, 2, \ldots$:

$\hat{\lambda} = \bar{\xi}_n$ (неизместена оценка);

д) за експоненциалното разпределение $f_\xi(x) = e^{-x/\lambda}/\lambda$, $x \geqq 0$:

$\hat{\lambda} = \bar{\xi}_n$ (неизместена оценка);

е) за нормалното разпределение $f_\xi(x) = \dfrac{1}{\sigma\sqrt{2\pi}}\exp\left[-\dfrac{(x-a)^2}{2\sigma^2}\right]$:

$$\hat{a} = \bar{\xi}_n; \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - \bar{\xi}_n)^2 \text{ (изместена);}$$

ж) за разпределението на Парето $f_\xi(x) = Cx^{-C-1}$, $x \geqq 1$, максимално правдоподобната оценка за C^{-1} е

$$\widehat{C^{-1}} = \frac{1}{n}(\ln\xi_1 + \cdots + \ln\xi_n);$$

з) за разпределението на Вайбул $f_\xi(x) = cb^{-c}x^{c-1}\exp\left[-\left(\frac{x}{b}\right)^c\right]$, $x \geqq 0$, максимално правдоподобните оценки $\hat{c}$ и $\hat{b}$ са решения на системата уравнения

$$\hat{b} = \left[\frac{1}{n}\sum_{i=1}^{n}\xi_i^{\hat{c}}\right]^{1/\hat{c}}, \quad \hat{c} = n\left[\frac{\sum_{i=1}^{n}\xi_i^{\hat{c}}\ln\xi_i}{\hat{b}^{\hat{c}}} - \sum_{i=1}^{n}\ln\xi_i\right]^{-1};$$

и) за равномерното разпределение $f_\xi(x) = 1/\theta$, $x \in [0,\theta]$:

$$\hat{\theta} = \max(\xi_1,\dots,\xi_n) \text{ (състоятелна оценка за } \theta).$$

11. Покажете, че за изброените закони на разпределение оценките за параметрите, получени по метода на моментите, са следните:

а) за дискретно равномерно разпределение $\mathbf{P}\{\xi = k\} = \frac{1}{b}$, където $k = a, a+1, \dots, a+b-1$:

$$\hat{a} = \bar{\xi}_n - \frac{1}{2}\left(2\sqrt{3\hat{\sigma}_n^2+1} - 1\right), \quad \hat{b} = 2\sqrt{3\hat{\sigma}_n^2+1};$$

б) за непрекъснато равномерно разпределение $f_\xi(x) = \frac{1}{b}$, $x \in [a, a+b]$, $b > 0$:

$$\hat{a} = \bar{\xi}_n - 3\sqrt{\hat{\sigma}_n^2}, \quad \hat{b} = 12\sqrt{\hat{\sigma}_n^2};$$

в) за гама-разпределението $f_\xi(x) = x^{\alpha-1}\lambda^{-\alpha}e^{-\lambda x}/\Gamma(\alpha)$, $x \geqq 0$:

$$\hat{\lambda} = \hat{\sigma}_n^2/\bar{\xi}_n; \quad \hat{\alpha} = \left(\bar{\xi}_n/\hat{\sigma}_n\right)^2;$$

г) за бета-разпределението $f_\xi(x) = x^{p-1}(1-x)^{q-1}\frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$, $x \in (0,1)$, $p, q > 0$:

$$\hat{p} = \bar{\xi}_n\left\{\left[\bar{\xi}_n(1-\bar{\xi}_n)/\hat{\sigma}_n^2\right] - 1\right\}, \quad \hat{q} = (1-\bar{\xi}_n)\left\{\left[\bar{\xi}_n(1-\bar{\xi}_n)/\hat{\sigma}_n^2\right] - 1\right\}.$$

12. Случайната величина ξ е нормално разпределена $\mathcal{N}(a, 3)$. По дадени 36 наблюдения над ξ покажете, че:

а) доверителният интервал за a с ниво на доверие $\gamma = 0,9$ е

$$\left[\bar{\xi}_{36} - 0,98; \bar{\xi}_{36} + 0,98\right];$$

б) минималният брой наблюдения за определяне на a чрез $\bar{\xi}_n$ с точност $\varepsilon = 0,9$ е 43.

13. Известно е, че времето на живот ξ на дадена генетична популация е нормално разпределено $\mathcal{N}(a, \sigma^2)$. При изследвания над 16 представители е установено, че средното им време на живот е $\bar{\xi}_{16} = 3000$, а средното квадратично отклонение е $\hat{\sigma}_{16}^2 = 400$. Покажете, че:

а) при доверителна вероятност $\gamma = 0,9$ доверителните интервали за a и σ съответно са $[2991,2; 3008,8]$ и [15,5;28,74];

б) доверителните вероятности, при които грешките са $\Delta_a = |\bar{\xi}_{16} - a| \leqq 10$ и $\Delta_\sigma = |\hat{\sigma}_{16} - \sigma| \leqq 2$, са съответно равни на 0,93 и 0,41.

14. Случайната величина ξ е експоненциално разпределена с плътност $f(x) = \theta^{-1}e^{-x/\theta}$. За оценка на θ е взето емпиричното средно $\bar{\xi}_n$. Покажете, че границите на доверителния интервал $[k_1\bar{\xi}_n, k_2\bar{\xi}_n]$ за θ, които удовлетворяват условията

$$\mathbf{P}\{\theta \leqq k_2\bar{\xi}_n\} = \mathbf{P}\{k_1\bar{\xi}_n \leqq \theta\} = \frac{1-\gamma}{2},$$

при $\gamma = 0,9$ са:

а) $k_1 = 0,64$, $k_2 = 1,83$ при $n = 10$;

б) $k_1 = 0,76$, $k_2 = 1,40$ при $n \geqq 30$.

Упътване. Приемете, че при $n \leqq 15$ случайната величина $\eta = 2n\bar{\xi}_n/\theta$ има χ^2-разпределение с $2n$ степени на свобода, а при $n \geqq 15$ случайната величина $\zeta = \sqrt{2\eta}$ е приблизително нормална $\mathcal{N}(\sqrt{2n-1}, 1)$.

15. Случайната величина ξ е разпределена по закона на Поасон с неизвестен параметър λ. За 50 единични интервала са наблюдавани 4 интервала с 0 реализации на ξ; 16 интервала с по 1 реализация; 20 интервала с по 2 реализации на ξ; 6 интервала с по 3 реализации и 4 интервала с по 4 реализации. Покажете, че доверителният интервал за λ при доверителна вероятност $\gamma = 0,9$ е $\hat{\lambda}_1 = 1,5$; $\hat{\lambda}_2 = 2,12$.

Упътване. Използвайте формулите за нормална апроксимация на χ^2-разпределението при степени на свобода над 30 и формулите за доверителни граници на параметъра на поасоновото разпределение (пример **2** от § 8).

16. При 30 опита, проведени по схема на Бернули, събитието A е настъпило 10 пъти. Покажете, че доверителният интервал за $p = \mathbf{P}(A)$ с ниво на доверие $\gamma = 0,95$ е $[\hat{p}_1, \hat{p}_2]$, където $\hat{p}_1 = 0,164$ и $\hat{p}_2 = 0,502$.

17. Нека $\xi_1, \ldots, \xi_n$ са независими наблюдения над случайна величина ξ с нормално разпределение. Покажете, че

$$\sigma^* = k\sqrt{\sum_{i=1}^{n} \frac{\left(\xi_i - \bar{\xi}_n\right)^2}{n}}$$

е неизместена оценка за $\sigma = \sqrt{\mathbf{D}\xi}$, когато $k = \sqrt{\frac{n}{2}}.\frac{\Gamma((n-1)/2)}{\Gamma(n/2)}$.

Упътване. Случайната величина $\sqrt{n}\sigma^*/(k\sigma)$ има χ^2-разпределение с $n-1$ степени на свобода.

18. „Двойно разпределение на Поасон“ с параметри λ_1 и λ_2 се дефинира с плътността

$$f(k, \lambda_1, \lambda_2) = \frac{\lambda_1^k e^{-\lambda_1} + \lambda_2^k e^{-\lambda_2}}{2k!}, \qquad k = 0, 1, 2, \ldots$$

Намерете оценки за λ_1 и λ_2 по метода на моментите.

19. Покажете, че ако случайната величина ξ има разпределение на Лаплас с плътност

$$f(x,\alpha,\mu) = \frac{1}{2\alpha}\exp\left[-\frac{|x-\mu|}{\alpha}\right],$$

то максимално правдоподобните оценки за μ и α съответно са $\hat{\mu} = \xi_{\left(\frac{n-1}{2}\right)}$, ако n е нечетно, и $\hat{\mu} \in [\xi_{\left(\frac{n}{2}\right)}, \xi_{\left(\frac{n}{2}\right)+1}]$, ако n е четно; $\hat{\alpha} = \frac{1}{n}\sum_{i=1}^{n}|\xi_i - \hat{\mu}|$. Тук $\xi_{(i)}$ са членове на вариационния ред на наблюденията, а n — броят на наблюденията.

20. Случайната величина ξ е нормално разпределена $\mathcal{N}(a,\sigma^2)$. По 25 независими наблюдения са получени оценките $\bar{\xi}_{25} = 100$, $\hat{\sigma}^2_{25} = 15$.

а) Покажете, че доверителният интервал с ниво на доверие $\gamma = 0{,}99$ за a е $[92{,}26; 107{,}74]$;

б) $\mathbf{P}\left\{|\bar{\xi}_{25} - a| \leqq 5\right\} = 0{,}95$.

21. Случайната величина ξ е нормално разпределена $\mathcal{N}(a, 2500)$. Покажете, че са необходими не по-малко от 11 наблюдения над ξ, за да бъде определено a с точност $\varepsilon = 25$ и надеждност $\alpha = 0,9$.

22. Над случайната величина ξ, разпределена по нормален закон $\mathcal{N}(a,\sigma^2)$ са проведени 100 наблюдения, при които е получено $\bar{\xi}_{100} = 5{,}5$; $\hat{\sigma}_{100} = 1{,}7$. Покажете, че доверителните граници за a и σ с ниво на доверие 0,85 са съответно $\hat{a}_1 = 5{,}25; \hat{a}_2 = 5{,}7; \hat{\sigma}_1 = 1{,}48; \hat{\sigma}_2 = 1{,}92$.

23. Случайната величина ξ има средно $\mathbf{E}\xi = a$. При 40 наблюдения над ξ са получени следните оценки за a и σ_ξ : $\hat{a} = 10400; \hat{\sigma}_\xi = 85$. Покажете, че доверителният интервал $[0{,}999.\hat{a}; 1,001.\hat{a}]$ покрива неизвестната стойност на a с вероятност $\gamma \approx 0{,}56$.

24. Грешките при измерванията на температурата на кипене на дадено вещество са нормално разпределени случайни величини $\mathcal{N}(0,\sigma^2)$. Покажете, че минималният брой измервания, при които оценката $\hat{\sigma}$ на средното квадратично отклонение σ не надвишава $\hat{\sigma}/5$ с вероятност $\gamma = 0{,}7$, е 15.

25. Случайната величина ξ е разпределена по експоненциален закон с параметър λ. Наблюдавани са 25 реализации на ξ с обща сума $\xi_1 + \cdots + \xi_{25} = 1600$. Докажете, че доверителният интервал за λ с ниво на доверие 0,8 при получените експериментални данни е $[50{,}75; 85{,}14]$.

Единадесета глава

Проверка на хипотези

§ 1. Статистически хипотези и критерии

Всяко решение, взето въз основа на обработени резултати от статистически наблюдения и експерименти, не е категорично и не може да бъде такова. То представлява една хипотеза, която ще наричаме статистическа хипотеза. Независимо от доводите и съображенията, които можем да приведем в подкрепа на истинността на нашето решение, остава известна доза съмнение, че освен него може да е вярно и нещо друго, различно от приетото. Как можем да дадем оценка на нашето твърдение? Въз основа на какво и с какви гаранции можем да решим да го приемем за вярно? Какви рискове допускаме, ако приемем (или не приемем) своето основно твърдение, а то се оказва невярно (или вярно, когато сме го отхвърлили)? Как да построим правилата за вземане на статистически решения, така че да минимизираме всички възможни рискове и последствия от неправилни решения? Всички тези въпроси и още редица други са свързани с теорията на статистическите критерии и правила за проверка на статистически хипотези. Проверката на хипотези е един от най-важните раздели на всички приложения. Ако задачата за оценяване на отделна вероятност, параметър или функция на разпределение е важна (особено в медицината, биологията, техниката), още по-важна е задачата за сравняване на две вероятности, на два параметъра или на две емпирични разпределения. Например, ако новото лекарство е изпитано върху редица пациенти и при това частта на излекуваните се е оказала по-голяма, отколкото при лекуваните с друго лекарство, свидетелства ли това за подобрено въздействие на новия препарат? Задачата не е тривиална, особено когато в едната експериментална група са лекувани n_1 пациенти и са излекувани k_1, а в другата от n_2 лекувани са излекувани k_2. Може ли различието в честотите k_1/n и k_2/n да се припише на различното действие на лекарствата, или то се дължи на обикновеното влияние на случайността при двата експеримента? А ако групите са повече от две и са изпробвани повече (например r) лекарствени препарата, кой от тях да предпочетем? Ако номерираме лекарствата с номера от 1 до r и започнем да твърдим, че i-тото лекарство е най-добро, ние ще сме изказали една хипотеза, която условно можем да означим с H_i. Така задачата се свежда до проверка, коя от хипотезите $H_1, \ldots, H_r$ е вярната. Към проблеми от проверка на хипотези водят често възникващи задачи от практиката като сравнителна оценка на различни технологични процеси по тяхната производителност, точност и икономичност, сравняване на конструктивни особености на машини и съоръжения, на организационни мероприятия, на методологии (на преподаване, на научно изследване). Такива проблеми са характерни за статистиката в агрономичните, медицинските, педагогическите, биологическите и хуманитарните науки. Въпросът за това, по какви признаци, по какви съотношения и показатели да се извършват сравняванията в постановките на задачите за проверка на хипотези, се отнася изцяло към онези специални области, в които са възникнали тези задачи. Обосновани изводи обаче могат да бъдат получени само по пътя на правилно конструиран научен анализ на данните от статистическите наблюдения.

Въпросът за проверка на статистически хипотези е изучен в основни линии през 30-те години на нашия век от Джерси Нейман и Егон Пирсън. В тази глава ще изложим основните идеи на построената от тях теория и ще покажем възможностите за нейното използване.

При проверка на хипотези се използва следният подход: Приема се, че изходен материал са наблюденията $\xi_1, \ldots, \xi_n$ над случайна величина ξ, които съдържат информация за истинността или неистинността на проверяваната хипотеза. На всяка група от наблюдавани стойности $\xi_1, \ldots, \xi_n$ се гледа като на единичен изход от един сложен опит, чието множество на възможни изходи е извадковото пространство $\mathfrak{X}_n$ — съвкупността от всички възможни стойности на вектора на наблюденията $\vec{\xi} = (\xi_1, \ldots, \xi_n)$. В извадковото пространство $\mathfrak{X}$ още преди извършване на наблюденията теоретично, вън и независимо от експеримента се конструира събитие E по подходящ начин. Когато проверяваната хипотеза е вярна, с достатъчно висока вероятност (пресметната по всички вероятностни закони, представящи хипотезата) настъпва E. Ако в резултат на опита събитието E не настъпи (това при нашата хипотеза е толкова малко вероятно, че може да се счита за практически невъзможно), той показва несъвместимост на предложената хипотеза с наблюденията. Обратно, ако събитието E настъпи, счита се, че резултатите от наблюденията не противоречат на нашата хипотеза. Обикновено не бива да се бърза с извода, че хипотезата се е потвърдила; може само да се признае нейната допустимост поне докато по-обстойни изследвания (например с помощта на повече материал или с по-силни критерии) не доведат до противоположни изводи.

Законността (разумността) на подхода се обяснява така: Данните от наблюденията съдържат информация за поведението на изследваните случайни величини, а то в условията на истинност на различните хипотези е различно и се отразява върху разпределението на извадката. По данните могат да се правят определени заключения за законите на разпределение на тези величини. В повечето случаи няма съмнение, че законът на разпределение на наблюдаваната величина при различните хипотези е от един и същ клас (например нормален, експоненциален, биномен и пр.). Обаче своеобразието на закона за всяка от хипотезите се изразява с различни стойности на параметрите му. Например при контрол на качеството хипотезата, че процентът на брак е p_0 (който се посочва в каталозите на предприятието производител), може да се проверява при контрахипотезата, че процентът на брак е $p_1 > p_0$. На двете твърдения съответстват биномни закони на разпределение на броя на дефектните изделия в извадка от n проверени изделия, но съответно с параметри p_0 и p_1.

По такъв начин проверката на хипотези се свежда към сравняване на статистическите характеристики със стойностите на параметри в законите на разпределение на наблюденията. Всяка статистическа хипотеза H е еквивалентна на някое предположение за закона на разпределение $F(x) = \mathbf{P}\{\xi < x\}$ на наблюдаваната случайна величина ξ, което е равносилно на предположение за вида на функцията на правдоподобие $L(x_1, \ldots, x_n)$. По тези предположения става реализуем посоченият подход в задачите за проверка на хипотези. По тях се правят вероятностни съждения (заключения) за поведението на някои статистически характеристики на извадката (критерии на проверката), които трябва да са верни, когато е вярна направената хипотеза H. Така на вече конструираното събитие E може да се съпостави една област $\overline{W} \subset \mathfrak{X}_n$ от стойности на извадката $\xi_1, \ldots, \xi_n$, в която тази извадка трябва да попадне след извършване на наблюденията при вярна хипотеза H. Областта $\overline{W}$ се нарича *област на приемане на хипотезата* H.

Допълнителната област $W = \mathfrak{X}_n \setminus \overline{W}$ се нарича *критична област* на критерия за проверка на хипотезата H. Когато наблюденията $\xi_1, \ldots, \xi_n$ в извадката се окажат в критичната област, хипотезата H се отхвърля като несъвместима с тези наблюдения.

Предложената за проверка хипотеза ще означаваме с H_0 и ще я наричаме *основна хипотеза*. Тя може да е вярна или не. Когато не е вярна, е вярна някоя друга хипотеза H_1, която наричаме *конкурираща* (*алтернативна*) *хипотеза* или просто *алтернатива*. Както на H_0, така и на H_1 съответстват някои предположения за функцията на правдоподобие (за закона на разпределение) на наблюдаваните величини. Ще ги означаваме съответно с $L_0(x)$ и $L_1(x)$ (или с $F_0(x)$ и $F_1(x)$).

Когато на H_0 съответства точно едно разпределение $L_0(x)$, се говори за *проста* основна *хипотеза*. Когато H_0 не се удовлетворява от едно единствено разпределение, говорим за *сложна* основна *хипотеза*. По аналогичен начин определяме понятията *проста* и *сложна алтернатива*.

Примери

а) Хипотезата H_0 е предположението, че процентът на негодните изделия при дадено производство е $p = p_0$, а алтернативата — че този процент е $p = p_1$, $p_1 \neq p_0$, където p_0 и p_1 са фиксирани числа между 0 и 1. Говори се за проверка на проста хипотеза срещу проста алтернатива.

б) Хипотезата H_0 е както в пример а), а хипотезата H_1 — че $p = p_1$ и $p_1 \neq p_0$, като p_0 е фиксирано число, а p_1 — кое да е друго число в (0,1). Говори се за проста хипотеза срещу сложна алтернатива.

в) Хипотезата H_0 е предположението, че $p_1 \leqq p_0$, докато в алтернативата H_1 се твърди, че $p > p_0$, където p_0 е дадено число в (0,1). Говори се за сложна хипотеза срещу сложна алтернатива.

Проверката на сложна хипотеза срещу проста алтернатива е безсмислена. В съответствие с класификацията от дадения пример се развиват и методите за проверка на статистически хипотези.

Трябва добре да се знае, че естественонаучните хипотези, възникващи в различни области на човешката дейност, нямат статистически характер. Изкуството на изследователя е да сътвори подходяща вероятностно-статистическа форма за тяхното изразяване. Това може да се прави по много начини. Необходимо е да се избира онзи начин, който позволява да се използва наличният фактически материал, или пък да се планира опит, предоставящ фактически материал.

След като задачата е статистически формализирана, проверката на хипотези може да се сведе към един от следните три вида проверки:

1. Проверка на хипотези относно параметри в разпределенията, принадлежащи на известни класове от закони на разпределение.

2. Проверка на хипотези относно класа на закона на разпределение на наблюдаваната случайна величина. Критериите за проверки от този тип се наричат критерии за съгласуваност и обикновено са т. нар. непараметрични критерии.

3. Проверка за отсъствие или наличие на връзка и взаимна зависимост между две и повече групи от случайни величини, за установяване вида на тази връзка и за изразяване количествената стойност на степента на зависимост. Този тип критерии се отнасят към съответните допълнителни раздели на статистическия анализ за търсене на причинно-следствени зависимости.

В тази глава ще разглеждаме предимно задачи от първия вид. Те най-добре илюстрират методологията за проверката на хипотези и са съществен дял от статистиката.

§ 2. Критични области
Грешки от първи и втори род

Видяхме, че за проверка на дадена хипотеза H са необходими наблюденията $\xi_1, \ldots, \xi_n$ над подходящо подбрана случайна величина ξ. Когато е вярна хипотезата H, на нея ще съответства разпределение $L(x \mid H)$ на извадката. Ако наблюденията $\xi_1, \ldots, \xi_n$ се окажат в построената област W (критерия) за проверката, хипотезата H се отхвърля; когато точката с координати $(\xi_1, \ldots, \xi_n)$ се окаже в областта $\mathfrak{X}_n \setminus W$, хипотезата H се приема като непротиворечаща на резултатите от наблюденията. Има вероятност хипотезата H да бъде отхвърлена дори когато е вярна. Тази вероятност е

$$(1) \qquad \alpha = \mathbf{P}\{\vec{\xi} \in W \mid H\} = \int_W L(x \mid H) dx$$

и винаги ще се допуска, щом $W \neq \varnothing$ (стига, разбира се, да имаме $\alpha > 0$). Желателно е подобна грешка да бъде много малко вероятна, т. е. α да е много малко число. Обикновено α се избира предварително — малко положително число, което се нарича ***ниво на съгласие.***

Вероятността α не дава еднозначна възможност за избор на критичната област W. По какви съображения да определим областта W? Ако има две подобласти (или просто две точки) Δ и Δ', такива че $\Delta \subset W$, а $\Delta' \subset \overline{W}$, и вероятността случайната точка $\vec{\xi} = (\xi_1, \ldots, \xi_n)$ да попадне в Δ е по-малка от α и равна на вероятността за попадане в Δ', то отнемайки от W областта Δ и прибавяйки към получената област областта Δ', ще получим нова критична област $W' = (W - \Delta) \cup \Delta'$. Вероятността да допуснем грешка при избрана критична област W' е същата както при W и двете не надминават α.

Отказът от проверяваната хипотеза H_0, когато тя е вярна, се нарича *грешка от първи род.*

Като вземем под внимание само грешката от първи род, критичните области не се определят еднозначно. Ако искаме да избегнем грешката от първи род въобще, би трябвало да избираме $W = \varnothing$. Но това означава винаги да приемаме безапелационно хипотезата H_0 — каквато и да е тя, и вероятността за грешка от първи род ще е нула. Обаче така не се постъпва никога, тъй като има опасност да се допусне и грешка от втори род.

Грешка от втори род се допуска, когато не се отхвърля проверяемата хипотеза, а в същото време тя не отговаря на истината.

Нека е избрана критична област W. Дори хипотезата H_0 да не е вярна, поради случайния характер на резултата $(\xi_1, \ldots, \xi_n)$ тази случайна точка може да не попадне в областта W. Тогава няма да бъде отхвърлена хипотезата H_0 и ще се допусне грешка от втори род. Затова е желателно критичната област W да бъде избрана така, че H_0 да не се отхвърля, когато е вярна, и да се отхвърля, ако не е вярна. Това ще бъде постигнато, ако е избрана такава критична област, че при нея да се минимизира вероятността β за грешка от втори род. Ясно е, че ако се стараем да направим по-малка вероятността за грешка от втори род (чрез разширяване на областта W и стесняване на $\overline{W}$), ще се увеличава вероятността за грешка от първи род и обратно.

Тук има обаче едно затруднение, тъй като вероятността за грешка от втори род не може да бъде определена предварително.

та ще зависи от това, коя алтернативна хипотеза H_1 ще се окаже правилна вместо невярната хипотеза H_0.

Нека предположим, че имаме само една алтернативна хипотеза H_1. *Мощност* π на критерия W относно хипотезата H_1 се нарича вероятността да се отхвърли хипотезата H_0, когато е вярна хипотезата H_1, т. е.

$$\pi = \mathbf{P}\{(\xi_1,\ldots,\xi_n) \in W \mid H_1\}.$$

При избрана критична област W вероятността да не бъде отхвърлена хипотезата H_0, когато е вярна H_1 (грешката от втори род), ще бъде $\beta = 1 - \pi$. Тъй като тази вероятност трябва да е колкото може по-малка, необходимо е да се направи критерият възможно по-мощен.

Пример 1. Ново лекарство трябва да бъде пуснато в производство. Възможни са две хипотези: H_0 — лекарството съдържа токсични съставки, т. е. съставки, вредни за човешкия организъм, и H_1 — лекарството не съдържа такива съставки.

Грешката от първи род е: решава се, че лекарството не съдържа вредни съставки, а всъщност то има и вредно въздействие.

Грешката от втори род е: решава се, че лекарството е вредно за човешкия организъм, а всъщност не е така.

Коя от двете грешки в случая е по-неприятна? Очевидно първата. Изборът на основната хипотеза се определя и от това, коя от двете грешки искаме да избегнем, или ако допускаме една от тях, това да бъде с възможно най-малка вероятност — вероятността, която можем да избираме предварително. Това обикновено е грешката от първи род и α се избира малко положително число — например $\alpha = 0,05$; $\alpha = 0,01$; $\alpha = 0,005$; $\alpha = 0,001$ и т. н. В зависимост от целите, които преследваме, и от последствията на взетите решения при проверката на хипотези се избира подходящата стойност α за вероятността на грешката от първи род, която определя нивото на съгласие.

Критерият W^*, имащ вероятност за грешка от първи род, равна на α, се нарича *най-мощен* относно алтернативната хипотеза H_1, ако сред всички критерии W с вероятност за грешка от първи род, ненадминаваща α, за W^* е изпълнено

$$\mathbf{P}\{\xi \in W^* \mid H_1\} \geqq \mathbf{P}\{\xi \in W \mid H_1\}. \tag{2}$$

Областта W^* се нарича *оптимална критична област.*

Задачата за определяне на най-мощния критерий за проверка на проста хипотеза H_0 срещу проста алтернатива H_1 се свежда до следното: Дадени са двете хипотези. На тях съответстват две възможности за разпределение на случайния вектор $(\xi_1,\ldots,\xi_n)$, определени с функции на правдоподобие $L_0(x_1,\ldots,x_n)$ и $L_1(x_1,\ldots,x_n)$. Търси се област W^*, за която вероятността

$$\mathbf{P}\{\vec{\xi}\in W^*\,|\,H_1\} = \int\cdots\int_{W^*\subset\mathfrak{X}_n} L_1(x_1,\ldots,x_n)dx_1\ldots dx_n$$

е максимална и е изпълнено условието

$$\mathbf{P}\{\vec{\xi}\in W^*\,|\,H_0\} = \int\cdots\int_{W^*\subset\mathfrak{X}_n} L_0(x_1,\ldots,x_n)dx_1\ldots dx_n \leqq \alpha.$$

Нивото на съгласие е най-голямата вероятност, с която сме съгласни да допуснем грешката от първи род — да отхвърлим основната хипотеза, въпреки че тя е вярна.

Подбирането на функциите на правдоподобие L_0 и L_1, съответстващи на простите хипотези H_0 и H_1, както и изборът на нивото на съгласие α не са математически задачи. Те са свързани с конкретните изисквания на решавания въпрос.

§ 3. Проста хипотеза срещу проста алтернатива. Лема на Нейман – Пирсън

По-нататък ще предполагаме, че на всяка проста хипотеза (ще я бележим с малка буква h вместо с голямата H) съответства едно единствено разпределение $L(\vec{x})$ на извадката $\xi_1,\ldots,\xi_n$. В случай на проста основна хипотеза h_0 срещу проста алтернатива h_1 ще считаме за известни съответните функции на правдоподобие $L = L_0(x)$ и $L = L_1(x)$, дефинирани в множеството $\mathfrak{X}_n$ на всички възможни стойности на извадката $\xi_1,\ldots,\xi_n$. При това едната от двете плътности L_0 и L_1 е истинската. За определеност ще предполагаме, че разпределението на извадката е непрекъснато.

Нека $\mathfrak{M}_\alpha$ е съвкупността на всички подмножества W на $\mathfrak{X}_n$, удовлетворяващи условието

$$\mathbf{P}\{(\xi_1,\ldots,\xi_n)\in W\,|\,h_0\} = \int_W L_0(x)dx = \alpha, \tag{1}$$

т. е. $\mathfrak{M}_\alpha$ е съвкупността от всички критични области, за които вероятността на грешката от първи род е равна на α — фиксирано число между 0 и 1. Съответната на α оптимална критична област W^* има свойството

$$\text{(2)} \qquad \mathbf{P}\{\vec{\xi} \in W^* \mid h_1\} = \int_{W^*} L_1(x)dx \geqq \int_W L_1(x)dx$$

за всяко $W \in \mathfrak{M}_\alpha$.

Вярно е следното твърдение:

Лема (Нейман – Пирсън). *При проверка на простата хипотеза*

$$h_0 : \ L(x) = L_0(x)$$

срещу простата алтернатива

$$h_1 : \ L(x) = L_1(x)$$

с ниво на съгласие α, ако критичната област $W^* \in \mathfrak{M}_\alpha$ е такава, че за някоя константа $K = K_\alpha > 0$ са изпълнени неравенствата

$$L_1(x) \geqq K L_0(x) \quad \text{за всяко } x = (x_1, \ldots, x_n) \in W^*,$$

$$L_1(x) \leqq K L_0(x) \quad \text{за всяко } x = (x_1, \ldots, x_n) \notin W^*,$$

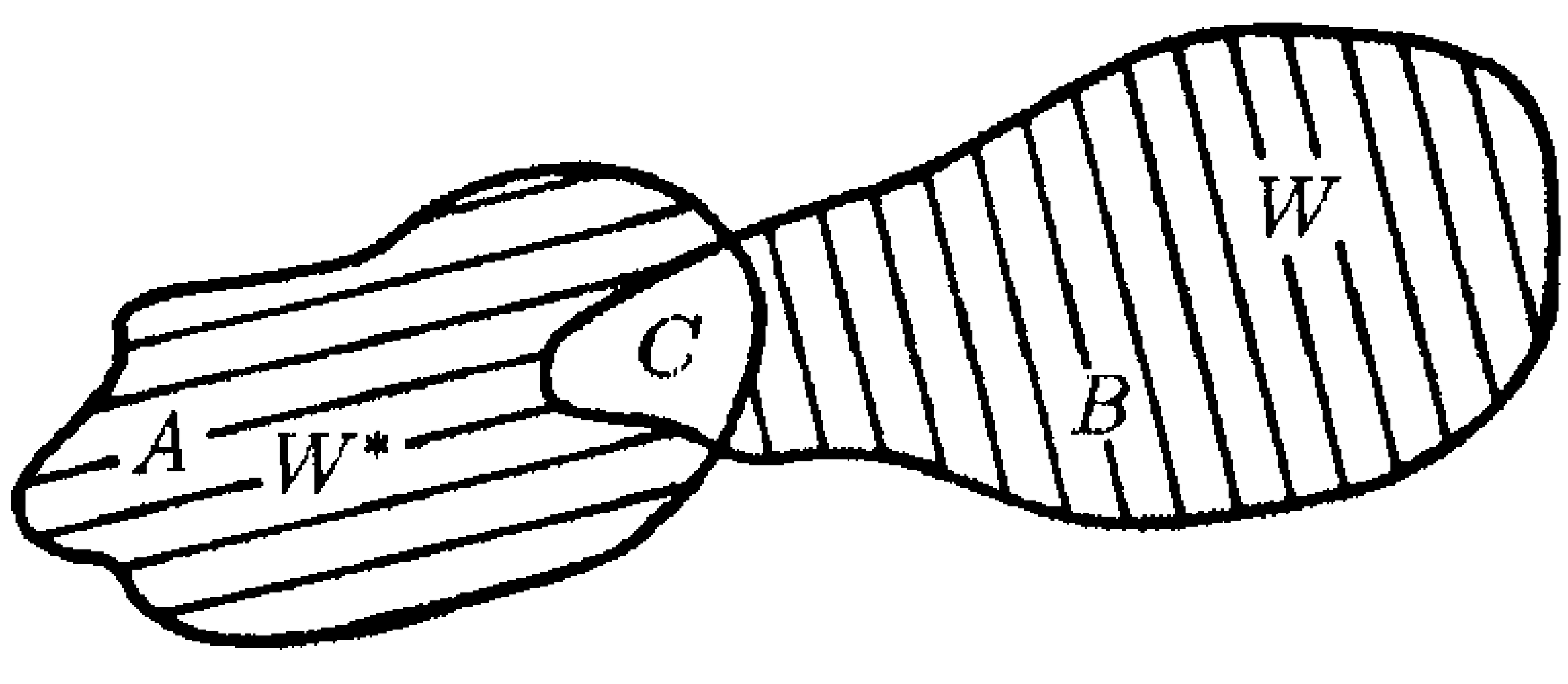
то W^* е оптимална критична област.

Константата K_α се избира по такъв начин, че да е изпълнено

$$\mathbf{P}\{\vec{\xi} \in W^* \mid h_0\} = \alpha.$$

Доказателство. Нека $W \in \mathfrak{W}_\alpha$ и $C = W \cap W^*$, $A = W^* \setminus C$, $B = W \setminus C$. Ще покажем, че за областта W^* е изпълнено (2).

Областта A е онази част от W^*, която няма общи точки с W (фиг. 19). Изпълнено е $W^* = A+C$, $W = B+C$, $W \cup W^* = A+B+C$, като A, B и C са непресичащи се подмножества на $\mathfrak{X}_n$.



Фиг. 19

Понеже W и W^* са от $\mathfrak{W}_\alpha$, имаме

$$\alpha = \int\limits_{W^*} L_0(x)dx = \int\limits_{C} L_0(x)dx + \int\limits_{A} L_0(x)dx;$$

$$\alpha = \int\limits_{W} L_0(x)dx = \int\limits_{C} L_0(x)dx + \int\limits_{B} L_0(x)dx,$$

откъдето намираме

$$\int\limits_{A} L_0(x)dx = \int\limits_{B} L_0(x)dx. \tag{3}$$

Да разгледаме сега разликата

$$\Delta = \int\limits_{W^*} L_1(x)dx - \int\limits_{W} L_1(x)dx = \left(\int\limits_{A} + \int\limits_{C}\right) L_1(x)dx - \left(\int\limits_{C} + \int\limits_{B}\right) L_1(x)dx$$

$$= \int\limits_{A} L_1(x)dx - \int\limits_{B} L_1(x)dx.$$

Обаче $L_1(x) \geqq KL_0(x)$ за всяко $\vec{x} \in A \subset W^*$ и $L_1(x) \leqq KL_0(x)$ за всяко $x \in B = W \setminus W^*$, откъдето

$$\Delta = \int\limits_{W^*} L_1(x)dx - \int\limits_{W} L_1(x)dx \geqq \int\limits_{A} KL_0(x)dx - \int\limits_{B} KL_0(x)dx$$

$$= K\left(\int\limits_{A} L_0(x)dx - \int\limits_{B} L_0(x)dx\right).$$

Поради (3) получаваме $\Delta \geqq 0$, т. е. W^* удовлетворява условието (2) за всяка друга критична област W с ниво на съгласие α. Следователно W^* е оптимална критична област.

Критична е областта W^* от всички точки $x = (x_1, \dots, x_n)$, за които $\dfrac{L_1(x)}{L_0(x)} \geqq K_\alpha$. Числото K_α е подбрано по такъв начин, че вероятността за грешка от първи род да е равна на α, т. е. ако вместо x положим случайната величина $\vec{\xi} = (\xi_1, \dots, \xi_n)$, при вярна хипотеза h_0 да е изпълнено

$$\mathbf{P}\left\{\frac{L_1(\xi_1, \dots, \xi_n)}{L_0(\xi_1, \dots, \xi_n)} \geqq K_\alpha \,|\, h_0\right\} = \alpha.$$

Намереният критерий W^* е най-чувствителен към алтернативата h_1 на хипотезата h_0. При друга основна хипотеза и същата алтернатива h_1 или при друга алтернатива и същата основна хипотеза h_0 критерият W^* може да се променя. Полученият критерий се нарича критерий, използващ отношението на функциите на правдоподобие. Той е най-мощен, когато алтернативната хипотеза е h_1. Целесъобразно е да бъде използван, когато се проверяват две хипотези h_0 и h_1 и с голяма сигурност може да се твърди, че точно една от тях е вярна.

След определяне на оптималната критична област W^* мощността на критерия е

$$\pi = \int_{W^*} L_1(x)dx.$$

Пример 1. Нека дадено случайно събитие съгласно хипотезата h_0 има вероятност p_0, а съгласно алтернативната хипотеза h_1 — вероятност $p_1 > p_0$. Нека при n независими опита това събитие се е осъществило X пъти. При кои X хипотезата h_0 трябва да се отхвърли при дадено ниво на съгласие α?

Според хипотезата h_0 вероятността за резултата X е биномна, т. е. можем да запишем

$$L_0(x) = \binom{n}{x} p_0^x (1-p_0)^{n-x}, \quad x = 0, 1, \ldots, n,$$

а според хипотезата h_1 тази вероятност е

$$L_1(x) = \binom{n}{x} p_1^x (1-p_1)^{n-x}, \quad x = 0, 1, \ldots, n.$$

Отношението на правдоподобие има вида

$$\frac{L_1(x)}{L_0(x)} = \left(\frac{p_1}{p_0}\right)^x \cdot \left(\frac{1-p_1}{1-p_0}\right)^{n-x},$$

откъдето критичната област W^* се дефинира с неравенството

$$\left(\frac{p_1}{p_0}\right)^x \cdot \left(\frac{1-p_1}{1-p_0}\right)^{n-x} \geqq K.$$

То е равносилно на неравенството

$$\left[\frac{p_0}{p_1} \cdot \frac{1-p_0}{1-p_1}\right]^x \geqq K_1,$$

където $K_1 = K\left(\frac{1-p_0}{1-p_1}\right)^n$ е константа. Оттук следва (след логаритмуването се взема предвид, че $p_1 > p_0$, $1-p_0 > 1-p_1$)

$$x \geqq \frac{\ln K_1}{\ln \frac{p_1(1-p_0)}{p_0(1-p_1)}} = C$$

и критичната област може да се определи с неравенството $x \geqq C$. Следователно трябва да отхвърляме хипотезата h_0, когато се окаже, че $X \geqq C$. При това C се определя по такъв начин, че вероятността на неравенството $X \geqq C$ при вярна хипотеза h_0 да не надмине α, т. е.

$$P\{X \geqq C \mid h_0\} = \sum_{k=C}^{n} \binom{n}{k} p_0^k (1-p_0)^{n-k} \leqq \alpha. \tag{4}$$

Така C_α е най-малкото цяло число, за което е изпълнено неравенството (4) и може да бъде определено след съответните изчисления. Ако след провеждането на n-те опита се окаже, че X е по-голямо или равно на така определеното число C_α, то хипотезата $h_0 : p = p_0$ трябва да се отхвърли и да се приеме алтернативата $h_1 : p = p_1$; ако се окаже, че $X < C_\alpha$, хипотезата $h_0 : p = p_0$ не се отхвърля.

Можем да определим асимптотично стойността на C_α, като използваме нормалното приближение за биномното разпределение. Неравенството $X \geqq C_\alpha$ е еквивалентно на $\frac{X - np_0}{\sqrt{p_0 q_0 n}} \geqq \frac{C_\alpha - np_0}{\sqrt{p_0 q_0 n}}$, в което лявата страна е приблизително $\mathcal{N}(0,1)$-разпределена. Следователно

$$\mathbf{P}\{X \geqq C_\alpha\} = \mathbf{P}\left\{\frac{X - np_0}{\sqrt{p_0 q_0 n}} \geqq \frac{C_\alpha - np_0}{\sqrt{p_0 q_0 n}}\right\} \approx 1 - \Phi\left(\frac{C_\alpha - np_0}{\sqrt{p_0 q_0 n}}\right).$$

Ако $t_{1-\alpha}$ е решение на уравнението $1 - \Phi(t) = \alpha$, то C_α е решение на $\frac{C_\alpha - np_0}{\sqrt{p_0 q_0 n}} = t_{1-\alpha}$, или $C_\alpha \approx np_0 + t_{1-\alpha}\sqrt{np_0 q_0}$.

Аналогично се пресмята, че асимптотичната мощност на този критерий е

$$\pi = \mathbf{P}\{X \geqq C_\alpha \mid h_1\} = 1 - \Phi\left(\sqrt{n}\frac{p_0 - p_1}{\sqrt{p_1 q_1}} + t_{1-\alpha}\sqrt{\frac{p_0 q_0}{p_1 q_1}}\right).$$

При $n \to \infty$ очевидно $\pi \to 1$ (поради $p_0 < p_1$).

Приведеният общ критерий често се среща в статистическия контрол на качеството — например в следния вид: Едно предприятие твърди, че произвежданата от него продукция не съдържа повече от 0,1% брак. Неговите потребители се съмняват в това и подлагат на щателна проверка 100 изделия, сред които забелязват X дефектни. Те не искат да допуснат грешка, като се откажат от предлаганите изделия, ако бракът наистина е не повече от 0,1%. Затова приемат, че вероятността за грешка от първи род α е равна на 0,05 и проверяват хипотезата $h_0 : p = 0,001$ срещу алтернативата $h_1 : p = 0,002$. Кога при описаните условия те трябва да приемат продукцията и кога не?

Според горните резултати, те трябва да определят такова минимално число C_α, че да е изпълнено неравенството

$$\sum_{k=C_\alpha}^{n} \binom{100}{k} (0,001)^k (0,999)^{100-k} \leqq 0,05.$$

Ако се окаже, че $X \leqq C_\alpha$, продукцията трябва да се приеме. Ако се окаже, че $X \geqq C_\alpha$, потребителите имат основание да се съмняват в уверенията на предприятието.

Пример 2. Нека $\mathfrak{X}_n$ е пространството на наблюденията $\xi_1, \ldots, \xi_n$, които съгласно хипотезата h_0 са независими, еднакво разпределени случайни величини със средно $a_0 = 0$ и дисперсия σ^2. Според алтернативната хипотеза h_1 същите наблюдения са независими и еднакво нормално разпределени, но със средно $a > 0$ и дисперсия σ^2. Съответните функции на правдоподобие са

$$L_0(x_1, \ldots, x_n) = (2\pi\sigma)^{-n/2} \exp\left\{-\frac{1}{2}\left(x_1^2 + \cdots + x_n^2\right)\right\},$$

$$L_1(x_1, \ldots, x_n) = (2\pi\sigma)^{-n/2} \exp\left\{-\frac{1}{2}\left[(x_1 - a)^2 + \cdots + (x_n - a)^2\right]\right\}.$$

Оптималната критична област W^* се определя с неравенството

$$W^* = \left\{\vec{x} : \frac{L_1(x)}{L_0(x)} = \exp\left\{a(x_1 + \cdots + x_n) - \frac{1}{2}na^2\right\} \geqq K\right\}.$$

Забелязваме, че отношението на правдоподобията е монотонно растяща функция от аргумента $\overline{x}_n = \dfrac{x_1 + \cdots + x_n}{n}$. Следователно неравенството

$$\exp\left\{na\overline{x}_n - \frac{1}{2}na^2\right\} \geqq K$$

е еквивалентно на неравенството

$$\overline{x}_n \geqq K_1,$$

където $K_1 = \dfrac{\ln K + na^2/2}{na}$. И така, критичната област W^* се определя с множеството на онези точки x с координати $x_1, \ldots, x_n$, за които $\dfrac{x_1 + \cdots + x_n}{n} \geqq K_1$.

Когато е вярна хипотезата h_0, случайната величина $\overline{\xi}_n$ е разпределена нормално със средно 0 и дисперсия σ^2/n. Така константата K_1 може да бъде определена при дадено ниво на съгласие α от условието

$$\mathbf{P}\left\{\frac{\xi_1 + \cdots + \xi_n}{n} \geqq K_1 \,|\, h_0\right\} = \alpha.$$

Според пример 10.4 обаче

$$\mathbf{P}\left\{\frac{\xi_1 + \cdots + \xi_n}{n} \geqq K_1 \,|\, h_0\right\} = 1 - \Phi\left(\frac{K_1\sqrt{n}}{\sigma}\right).$$

Ако $t_{1-\alpha}$ е решението на уравнението $\Phi(t) = 1 - \alpha$, константата K_1 се определя от равенството $K_1 = \dfrac{\sigma t_{1-\alpha}}{\sqrt{n}}$.

Ако резултатът от наблюденията покаже, че $\overline{\xi}_n \geqq K_1$, хипотезата $h_0 : \ a = 0$ се отхвърля и се приема хипотезата $h_1 : \ a = a_1 > 0$. При $\overline{\xi}_n < K_1$ хипотезата h_0 не се отхвърля.

Критериите, получени в тези примери, са забележителни с това, че не зависят от конкретната стойност на p_1 и a_1; достатъчно е само, че $p_1 > p_0$ и $a_1 > a_0$.

§ 4. Проста хипотеза срещу сложна алтернатива. Равномерно най-мощен критерий

А. Общи положения. В редица задачи за проверка на хипотези може да се срещне ситуация, в която на основната хипотеза H_0

съответства точно една плътност $L_0(x)$ на извадката, а на алтернативната хипотеза H_1 — цяла фамилия $L_1(x\,|\,\theta)$ от плътности, където $\theta \in \Theta$. Така се получава задача за проверка на проста хипотеза срещу сложна алтернатива.

За удобство се счита (и на практика това е обичайната ситуация), че $L_0(x)$ е от същото семейство $L_1(x\,|\,\theta)$, като $L_0(x)$ се получава за $\theta = \theta_0$. Задачата за проверка на хипотези се свежда към схемата:

Да се провери основната хипотеза

$$H_0 : \; L_0(x) = L(x\,|\,\theta_0)$$

срещу алтернативата

$$H_1 : \; L(x) = L(x\,|\,\theta), \quad \theta \in \Theta \setminus \theta_0,$$

с ниво на съгласие α $(0 < \alpha < 1)$.

Търсим критична област $W \subset \mathfrak{X}_n$, такава че: ако събитието $\left\{\vec{\xi} \in W\right\}$ настъпи, то хипотезата H_0 да се отхвърля с вероятност за грешка от първи род, равна на α, а ако настъпи $\left\{\vec{\xi} \notin W\right\}$, хипотезата H_0 да не се отхвърля.

Нека отново $\mathfrak{W}_\alpha$ е съвкупността на всички критични области, чиято вероятност за грешка от първи род не надминава α. С H_θ ще бележим онази проста хипотеза от групата на алтернативните хипотези, на която съответства точно плътността $L(x\,|\,\theta)$, $\theta \neq \theta_0$, $\theta \in \Theta$. Ясно е, че когато с основание се отхвърля проверяемата хипотеза H_0, ще е вярна някоя от простите хипотези H_θ.

Нека е избрана критичната област $W \in \mathfrak{W}_\alpha$. Да положим

$$\pi(W,\theta) = \mathbf{P}\left\{\vec{\xi} \in W \,|\, H_\theta\right\} = \int\limits_W L(x_1, \ldots, x_n \,|\, \theta)\, d\vec{x}.$$

За всяко $W \in \mathfrak{W}_\alpha$ е изпълнено

$$\pi(W,\theta_0) \leqq \alpha. \tag{1}$$

При условието (1) искаме да изберем такава област W^*, че да минимизираме вероятността за грешка от втори род, т. е. за всяко $\theta \in \Theta \setminus \theta_0$ да имаме

$$\mathbf{P}\left\{\vec{\xi} \notin \overline{W^*} \,|\, H_\theta\right\} = 1 - \mathbf{P}\left\{\vec{\xi} \in W^* \,|\, H_\theta\right\} = \inf_{W \in \mathfrak{W}_\alpha} \mathbf{P}\left\{\vec{\xi} \notin \overline{W} \,|\, H_\theta\right\}.$$

Това е равносилно на максимизиране на вероятността за отхвърляне на хипотезата H_0, когато е вярна алтернативата H_θ, т. е. за всяко $\theta \neq \theta_0$ да е изпълнено

$$(2) \qquad \mathbf{P}\left\{\vec{\xi} \in W^* \,|\, H_\theta\right\} = \pi(W^*, \theta) = \max_{W \in \mathfrak{W}_\alpha} \pi(W, \theta).$$

Вероятността $\pi(W, \theta)$ като функция на θ се нарича *функция на мощността на критерия* (на критичната област) $W \in \mathfrak{W}_\alpha$.

И така, търсим областта $W^* \in \mathfrak{W}_\alpha$ (евентуално зависеща от θ), такава че за всяко $W \in \mathfrak{W}_\alpha$ да е изпълнено (1). Може да се случи, че за всички допустими $\theta \neq \theta_0$ да получим една и съща област W^*. В такъв случай казваме, че измежду всички критерии с ниво на съгласие α критичната област W^* е *равномерно най-мощна критична област* (РНМКО) за проверка на хипотезата H_0 относно цялата съвкупност $H_1 = \{\theta \in \Theta \setminus \theta_0\}$ от допустими хипотези. Когато РНМКО съществува, тя може да се приеме като най-добра в сравнение с всяка друга критична област със същото ниво на съгласие α. Тогава $\pi(W^*, \theta) \geqq \pi(W, \theta)$ за всяко $\theta \neq \theta_0$ и всяко $W \in \mathfrak{W}_\alpha$.

Б. Метод за намиране на РНМКО. Въз основа на лемата на Нейман – Пирсън може да се предложи следната процедура за търсене на РНМКО:

Избира се някакво $\theta_1 \in \Theta$, $\theta_1 \neq \theta_0$. Разглежда се простата хипотеза $h_0 : \ \theta = \theta_0$ срещу проста алтернатива $h_1 : \ \theta = \theta_1$ с ниво на съгласие α и се търси оптимална критична област W^* за тази задача.

Ако W^* не зависи от θ_1, то W^* ще е РНМКО за проверка на простата хипотеза $h_0 : \ \theta = \theta_0$ срещу сложната алтернатива $H_1 : \ \theta \neq \theta_0$, $\theta \in \Theta$, с ниво на съгласие α, защото тогава (2) винаги е изпълнено.

Така задачата за намиране на РНМКО при проверка на проста хипотеза h_0 срещу сложна алтернатива H_1 се свежда до намиране на оптимална критична област W^* за проверка на простата хипотеза h_0 срещу коя да е от алтернативите $h_1 \in H_1$. Трябва да се покаже още, че W^* не зависи от алтернативата h_1.

Пример 1. В пример 1 от предишния параграф за проверка на простата хипотеза $h_0 : \ p = 0,001$ срещу простата алтернатива $h_1 : \ p = p_1 = 0,002$ получената критична област W^* с ниво на съгласие $\alpha = 0,05$ е РНМКО за проверка на простата хипотеза

$h_0: \; p = 0,001$ срещу сложната алтернатива $H_1: \; p > 0,001$. Критичната област W^* се определя от условието $X \geqq K_\alpha(n)$, където $K_\alpha(n)$ е най-малкото цяло число, решение на неравенството

$$\sum_{k=K_\alpha(n)}^{n} \binom{n}{k} (0,001)^k (0,999)^{n-k} \leqq 0,05.$$

Областта W^* не зависи от алтернативата h_1.

Пример 2. Наблюдава се случайна величина ξ с експоненциална плътност на разпределение

$$f(x \mid \theta) = \begin{cases} \dfrac{1}{\theta} e^{-x/\theta} & \text{за} \quad x \geqq 0; \\ 0 & \text{за} \quad x < 0. \end{cases}$$

Нека $\xi_1, \ldots, \xi_n$ са независими наблюдения над ξ. Да се построи РНМКО за проверка на простата хипотеза $H_0: \; \theta = 1$ срещу сложната алтернатива $H_1: \; \theta > 1$ с ниво на съгласие $\alpha = 0,001$.

Решение. Функцията на правдоподобие в случая е

$$L(x \mid \theta) = \theta^{-n} \exp\left[-\frac{1}{\theta} \sum_{i=1}^{n} x_i\right].$$

Нека $\theta > 1$. Проверяваме простата хипотеза $h_0: L(x \mid \theta) = L(x \mid 1)$ срещу простата алтернатива $h_1: \; L(x \mid \theta) = L(x \mid \theta_1), \; \theta_1 > 1$.

Търсената W^* е област в $\mathfrak{X}_n$, определена от условията на лемата на Нейман – Пирсън. За някое $K > 0$ имаме

$$\text{(3)} \qquad \left| \begin{array}{ll} L(x \mid \theta_1) \geqq K L(x \mid 1) & \text{за } x \in W^*, \\ L(x \mid \theta_1) \leqq K L(x \mid 1) & \text{за } x \notin W^*. \end{array} \right.$$

Условието (3) дава

$$\theta_1^{-n} \exp\left[-\frac{1}{\theta} \sum_{i=1}^{n} x_i\right] \geqq K \exp\left[-\sum_{i=1}^{n} x_i\right],$$

което е равносилно на

$$-n \ln \theta_1 - \frac{1}{\theta_1} \sum_{i=1}^{n} x_i \geqq \ln K - \sum_{i=1}^{n} x_i,$$

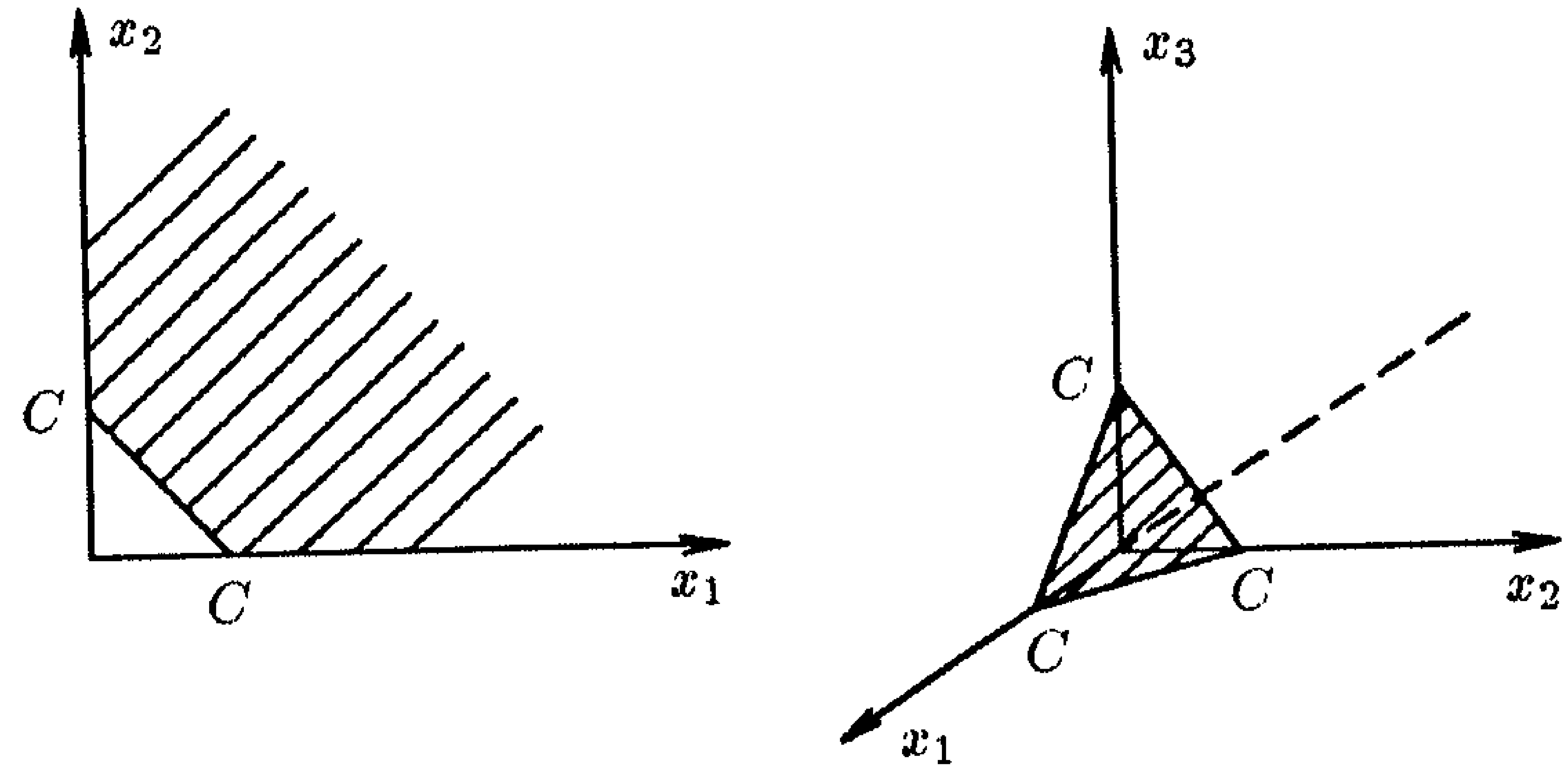
т. е. $(1-\theta^{-1})(x_1 + \cdots + x_n) \geqq \ln K + n \ln \theta_1 = K_1$. Окончателно имаме

$$\text{(4)} \qquad W^* = \left\{ \vec{x}: \; x_1 + \cdots + x_n \geqq \frac{K_1}{1-\theta^{-1}} = C_1, \quad x_i \geqq 0 \right\}.$$

Константата C се определя от условието

(5) $$\mathbf{P}\{\xi_1 + \dots + \xi_n \geqq C \mid h_0\} = 0{,}001.$$

Забележка 1. Към областта W^* принадлежат и всички точки $\vec{x} \in \mathfrak{X}_n$, за които $L(x \mid \theta_1) = L(x \mid 1) = 0$. Това е област A, определена с условието $\min\limits_{0 \leqq i \leqq n} x_i < 0$, тъй като за $\vec{x} \in A$ е изпълнено условието (3). Обаче $\mathbf{P}\left\{\vec{\xi} \in A \mid h_0\right\} = \mathbf{P}\left\{\vec{\xi} \in A \mid h_1\right\} = 0$, т. е. вероятността да получим експериментален резултат в областта A е 0, независимо дали е вярна хипотезата h_0 или не. Следователно критичната област W^* се състои само от точките $\vec{x}$, определени с (4). Това е област, лежаща в първи октант, опряна на някакъв симплекс, отделящ началото (фиг. 20).



Фиг. 20

Забележка 2. Очевидно е, че получената оптимална критична област W^* не зависи от $\theta_1 > 1$, защото

(6) $$W^* = \{\vec{x} : \ x_1 + \dots + x_n \geqq C > 0, \ x_i \geqq 0\},$$

а C зависи само от h_0. Следователно получената W^* ще е РНМКО за проверка на хипотезата $h_0 : \theta = 1$ срещу алтернативата $H_1 : \theta > 1$.

Тъй като ξ_i са независими, еднакво разпределени експоненциални случайни величини, то $\eta = \xi_1 + \dots + \xi_n$ е гама-разпределена случайна величина с плътност

$$f_\eta(x \mid \theta) = \frac{x^{n-1}\,\theta^{-n}\,e^{-x/\theta}}{\Gamma(n)}, \quad x \geqq 0.$$

Следователно C се определя от уравнението, получено от (5):

$$\mathbf{P}\{\eta_n \geqq C \mid h_0\} = \pi(1) = \frac{1}{(n-1)!} \int_C^\infty e^{-t} t^{n-1} dt = 0,001.$$

Стойността на C може да бъде получена от таблиците за непълната гама-функция и очевидно не зависи от θ_1.

За фиксирано n мощността на критерия като функция на θ е

$$\pi(\theta) = \mathbf{P}\{\eta_n \geqq C \mid h_0\} = \frac{\theta^{-n}}{(n-1)!} \int_C^\infty t^{n-1} e^{-t/\theta} dt = \frac{1}{(n-1)!} \int_{C/\theta}^\infty e^{-x} x^{n-1} dx.$$

B. Асимптотично решение. Както се вижда от § 3, в определянето на критичните области W^* играят роля разпределението на статистиката $t_n = \dfrac{L_1(\vec{\xi})}{L_0(\vec{\xi})}$ и константата K_α. За крайни n (брой на наблюденията в извадката) разпределението на статистиката t_n може да е трудно за определяне, както и не винаги неравенството $t_n \geqq K_\alpha$ може да се преобразува в еквивалентно неравенство, в което да участва друга, по-проста статистика t'_n с известно разпределение. В повечето такива случаи помагат знанията от теорията на вероятностите за граничното поведение на редиците от случайни величини $\{t_n\}$ или $\{t'_n\}$ при големи n. Точните разпределения се заменят с асимптотични, така че уравненията $\mathbf{P}\{t_n \geqq K_\alpha \mid h_0\} = \alpha$ да са достатъчни за определяне на необходимата стойност на константата K_α.

Пример 2. (продължение). При големи n стойностите на C и на $\pi(\theta)$ могат да бъдат определени чрез използване на централната гранична теорема. Тъй като ξ_i са независими и еднакво разпределени, $\mathbf{E}\xi_i = \theta$, $\mathbf{D}\xi_i = \theta^2$, то

$$\pi(\theta) \approx \lim_{n\to\infty} \mathbf{P}\left\{\frac{\xi_1 + \cdots + \xi_n - n\theta}{\theta\sqrt{n}} \geqq \frac{C - n\theta}{\theta\sqrt{n}} \mid H_\theta\right\} = 1 - \Phi\left(\frac{C - n\theta}{\theta\sqrt{n}}\right).$$

При $\theta = 1$ трябва да е изпълнено

$$\pi(1) \approx 0,02 = 1 - \Phi\left(\frac{C-n}{\sqrt{n}}\right), \quad \text{или} \quad \Phi\left(\frac{C-n}{\sqrt{n}}\right) = 0,98.$$

Уравнението $\Phi(t) = 0,98$ има корен $t_{1-\alpha} = 2,055$. Следователно $\dfrac{C-n}{\sqrt{n}} = 2,055$, откъдето $C = n + 2,055\sqrt{n}$. Мощността на критерия се изразява с приблизителното равенство

$$\pi(\theta) \approx 1 - \Phi\left(\frac{(1-\theta)n + 2,055\sqrt{n}}{\theta\sqrt{n}}\right) = 1 - \Phi\left(\frac{2,055}{\theta} - \left(1 - \frac{1}{\theta}\right)\sqrt{n}\right).$$

При дадена стойност на θ лесно може да се определят достатъчен брой точки от графиката на функцията $\pi(\theta)$. Например при $n = 25$, $\theta = 2$ имаме

$$\frac{2,055}{\theta} - \left(1 - \frac{1}{\theta}\right)\sqrt{n} = -1,4725,$$

$$1 - \Phi(-1,4725) = \Phi(+1,4725) = 0,9292.$$

Окончателно отхвърляме $h_0 : \theta = 1$, когато

$$\xi_1 + \cdots + \xi_n \geqq 2,055\sqrt{n} + n.$$

Ако $\xi_1 + \cdots + \xi_n < 2,055\sqrt{n} + n$, нямаме основание да отхвърляме хипотезата h_0. Ръководейки се от тази процедура, ще допускаме грешка от първи род в 2% от случаите.

Забележка 3. Ако областта Θ не е интервалът $[1, \infty)$, а $(0, \infty)$, получената оптимална критична област W^* зависи от θ_1. Наистина при $\theta_1 < 1$ тя ще се определя от условията

$$W_1^* = \{\vec{x} : \ x_1 + \cdots + x_n \leqq C_1'\};$$

$$\mathbf{P}\{\xi_1 + \cdots + \xi_n \leqq C_1 \,|\, h_1\} \leqq 0,02.$$

Следователно не можем да получим равномерно най-мощен критерий за $H_0 : \theta = 1$ срещу алтернативата $H_1 : \theta \neq 1$. Тогава задачата се решава с методите за проверка на сложни хипотези.

Търсенето на РНМКО често изглежда като въпрос на налучкване, но всъщност е въпрос на опит и умение да се работи с разпределенията, изучени в теорията. Особено важно е да се преодолее т. нар. разпределителен проблем, т. е. да се определи стойността на константата K_α, участваща при определянето на критичната област в отношението на двете функции на правдоподобие. Разпределителният проблем става още по-труден в задачите за проверка на сложни хипотези, тъй като там лемата на Нейман – Пирсън не е приложима и се прибягва към по-сложни начини за определяне на критичните области.

§ 5. Сложна хипотеза срещу сложна алтернатива. Критерий с отношенията на правдоподобията

При проверка на сложни хипотези е удобно да си представяме разпределенията (функциите на правдоподобие), съответни на основната хипотеза H_0 и на алтернативната H_1, като принадлежащи на едно и също семейство $L(x\,|\,\theta)$, $\theta \in \Theta$. При това на основната сложна хипотеза H_0 съответстват разпределенията при $\theta \in \Theta_0 \subset \Theta$, а на сложната хипотеза H_1 съответни са разпределенията, за които $\theta \in \Theta \setminus \Theta_0$. Множествата Θ_0 и $\Theta \setminus \Theta_0$ могат да се отъждествяват съответно с хипотезите H_0 и H_1, като множеството Θ се нарича ***множество на допустимите хипотези.*** За краткост проверяваната хипотеза бележим $H_0 : \ \theta \in \Theta_0$, а алтернативата — $H_1 : \ \theta \in \Theta \setminus \Theta_0$. Нивото на съгласие α сега има смисъл, че която и проста хипотеза h_0 от H_0 да е истинската, вероятността за отхвърляне на h_0 да не е по-голяма от α.

Пример 1. Наблюдавана е случайна величина $\xi \in \mathcal{N}(a, \sigma^2)$. Множеството на допустимите хипотези може да се зададе с областта $\Theta = \{\theta = (a, \sigma^2), a \in \mathbf{R}^1, \sigma > 0\}$ в двумерното евклидово пространство $\mathbf{R}^2$. Основната хипотеза би могла да бъде $H_0 : \ a = \sigma^2$ срещу алтернативите $a \neq \sigma^2$. В случая и хипотезата, и алтернативите са сложни.

При проверката на сложни хипотези се използват различни интуитивни съображения, които водят към решаване на поставената задача — построяване на критична област за проверка на основната хипотеза H_0 срещу алтернативата H_1 с ниво на съгласие α. Най-често се използва т. нар. *критерий с отношението на правдоподобията.*

За целта се предполага, че навсякъде в извадковото пространство $\mathfrak{X}_n$ (множеството от възможните стойности на вектора $\vec{\xi} = (\xi_1, \ldots, \xi_n)$ от наблюденията) поне една от предложените хипотези е възможна, т. е. че

$$\sup_{\theta \in \Theta} L(x_1, \ldots, x_n\,|\,\theta) \neq 0.$$

Дефинираме следната функция на n аргумента:

$$\lambda_n(\vec{x}) = \lambda_n(x_1, \ldots, x_n) = \frac{\sup\limits_{\theta \in \Theta_0} L(x_1, \ldots, x_n\,|\,\theta)}{\sup\limits_{\theta \in \Theta} L(x_1, \ldots, x_n\,|\,\theta)}.$$

Критерий с отношението на правдоподобията. Критичната област $W \subset \mathfrak{X}_n$ е такава, че за някоя константа $K \in (0,1)$ е изпълнено: $\lambda_n(\vec{x}) \leqq K$ за всяко $\vec{x} \in W$ и $\lambda_n(\vec{x}) > K$ за всички $\vec{x} \notin W$. Константата K е най-голямото число, за което $\mathbf{P}\left\{\vec{\xi} \in W \mid \theta\right\} \leqq \alpha$ за всички $\theta \in \Theta_0$. Когато W е определена и в резултат на експеримента се окаже, че $\vec{\xi} \in W$, хипотезата $H_0: \ \theta \in \Theta_0$ се отхвърля като несъгласуваща се с данните; ако $\vec{\xi} \notin W$, няма основания при тези данни да се отхвърли H_0.

Следните евристични разсъждения доказват логическата и интуитивната естественост на критерия с отношението на правдоподобията. Стойността θ_0, която е неизвестна и точно съответства на разпределението $L(x \mid \theta_0)$ на извадката $\vec{\xi}$, е близка до онази стойност $\widehat{\theta} = \widehat{\theta}(\xi_1, \ldots, \xi_n)$ на параметъра θ, която максимизира функцията на правдоподобие, т. е. максимално правдоподобната оценка $\widehat{\theta}$ за θ_0 ще се намира някъде близо около θ_0. Това е едно от свойствата на максимално правдоподобните оценки. Следователно, ако $\theta_0 \in \Theta_0$, много е вероятно и $\widehat{\theta} \in \Theta_0$. Тогава в отношението $\lambda_n(\vec{\xi})$ супремумите в числителя и знаменателя ще се достигат в близки точки и ще са близки помежду си числа, т. е. $\lambda_n(\vec{\xi})$ ще е близко до 1. Ако обаче H_0 не е вярна, т. е. $\theta_0 \notin \Theta_0$, най-вероятно е и $\widehat{\theta} \notin \Theta_0$. Тогава $L(\vec{\xi} \mid \widehat{\theta})$ ще приема големи стойности при $\widehat{\theta} \in \Theta \setminus \Theta_0$, значително по-големи от случаите, когато $\widehat{\theta} \in \Theta_0$. Следователно много е вероятно частното $\lambda_n(\vec{\xi})$ да е далеч от 1, т. е. да е близко до 0. Но в тези случаи за малки стойности на $\lambda_n(\vec{\xi})$ бихме искали да отхвърляме хипотезата H_0, т. е. да причислим такива аргументи на $\lambda(\vec{x})$ към критичната област W.

З а б е л е ж к а 1. Критерият с отношение на правдоподобията се практикува и в случаите, когато се проверява проста хипотеза срещу сложна алтернатива и не съществува РНМКО.

Пример 2. Нека $\xi \in \mathcal{N}(a, 1)$, където a е неизвестен параметър. Да се построи критерий за проверка на простата хипотеза $h_0: \ a = a_0$ срещу сложната алтернатива $H_1: \ a \neq a_0$ с ниво на съгласие $\alpha = 0,02$.

Ако разглеждаме простата хипотеза $h_0: \ a = a_0$ срещу простата алтернатива $h_1: \ a = a_1 \neq a_0$ с ниво на съгласие α, ще видим, че при $a_1 < a_0$ критичната област W^* има вида

$$W^* = \{\vec{x}: \ x_1 + \cdots + x_n \leqq K\},$$

а при $a_1 > a_0$ критичната област е от вида

$$W_1^* = \{\vec{x}:\ x_1 + \cdots + x_n \geqq K\}.$$

Двете критични области са от различен тип и явно зависят от избора на конкуриращата хипотеза. Следователно не съществува РНМКО.

Ще използваме критерия с отношението на правдоподобията. При независими наблюдения имаме

$$\lambda_n(\vec{x}) = \frac{(2\pi)^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i - a_0)^2\right\}}{(2\pi)^{-n/2} \sup_a \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i - a)^2\right\}}.$$

Знаменателят достига своя максимум в точката $\widehat{a}$, в която функцията $h(a) = \sum_{i=1}^{n}(x_i - a)^2$ има минимум. Тя се определя от условието $h'(a) = 0$, т. е. $h'(a) = -2\sum_{i=1}^{n}(x_i - a) = 0$, или $\widehat{a} = \dfrac{(x_1 + \cdots + x_n)}{n}$.

Понеже $h''(\widehat{a}) = 2n > 0$, то $h(a)$ има абсолютен минимум за $a = \widehat{a}$. Следователно

$$\lambda_n(\vec{x}) = \frac{L(x \mid a_0)}{L(x \mid \widehat{a})} = \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i - a_0)^2 + \frac{1}{2}\sum_{i=1}^{n}(x_i - \widehat{a})^2\right].$$

Като преобразуваме

$$\begin{aligned}\sum_{i=1}^{n}(x_i - \widehat{a})^2 &= \sum_{i=1}^{n}(x_i - a_0)^2 - 2\sum_{i=1}^{n}(x_i - a_0)(\widehat{a} - a_0) + \sum_{i=1}^{n}(\widehat{a} - a_0)^2 \\ &= \sum_{i=1}^{n}(x_i - a_0)^2 - n(\widehat{a} - a_0)^2,\end{aligned}$$

ще получим

$$\lambda_n(\vec{x}) = \exp\left[-\frac{n}{2}(\widehat{a} - a_0)^2\right].$$

Условието за W^*, определено с отношението на правдоподобията, е $W^* = \{\vec{x}:\ \lambda_n(\vec{x}) \leqq K\}$ и е равносилно на $(\widehat{a} - a_0)^2 \geqq K_1$, т. е.

$$W^* = \{\vec{x}:\ |\widehat{a} - a_0| \geqq C > 0\} = \left\{\vec{x}:\ \left|a_0 - \frac{x_1 + \cdots + x_n}{n}\right| \geqq C\right\}.$$

След като определихме вида на критичната област, започваме с разпределителния проблем. Той се състои в определяне на константата C така, че

$$\mathbf{P}\left\{\vec{\xi} \in W^* \,|\, a_0\right\} = \mathbf{P}\left\{|\bar{\xi}_n - a_0| \geqq C \,|\, a_0\right\} = \alpha = 0,02.$$

Ако е вярна h_0, то $\xi_i \in \mathcal{N}(a_0, 1)$, следователно

$$\bar{a}(\vec{\xi}) = \frac{\xi_1 + \cdots + \xi_n}{n} - a_0 \in \mathcal{N}\left(0, \frac{1}{n}\right).$$

Тогава $\sqrt{n}(\bar{\xi}_n - a_0) \in \mathcal{N}(0,1)$. Следователно константата C може да се определи от условието

$$\begin{aligned}\alpha &= x\{|\bar{\xi}_n - a_0| \geqq C \,|\, a = a_0\} = \mathbf{P}\{\sqrt{n}\,|\bar{\xi}_n - a_0| \geqq C\sqrt{n} \,|\, a = a_0\} \\ &= 1 - \mathbf{P}\{|\mathcal{N}(0,1)| < C\sqrt{n}\} = 1 - [\Phi(C\sqrt{n}) - \Phi(-C\sqrt{n})],\end{aligned}$$

където $\Phi(x)$ е стандартната нормална функция на разпределение. Като използваме, че $\Phi(-a) = 1 - \Phi(a)$, за C ще получим уравнението

$$\Phi(C\sqrt{n}) = 1 - \frac{\alpha}{2}.$$

При $\alpha = 0,02$ корен на уравнението $\Phi(t) = 0,99$ е $t_{1-\frac{\alpha}{2}} = 2,33$, откъдето $C = 2,33/\sqrt{n}$.

Следователно ще отхвърлим хипотезата $h_0 : \; a = a_0$, ако настъпи събитието

$$\left\{|\bar{\xi}_n - a_0| \geqq \frac{2,33}{\sqrt{n}}\right\} = \left\{\left|\frac{\xi_1 + \cdots + \xi_n}{n} - a_0\right| > \frac{2,33}{\sqrt{n}}\right\}.$$

Ако сумата $\xi_1 + \cdots + \xi_n$ се окаже в границите

$$na_0 - 2,33\sqrt{n} \leqq \xi_1 + \cdots + \xi_n \leqq na_0 + 2,33\sqrt{n},$$

нямаме основание да се отказваме от предположението, че $a = a_0$.

Още веднъж в резюме да повторим основните резултати и етапи от проверката на хипотези:

1. Избираме критерий, по който бихме искали да отхвърлим или не основната хипотеза H_0, като при това:

а) при прости хипотези срещу прости алтернативи основната лема на Нейман – Пирсън дава метод за определяне оптималната критична област;

б) при сложни хипотези има две възможности:

б$_1$) да се търси РНМКО с помощта на лемата на Нейман – Пирсън;

б$_2$) да се търси оптимален критерий с отношението на правдоподобията;

2. Избраният критерий W дава възможност да отхвърлим H_0, когато за функцията $\lambda_n(\vec{\xi})$ на наблюденията $(\xi_1+\cdots+\xi_n)$ е изпълнено $\lambda_n(\vec{\xi}) \leqq C$ за някоя константа C, избрана така, че

$$\mathbf{P}\left\{\lambda_n(\vec{\xi}) \leqq C \,|\, H_0\right\} \leqq \alpha.$$

3. Основна трудност е разпределителният проблем — да се намери разпределението на $\lambda_n(\vec{\xi})$ при условие, че е вярна H_0, което дава възможност да се определи константата C. Не винаги при крайни n може да се определи точното разпределение на $\lambda_n(\vec{\xi})$. В повечето случаи се доказва някакво гранично приближение за разпределението на $\lambda_n(\vec{\xi})$ при $n \to \infty$. Тази апроксимация, за която обикновено разпределителният проблем е по-лек, се ползва вместо точното разпределение на $\lambda_n(\vec{\xi})$ при определяне на константата C_α от условието $\mathbf{P}\left\{\lambda_n(\vec{\xi}) \leqq C_\alpha\right\} = \alpha$.

§ 6. Проверка на хипотези и доверителни множества

А. Развитата в § 1.9 теория на доверителните интервали за неизвестния едномерен параметър θ се обобщава лесно за многомерния случай. Нека да наблюдаваме случайна величина ξ, чието разпределение $F(x,\theta)$ зависи от многомерния параметър $\theta \in \Theta$, $\theta = (\theta_1, \dots, \theta_k)$. Тогава на всяка стойност на θ в извадковото пространство $\mathfrak{X}_n$ от стойностите на наблюденията $(\xi_1, \dots, \xi_n)$ ще съответства едно разпределение $L(x_1, \dots, x_n \,|\, \theta)$ на извадката. Нека на всяко $\vec{x} = (x_1, \dots, x_n)$ от $\mathfrak{X}_n$ сме съпоставили по едно подмножество $S(\vec{x}) \subset \Theta$. Когато вместо $\vec{x}$ вземем вектора на наблюденията $\vec{\xi}$, множеството $S(\vec{\xi})$ ще зависи от резултатите $\xi_1, \dots, \xi_n$ и следователно ще е случайно подмножество на Θ. Ако $\theta_0 = (\theta_1^0, \dots, \theta_k^0)$ е истинската стойност на параметъра θ, която съответства на наблюдаваната случайна величина ξ, то съществува определена вероятност $\mathbf{P}_\theta\left\{\theta_0 \in S(\vec{\xi})\right\}$ случайното множество $S(\vec{\xi})$ да съдържа и точката θ_0. Когато $\mathbf{P}_{\theta_0}\left\{\theta_0 \in S(\vec{\xi})\right\}=\gamma$, казваме, че $S(\vec{x})$ е система от γ-доверителни множества и означаваме с долен индекс γ, т. е. $S(\vec{x}) = S_\gamma(\vec{x})$. Обикновено за всяко $\gamma \in (0,1)$ съществуват много системи доверителни множества с доверителна вероятност γ.

Представлява интерес задачата по дадена доверителна вероятност γ да се построи система от γ-доверителни множества $S_\gamma(\vec{x})$. Пътят на тяхното построяване следва идеите на построяване на доверителните интервали за едномерен параметър θ и може да се използва същевременно методът с отношенията на правдоподобията от проверката на сложни хипотези.

Б. За всяко $\theta \in \Theta$ може да се определи такова множество $t_\gamma(\theta)$ от точки $\vec{x}$ в $\mathfrak{X}_n$, че

(1)
$$\mathbf{P}_\theta\left\{\vec{\xi} \in t_\gamma(\theta)\right\} = \gamma,$$

където γ е дадената доверителна вероятност. Множеството $t_\gamma(\theta)$ съответства на интервала $\left[\widehat{t}_2(\gamma), \widehat{t}_1(\gamma)\right]$, който имахме в случая на едномерен параметър θ в § 1.9. Съотношението (1) съответства на изискването $\mathbf{P}\left\{t_n(\vec{\xi}) \in \left[\widehat{t}_2(\gamma), \widehat{t}_1(\gamma)\right]\right\} = \gamma$ от § 10.9. На областта $\mathfrak{D}_{(0,t)}(\gamma)$ съпоставяме в случая подмножеството $\mathfrak{D}(\gamma)$ от декартовото произведение $\Theta \times \mathfrak{X}_n$, за което е изпълнено съотношението $\vec{x} \in t(\theta)$, т. е.

$$\mathfrak{D}(\gamma) = \{(\theta, \vec{x}),\ \theta \in \Theta,\ \vec{x} \in t_\gamma(\theta)\}.$$

Сега на всяко $\vec{x} \in \mathfrak{X}_n$ съпоставяме съвкупността $S_\gamma(\vec{x})$ от всички точки θ на Θ, за които $(\theta, \vec{x}) \in \mathfrak{D}(\gamma)$. Така $S_\gamma(\vec{x})$ съответства на интервала $\left[\widehat{\theta}_1(\vec{x}), \widehat{\theta}_2(\vec{x})\right]$ от § 1.9. От тази конструкция е ясно, че трите събития

$$A = \left\{(\theta_0, \vec{\xi}) \in \mathfrak{D}(\gamma)\right\}; \quad B = \left\{\vec{\xi} \in t_\gamma(\theta)\right\}; \quad C = \left\{\theta_0 \in S_\gamma(\vec{\xi})\right\}$$

са еквивалентни, защото отразяват едно и също нещо — принадлежност на точката $(\theta_0, \vec{\xi})$ към подмножеството $\mathfrak{D}(\gamma)$ на $\Theta \times \mathfrak{X}_n$. Така получаваме, че за аналога на доверителния интервал — доверителното множество $S_\gamma(\vec{\xi})$, е изпълнено

(2)
$$\mathbf{P}_{\theta_0}\{C\} = \mathbf{P}_{\theta_0}\left\{\theta_0 \in S_\gamma(\vec{\xi})\right\} = \gamma.$$

В частност, ако множеството $S_\gamma(\vec{\xi})$ е интервалът $\left[\widehat{\theta}_1^i, \widehat{\theta}_2^i\right]$, равенството (2) е еквивалентно на

(3)
$$\mathbf{P}\left\{\widehat{\theta}_1^i \leqq \theta_i^0 \leqq \widehat{\theta}_2^i\right\} = \gamma.$$

Тук i е някой от индексите 1, 2, ..., k, а $\widehat{\theta}_1^i$ и $\widehat{\theta}_2^i$ не зависят от $\theta_1^0, \ldots, \theta_k^0$. Тогава $S_\gamma(\vec{\xi})$ е γ-доверителен интервал за θ_i^0.

Множествата $S_\gamma(\vec{\xi})$ не са определени еднозначно, но и при тяхното определяне може да се използва методът на максималното правдоподобие.

В. Системата от γ-доверителни множества $S_\gamma(\vec{x}) \subset \Theta$ се нарича *система от множества на максималното правдоподобие*, когато е изпълнено (2) и от $\theta' \in S_\gamma(\vec{x})$, $\theta'' \in \Theta$ и $L(x \mid \theta'') > L(x \mid \theta')$ следва, че $\theta'' \in S_\gamma(\vec{x})$. Така за всяка система от множества на максималното правдоподобие и за някоя функция $C(\vec{x})$ на $\vec{x}$ е изпълнено

$$S_\gamma(\vec{x}) = \{\theta : \ L(x \mid \theta) \geqq C(\vec{x})\}, \qquad \vec{x} \in \mathfrak{X}_n.$$

Ясно е оттук, че максимално правдоподобната оценка $\widehat{\theta}$ за θ_0 за всяко $x \in \mathfrak{X}_n$ ще принадлежи на $S_\gamma(\vec{x})$.

Методът, основан на принципа на отношението на правдоподобие за проверка на сложни хипотези, е полезен и се използва за получаване на доверителни множества на максималното правдоподобие.

Нека да конструираме критерий за проверка на хипотезата $h_0 : \ \theta = \theta_0$ срещу алтернативата $H_1 : \ \theta \neq \theta_0$ (т. е. $\theta \in \Theta \setminus \theta_0$). Тогава критичната област W^* с ниво на съгласие α се определя от изискването: за някоя константа $K \in (0,1)$ да е изпълнено

$$\lambda(\vec{x} \mid \theta_0) = \frac{L(\vec{x} \mid \theta_0)}{\sup\limits_{\theta \in \Theta} L(\vec{x} \mid \theta)} \geqq K \quad \text{за} \quad \vec{x} \notin W^* \tag{4}$$

и $\lambda(\vec{x} \mid \theta_0) \leqq K$ за всички $\vec{x} \in W^*$. Тук константата K се определя така, че $\mathbf{P}_{\theta_0}\left\{\lambda(\vec{\xi}) \leqq K\right\} = 1 - \alpha$.

Ако положим $\gamma = 1 - \alpha$ и дефинираме множествата

$$S_\gamma(\vec{x}) = \{\theta : \ \lambda(\vec{x} \mid \theta) \geqq K_\alpha, \quad \theta \in \Theta\}, \tag{5}$$

където K_α се определя от изискването $\mathbf{P}_{\theta_0}\left\{\theta_0 \in S_\gamma(\vec{\xi})\right\} \geqq \gamma = 1 - \alpha$, еквивалентно на условието $\mathbf{P}_{\theta_0}\left\{\lambda(\vec{\xi}) \geqq K_\alpha\right\} = \gamma$, ще сме определили γ-доверителното множество $S_\gamma(\vec{x})$ за неизвестния параметър θ_0. При това $S_{\theta_0}(\vec{\xi})$ е множество на максималното правдоподобие.

В този метод един от трудните моменти е пресмятането на константата K_α за всяко ниво на съгласие α. Обаче при достатъчно общи предположения може да се използва асимптотичното разпределение на величината $-2\ln\lambda(\vec{\xi}\,|\,\theta)$.

Теорема. *Когато са налице условията на теорема 7.2 за асимптотична нормалност на максимално правдоподобната оценка $\widehat{\theta}_n$, е изпълнено*

$$\lim_{n\to\infty} \mathbf{P}_{\theta_0}\left\{-2\ln\lambda(\vec{\xi}\,|\,\theta_0) < t\right\} = \mathbf{P}\left\{\chi^2(k) < t\right\},$$

където $k = \dim(\Theta)$ е размерността на неизвестния параметър θ_0.

Доказателство. Като използваме (4) и забележим, че супремумът в знаменателя се достига за $\widehat{\theta}_n$, ще получим

$$-2\ln\lambda(\vec{\xi}\,|\,\theta_0) = 2\left[\ln L(\vec{\xi}|\widehat{\theta}_n) - \ln L(\vec{\xi}\,|\,\theta_0)\right]. \tag{6}$$

Тейлоровото развитие на разликата в дясната страна може да се запише чрез скаларно произведение на вектори $(\vec{x},\vec{y}) = \sum_i x_i y_i$ по следния начин:

$$\begin{aligned}\ln L(\vec{\xi})\,|\,\widehat{\theta}_n) - \ln L(\vec{\xi}\,|\,\theta_0) &= \left(\frac{\partial}{\partial\theta}\ln L(\vec{\xi}\,|\,\widehat{\theta}_n), \theta_0 - \widehat{\theta}_n\right)\\ &+ \frac{1}{2!}\left(n\mathbf{B}_0^{-1}(\theta_0 - \widehat{\theta}_n), \theta_0 - \widehat{\theta}_n\right) + R.\end{aligned} \tag{7}$$

Тук $\mathbf{B}_0^{-1}$ е матрицата, обратна на $\mathbf{B}_0 = (b_{ij}(\theta))$, с елементи

$$b_{ij}(\theta) = \mathbf{E}_{\theta_0}\left\{\frac{\partial\ln f(\xi\,|\,\theta)}{\partial\theta_i}\cdot\frac{\partial\ln f(\xi\,|\,\theta)}{\partial\theta_j}\right], \quad i,j = 1,\,\ldots,\,k$$

(вж. § 10.7). Отново използваме, че $\widehat{\theta}_n$ е максимално правдоподобна оценка, така че първото събираемо в дясната страна на (7) е 0. Освен това:

— остатъчният член $R \to 0$ по вероятност при $n \to \infty$;

— случайната величина $\sqrt{n}(\widehat{\theta}_n - \theta_0)$ клони по разпределение към случайна величина η_∞ с нормално разпределение $\mathcal{N}(0, \mathbf{B}_0^{-1})$. Затова по (6) и (7) случайната величина $-2\ln\lambda(\vec{\xi}\,|\,\theta_0)$ клони по разпределение към случайна величина $(\mathbf{B}_0^{-1}\eta, \eta)$.

Остава да се покаже, че величината $(\mathbf{B}_0^{-1}\eta, \eta)$ има $\chi^2(k)$–разпределение. Наистина от положителната дефинитност на матрицата $\mathbf{B}_0^{-1}$ следва, че тя се представя във вида $\mathbf{B}_0^{-1} = \mathbf{D}^2$, където транспонираната на $\mathbf{D}$ съвпада с $\mathbf{D}$, т. е. $\mathbf{D}' = \mathbf{D}$. Ако $\zeta = \mathbf{D}\eta$, ще имаме $\zeta = \mathbf{D}'\eta$ и $(\mathbf{B}_0^{-1}\eta, \eta) = (\zeta, \zeta)$.

Случайният вектор ζ има разпределение $\mathcal{N}(0, I_k)$, където I_k е единична матрица с размерност $k \times k$. Тук се използва фактът, че ако някакъв случаен вектор η има многомерно (k-мерно) нормално разпределение $\mathcal{N}(\mu, \Sigma)$ и $\mathbf{B}$ е неизродена реална матрица с размерност $k \times k$, случайният вектор $\zeta = \mathbf{B}\eta$ има k-мерно нормално разпределение $\mathcal{N}(\mathbf{B}\mu, \mathbf{B}\Sigma\mathbf{B}')$. Следователно (ζ, ζ) има $\chi^2(k)$-разпределение.

Граничното разпределение не зависи даже от $L(\vec{x}, \theta_0)$, а не само от θ_0. Затова при достатъчно голямо n константата K_α в (5) може да бъде определена просто поради верността на следните равенства:

$$\begin{aligned}\gamma = \mathbf{P}_{\theta_0}\left\{\theta_0 \in S_\gamma(\vec{\xi})\right\} &= \mathbf{P}_{\theta_0}\left\{\lambda(\vec{\xi} \mid \theta_0) \geqq K_\alpha\right\}\\ &= \mathbf{P}_{\theta_0}\left\{-2\ln\lambda(\vec{\xi} \mid \theta_0) \leqq C_\gamma\right\} = \mathbf{P}\{\chi^2(k) \leqq C_\gamma\}.\end{aligned}$$

Ако определим C_γ от условието $\mathbf{P}\{\chi^2(k) \leqq C_\gamma\} = \gamma$ и положим $K_\alpha = \exp(-2C_\gamma)$, то областта $S_\gamma(\vec{\xi})$ ще е множеството от всички $\theta \in \Theta$, за които $\lambda(\vec{\xi} \mid \theta) \geqq K_\alpha$.

Г. Пример. Нека $\xi \in \mathcal{N}(a, \sigma^2)$ с неизвестни a_0 и σ^2. Да се построи γ-доверително множество за неизвестния параметър a_0 по n независими наблюдения $\xi_1, \dots, \xi_n$ над ξ.

Решение. Съвместната плътност на $\xi_1, \dots, \xi_n$ е

$$L(x \mid a, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - a)^2\right].$$

Да построим критерия с отношение на правдоподобията за проверка на хипотезата (тя е сложна) $H_0: \; a = a_0$, $\sigma^2 > 0$ срещу алтернативата $H_1: \; a \neq a_0$, $\sigma^2 > 0$. Отношението на правдоподобията има вида

$$\lambda(\vec{x} \mid a_0) = \frac{\sup\limits_{\sigma} L(x \mid a_0, \sigma^2)}{\sup\limits_{a,\sigma} L(x \mid a, \sigma^2)}.$$

За пресмятане на числителя приравняваме на нула частната производна $\dfrac{\partial}{\partial\sigma^2} L(x \mid a_0, \sigma^2)$ и решаваме полученото уравнение относно σ^2. Решението полагаме в $L(x \mid a_0, \sigma^2)$ и откриваме, че числителят на $\lambda(x \mid a_0)$ е максимален, когато

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - a_0)^2.$$

За изчисляване на знаменателя решаваме съвместно уравненията $\frac{\partial}{\partial a} L(x \mid a, \sigma^2) = 0$ и $\frac{\partial}{\partial \sigma^2} L(x \mid a, \sigma^2) = 0$. Решенията им относно a и σ^2 са

$$a = \overline{x}_n = \frac{1}{n}(x_1 + \cdots + x_n); \qquad \sigma^2 = \widehat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x}_n)^2.$$

Следователно

$$\lambda(\vec{x} \mid a_0) = \left[\frac{\sum\limits_{i=1}^{n} (x_i - \overline{x}_n)^2}{\sum\limits_{i=1}^{n} (x_i - a_0)^2} \right]^{n/2}.$$

Критичната област се състои от всички $\vec{x}$, удовлетворяващи неравенството $\lambda(\vec{x} \mid a_0) \leqq K$, което е еквивалентно на

$$\frac{\sum\limits_{i=1}^{n} (x_i - a_0)^2}{\sum\limits_{i=1}^{n} (x_i - \overline{x}_n)^2} \geqq K_1. \tag{8}$$

След преобразуването

$$\sum_{i=1}^{n} (x_i - a_0)^2 = \sum_{i=1}^{n} (x_i - \overline{x}_n + \overline{x}_n - a_0)^2 = \sum_{i=1}^{n} (x_i - \overline{x}_n)^2 + n(\overline{x}_n - a)^2$$

получаваме, че (8) е еквивалентно на

$$\frac{(x_n - a_0)^2}{\sum\limits_{i=1}^{n} (x_i - \overline{x}_n)^2} \geqq K_2$$

или на

$$\left| \frac{\sqrt{n}(\overline{x}_n - a_0)^2}{\widehat{s}_n} \right| \geqq C; \quad \widehat{s}_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x}_n)^2.$$

Константата C се избира така, че да е изпълнено

$$\mathbf{P}_{a_0}\left\{\left|\frac{\sqrt{n}(\overline{\xi}_n - a_0)^2}{\widehat{s}_n}\right| \geqq C\right\} = 1-\gamma.$$

Обаче, когато хипотезата H_0 е вярна, случайната величина под знака за абсолютна стойност има t-разпределение на Стюдънт с $n-1$ степени на свобода (§ 10.5). Ако изберем C от таблиците за t-разпределението с $n-1$ степени на свобода при ниво на съгласие $\alpha = 1-\gamma$, т. е. от условието $1-\gamma = \mathbf{P}\{|\,t(n-1)| \geqq C_\gamma\}$, тогава C_γ ще е решение на уравнението

$$\mathbf{P}\{\,t(n-1) \leqq C_\gamma\} = \frac{1+\gamma}{2}.$$

В съответствие с (5) определяме доверителната област

$$S_\gamma(\vec{x}\,) = \{a_0 :\ \lambda(\vec{x}\,|\,a_0) \geqq C_\gamma\} = \left\{a_0 :\ \left|\frac{\sqrt{n}(\overline{x}_n - a_0)}{\widehat{s}_n}\right| \leqq C_\gamma\right\}$$

$$= \left\{\overline{x}_n - \frac{C_\gamma \widehat{s}_n}{\sqrt{n}} \leqq a_0 \leqq \overline{x}_n + \frac{C_\gamma \widehat{s}_n}{\sqrt{n}}\right\}.$$

Така получихме, че γ-доверителният интервал $S_\gamma(\vec{x}\,)$ е доверително множество на максималното правдоподобие за a_0 с ниво на съгласие $\alpha = 1-\gamma$.

Други примери от проверка на сложни хипотези се разглеждат в дванадесета глава.

Задачи

1. Нека $f_0(x)$ и $f_1(x)$ са две различни вероятностни плътности на разпределение и случайната величина ξ да има плътност на разпределение $f(x) = (1-\theta)\,f_0(x) + \theta f_1(x)$. Покажете, че оптималната критична област W_α за проверка на хипотезата $h_0 :\ \theta = 0$ срещу алтернативата $h_1 :\ \theta = 1$ с ниво на съгласие α се определя като множеството на всички точки $\vec{x}$, за които $\prod_{i=1}^{n} f_1(x_i) \geqq K_\alpha \prod_{i=1}^{n} f_0(x_i)$. Константата K_α се определя така, че

$$\int_W \prod_{i=1}^{n} f_0(x_i)d\vec{x} = 1-\alpha.$$

2. Нека π^* е мощността на най-мощния критерий W^* с ниво на съгласие α, получен по лемата на Нейман – Пирсън за проверка на простата хипотеза

h_0 срещу простата алтернатива h_1. Покажете, че $\alpha < \pi$, с изключение на случая, когато $L_0(x) \equiv L_1(x)$.

3. Наблюдават се 20 стойности на случайна величина ξ. Трябва да се провери хипотезата h_0 : $f_\xi(x) = \frac{1}{20}$ за всички $x = 1, 2, \ldots, 20$ (т. е. че ξ е равномерно разпределена върху числата от 1 до 20) срещу алтернативата

$$h_1 : f_\xi(x) = \begin{cases} 0,60 & \text{за } x = 1, \\ 0,15 & \text{за } x = 2, 3, \\ 0,1 & \text{за } x = 4, 5, \ldots, 20 \end{cases}$$

с ниво на съгласие $\alpha = 0,10$. Покажете, че $W_1 = \{1, 2\}$, $W_2 = \{1, 3\}$ и $W_3 = \{1, 4\}$ са критични области за W_α. Докажете, че W_1 и W_2 са оптимални критични области.

4. Нека е направено едно наблюдение над случайната величина ξ. Възможни са две хипотези:

h_0 : ξ е равномерно разпределена в интервала $[0, 10]$,

h_1 : разпределението на ξ в интервала $[0, 10]$ е неравномерно и има плътност

$$f_1(x) = \begin{cases} 1,2 & \text{за } x \in [0; 0,5], \\ 0,3 & \text{за } x \in [0,5; 1,5], \\ \frac{1}{85} & \text{за } x \in [1,5; 10], \\ 0 & \text{за } x \notin [0, 10]. \end{cases}$$

Покажете, че $W_1 = [0; 1)$, $W_2 = [0; 0,5) + [1; 1,5)$, $W_3 = [0; 0,5) + [0,75; 1,25)$, $W_4 = [0; 0,5) + [1,5; 2)$, $W_5 = [5,2; 6,2)$ са 0,10-критични области, сред които W_1, W_2 и W_3 са оптимални критични области.

5. Извършва се статистически контрол на качеството на голяма партида от N изделия ($N \gg 1$). Проверяват се n ($n \ll N$) изделия, като ν_n от тях се оказват дефектни. Покажете, че оптималната критична област за проверка на хипотезата „Процентът на дефектните изделия в партидата е $p = p_0 = 0,04$", срещу алтернативата „Процентът на дефектните изделия е $p = p_1 = 0,10$" с ниво на съгласие $\alpha = 0,02$ има вида $W^* = \{\nu_n \geqq K_\alpha\}$, където K_α се определя като минималното цяло число, удовлетворяващо условието

$$\sum_{k=K_\alpha}^{n} \binom{n}{k} (0,04)^k (0,96)^{n-k} \leqq 0,02.$$

Асимптотично при $n \to \infty$ критичната област W^* може да се представи с неравенството

$$W^* : \ \nu_n \geqq 0,04n + 2,05\sqrt{0,04.0,96n}.$$

Асимптотичната мощност на построения критерий е

$$\pi \approx \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x_0} \exp\left\{-\frac{u^2}{2}\right\} du, \quad x_0 = \frac{0,06 + 2,05\sqrt{0,04.0,96n}}{\sqrt{0,10.0,90n}}.$$

6. В условията на зад. 5 определете приблизително какъв трябва да бъде минималният брой наблюдения в извадката, при който мощността на оптималния критерий е не по-малка от 0,95. Изследвайте мощността на критерия в зависимост от n (пресметнете мощността при $n = 10$; 20; 30). Покажете, че при $n \to \infty$ мощността клони към 1.

7. В условията на зад. 5 покажете, че в случай на основна хипотеза $h_0 : p = p_0 = 0,10$ срещу алтернативата $h_1 : p = p_1 = 0,04$, критичната област W^* има вида $W^* = [0, K_\alpha]$, където K_α се определя като максималното цяло число, за което е изпълнено условието

$$\sum_{k=0}^{K_\alpha - 1} \binom{n}{k} (0,10)^k (0,90)^{n-k} \leqq \alpha.$$

8. Партида съдържа N изделия. По случаен начин се избират n изделия за проверка. Нека X е броят на дефектните изделия сред проверените. Докажете, че равномерно най-мощната критична област за проверка на хипотезата „Броят на дефектните изделия в партидата е $D = D_0$“ срещу алтернативата „Броят на дефектните изделия в партидата е $D > D_0$“ с ниво на съгласие α се определя с условието

$$\frac{D_0 + 1}{N - D_0} \cdot \frac{N - D_0 - n + X}{D_0 + 1 - X} \geqq K_\alpha.$$

Упътване. Изберете простата алтернатива $h_1 : D = D_0 + 1$.

9. Партида изделия се приема, ако вероятността произволно избрано изделие да е дефектно не превишава 0,02. Проверени са 480 изделия, сред които се оказали 12 дефектни. Може ли да се приеме партидата при ниво на съгласие $\alpha = 0,05$?

10. Наблюдавани са $n = 100$ независими реализации на случайна величина $\xi \in \mathcal{N}(\theta, \sigma^2)$, за която е известно, че $\sigma = 5,2$. Получено е $\overline{\xi}_n = 27,56$. Докажете, че при ниво на съгласие $\alpha = 0,05$ основната хипотеза $h_0 : \mathbf{E}\xi = \theta_0 = 26$ трябва да бъде отхвърлена при конкуриращи хипотези $H_1 : \mathbf{E}\xi = \theta_1 \neq 26$.

11. Покажете, че в условията на зад. 10 при емпирични данни $n = 16$; $\overline{\xi}_n = 118,2$; $\sigma^2 = 3,6$; $\alpha = 0,05$; $h_0 : \theta = \theta_0 = 120$; $H_1 : \theta = \theta_1 \neq 120$ няма основание за отхвърляне на основната хипотеза h_0.

12. Нека ξ е случайна величина, равномерно разпределена върху числата $1, 2, \ldots, N$, където N е неизвестно. Извършени са n независими наблюдения $\xi_1, \ldots, \xi_n$ над ξ. Покажете, че равномерно най-мощната критична област

W^* за проверка на хипотезата $h_0 : \ N = N_0$ срещу сложната алтернатива $H_1 : \ N > 1, \ N \neq N_0$, с ниво на съгласие α, има вида

$$W^* = \left\{ \vec{x} = (x_1, \ldots, x_n) : \max_{1 \leqq i \leqq n} x_i > N_0 \cup \max_{1 \leqq i \leqq n} x_i \leqq N_0 \alpha^{1/n} \right\}.$$

Мощността на критерия е $\pi(N) = \mathbf{P}\{\vec{\xi} \in W^* \mid N\} = 1 - (1-\alpha)\left(\frac{N_0}{N}\right)^n$.

Начертайте графиката на $\pi(N)$ като функция на N.

В следващите задачи ξ е случайна величина, а $\xi_1, \ldots, \xi_n$ — независими наблюдения над ξ.

13. Разпределението на ξ е $\mathcal{N}(\theta, 1)$.

а) Покажете, че равномерно най-мощната критична област за проверка на хипотезата $h_0 : \ \theta = 0$ срещу алтернативата $H_1 : \ \theta < 0$ с ниво на съгласие $\alpha = 0,01$ има вида $W^* = \{\vec{x} : \ x_1 + \cdots + x_n < -2,33\sqrt{n}\}$.

б) При $n = 9$ определете най-малката стойност θ_1 на θ, за която вероятността да се отхвърли хипотезата h_0 при вярна хипотеза $h_1 : \ \theta = \theta_1$ не е по-малка от $0,96$.

14. Нека ξ е експоненциално разпределена с плътност $f(x) = \theta^{-1} e^{-x\theta^{-1}}$, $x \geqq 0$, $\theta > 0$. Покажете, че равномерно най-мощната критична област за проверка на $h_0 : \ \theta = 1$ срещу алтернативата $H_1 : \ \theta > 1$ с ниво на съгласие $\alpha = 0,03$ има вида $W^* = \{\vec{x} : \ x_1 + \cdots + x_n \geqq K\}$, където K е решение на уравнението

$$\frac{1}{(n-1)!} \int_K^\infty t^{n-1} e^{-t} dt = 0,03.$$

При големи n е изпълнено приблизително $W^* = \{\vec{x} : \ x_1 + \cdots + x_n \leqq n + 1,881\sqrt{n}\}$ и мощността на критерия е

$$\pi(\theta) = 1 - \Phi\left(\frac{1,881}{\theta} - \frac{(\theta - 1)\sqrt{n}}{\theta}\right) \longrightarrow 1 \quad \text{при} \quad n \to \infty.$$

15. Нека $f_\xi(x) = \dfrac{1}{\theta}$ при $x \in (0, \theta)$ и $f_\xi(x) = 0$ при $x \notin (0, \theta)$. Покажете, че равномерно най-мощната критична област за проверка на $h_0 : \ \theta = 1$ срещу алтернативата $H_1 : \ \theta > 1$ с ниво на съгласие α има вида

$$W^* = \left\{ \vec{x} : \left[\max_{1 \leqq i \leqq n} x_i \leqq \sqrt[n]{\alpha} \right] \cup \left[\max_{1 \leqq i \leqq n} x_i > 1 \right] \right\}.$$

Мощността на критерия е $\pi(\theta) = 1 - (1-\alpha)\theta^{-n}$.

16. Нека $f(x) = \dfrac{\theta^x e^{-\theta}}{x!}$, $x = 0, 1, 2, \ldots$, $\theta > 0$ (разпределение на Поасон). Покажете, че равномерно най-мощната критична област за проверка на хипотезата $h_0 : \ \theta = \theta_0$ срещу алтернативата $H_1 : \ \theta > \theta_0$ с ниво на съгласие а има вида $\{\vec{x}: x_1 + \cdots + x_n > K_\alpha\}$, където K_α се определя от условието $\sum_{x=1}^{K_\alpha} \frac{[n\theta_0]^x e^{-n\theta_0}}{x!} = 1 - \alpha$. При големи n е изпълнено $K_\alpha \approx n\theta_0 + 1,75\sqrt{n\theta_0}$. Определете мощността на критерия при $\theta_0 = 1$ и $n = 25$.

17. Нека $f(x) = e^{-(x-\theta)}$ за $x \in [\theta, \infty)$ и $f(x) = 0$ при $x < \theta$. Покажете, че равномерно най-мощната критична област за проверка на хипотезата $h_0: \theta = \theta_0$ срещу алтернативата $H_1: \theta < \theta_0$ с ниво на съгласие α има вида $W^* = \{\vec{x}: (\inf x_i < \theta_0) \cup (\sup x_i \leqq K_\alpha)\}$, където $K_\alpha = \theta_0 - \ln(1 - \sqrt{\alpha})$.

Ако $h_1: \theta > \theta_0$, то $W^* = \{\inf x_i \geqq K_\alpha\}$, където $K_\alpha = \theta - \frac{1}{n} \ln \alpha$.

18. Нека $f(x) = C \exp[a(\theta) b(x) + C(\theta) + d(x)]$ (**експоненциална фамилия плътности**). Покажете, че оптималната критична област W^* за проверка на хипотезите $h_0: \theta = \theta_0$ и $h_1: \theta = \theta_1$ с ниво на съгласие α при $a(\theta_1) - a(\theta_0) > 0$ има вида $\sum_{i=1}^{n} b(x_i) \geqq K_\alpha$, където K_α се определя с $\mathbf{P}\{\vec{\xi} \in W^* \mid \theta = \theta_0\} = \alpha$.

19. Извършено е едно наблюдение над случайна величина $\xi \in \mathcal{N}(\theta, 1)$, при което е регистрирана стойността $\xi_1 = \zeta$. Хипотезите са $h_0: \theta_0 = 0$ и $h_1: \theta = 10$. При ниво на съгласие $\alpha = 0,02$ следва ли да се отхвърли h_0? Въз основа на данните решете коя от двете хипотези е за предпочитане.

20. Нека случайната величина ξ има еднопараметрична плътност на разпределение $f(x, \theta) = c(\theta) h(x) \exp[-Q(\theta) T(x)]$, където θ е реален параметър, а функцията $Q(\theta)$ е строго монотонна. Покажете, че при проверка на хипотезата $h_0: \theta = \theta_0$ срещу алтернативата $H_1: \theta > \theta_0$ съществува равномерно най-мощна критична област W^*, която при проста извадка $\xi_1, \ldots, \xi_n$ изглежда така:

а) ако $Q(\theta)$ е растяща, то

$$W^* = \left\{ \vec{x}: \sum_{i=1}^{n} T(x_i) < K \right\};$$

б) ако $Q(\theta)$ е намаляваща, то

$$W^* = \left\{ \vec{x}: \sum_{i=1}^{n} T(x_i) > K \right\}.$$

Тук $K = K_\alpha$ се определя от θ_0 и от нивото на съгласие α от уравнението $\int_K^\infty g(x, \theta) dx = \alpha$, където $g(x, \theta)$ е вероятностната плътност на разпределение на статистиката $Z = \sum_{i=1}^{n} T(\xi_i)$.

21. Нека простата хипотеза $h_0 : \vec{p} = \vec{p}_0 = (p_1, \dots, p_N)$ изразява, че полиномното разпределение на пълна система от N случайни събития $\{A_i\}_{i=1}^N$ се задава с фиксирания вектор $\vec{p}_0$. Нека сложната алтернатива $H_1 : \vec{p} \neq \vec{p}_0$ изразява, че $\vec{p} \in \mathfrak{B}^+$, където $\mathfrak{B}^+$ е симплексът на всички вектори с неотрицателни координати $p_i \geqq 0$, чиято сума е $p_1 + \dots + p_N = 1$. Нека са проведени n независими опита, в които всяко от събитията A_i се е наблюдавало n_i пъти ($\sum_{i=1}^{N} n_i = n$, $i = 1, \dots, N$), на което съответства функция на правдоподобие

$$L(p) = \frac{n!}{n_1! \dots n_N!} p_1^{n_1} \dots p_N^{n_N}.$$

Покажете, че при $n \to \infty$ и вярна хипотеза h_0 статистиките $-2 \ln \hat{\lambda}_n(p_0)$ и $\hat{\chi}^2 = \sum_{i=1}^{N} \frac{(n_i - np_i)^2}{np_i}$ клонят към една и съща случайна величина, имаща χ^2-разпределение с $N-1$ степени на свобода. (Тук $\hat{\lambda}_n(p_0) = L(p_0) / \max\limits_{p \in \mathfrak{B}^+} L(p)$ е статистиката, получена по метода с отношенията на правдоподобията.)

Упътване. Покажете, че е изпълнено $-2 \ln \hat{\lambda}_n(p_0) = 2 \sum n_i \ln \frac{n_i}{np_i}$. Като положите $n_i = np_i + Z_i \sqrt{np_i}$, покажете, че $Z_i \xrightarrow[n \to \infty]{} \mathcal{N}(0, q_i)$, $q_i = 1 - p_i$, и е вярно асимптотичното равенство

$$-2 \ln \hat{\lambda}_n(p_0) = 2 \sum_{i=1}^{N} Z_i \sqrt{np_i} + \sum_{i=1}^{N} Z_i^2 + o\left(\frac{1}{\sqrt{n}}\right),$$

в което първото събираемо е равно на 0.

Дванадесета глава

Емпирична функция на разпределение и свързани с нея статистики

§ 1. Вариационен ред и емпирична функция на разпределение

Видяхме, че в задачите за проверка на хипотези основно значение има правилният подбор на разпределенията в извадковото пространство, съответни на основната хипотеза H_0 и на алтернативата H_1. Въобще в задачите на математическата статистика главният проблем е как да определим разпределението, съответстващо на наблюдаваната случайна величина. Теоретичните познания за свойствата на разпределенията, придобити в курса по теория на вероятностите, тук оказват неоценима помощ. Но докато се добере изследователят до аналитичния вид на разпределението на наблюдаваната величина, изминава дълъг път. Той минава непременно през задачите за проверка на хипотези, но започва преди всичко с търсене на оценка за неизвестната функция на разпределение, на графична представа за нейния вид и вида на съответната ѝ плътност на разпределение, докато се идентифицира класът, към който тя ще принадлежи, и се премине към пълно определяне на нейния представител в този клас.

Отново изходен материал за следващите разсъждения са наблюденията

$$\xi_1, \xi_2, \ldots, \xi_n \tag{1}$$

над интересуващата ни случайна величина ξ. Ще предполагаме, че наблюденията представляват проста случайна извадка от стойностите на ξ и че ξ има функция на разпределение $F(x)$. Ще отбележим, че ξ може да бъде k-мерен случаен вектор с функция на разпределение $F(x_1, \ldots, x_k)$, определена в $\mathbf{R}^k$. В този случай извадката $\xi_1, \ldots, \xi_n$ е kn-мерен случаен вектор с функция на разпределение $\prod_{i=1}^{n} F(x_{1i}, \ldots, x_{ki})$, определена в $\mathbf{R}^{nk} = R^k_{(1)} \times \cdots \times R^k_{(n)}$. Извадковото пространство $\mathfrak{X}$ е геометричното място на множеството от точки $\{(\xi_1, \ldots, \xi_n)\}$. В тази глава ще се занимаваме предимно с емпирични изследвания на едномерна случайна величина ξ, в които извадката $\xi_1, \ldots, \xi_n$ е проста.

Наредените във възходящ ред наблюдения

$$\xi_{(1)} \leqq \xi_{(2)} \leqq \cdots \leqq \xi_{(n)} \tag{2}$$

от редицата (1) се наричат *вариационен ред* на наблюденията (наредени, поредни статистики). Величината $\xi_{(k)}$ се нарича *k-та поредна статистика.*

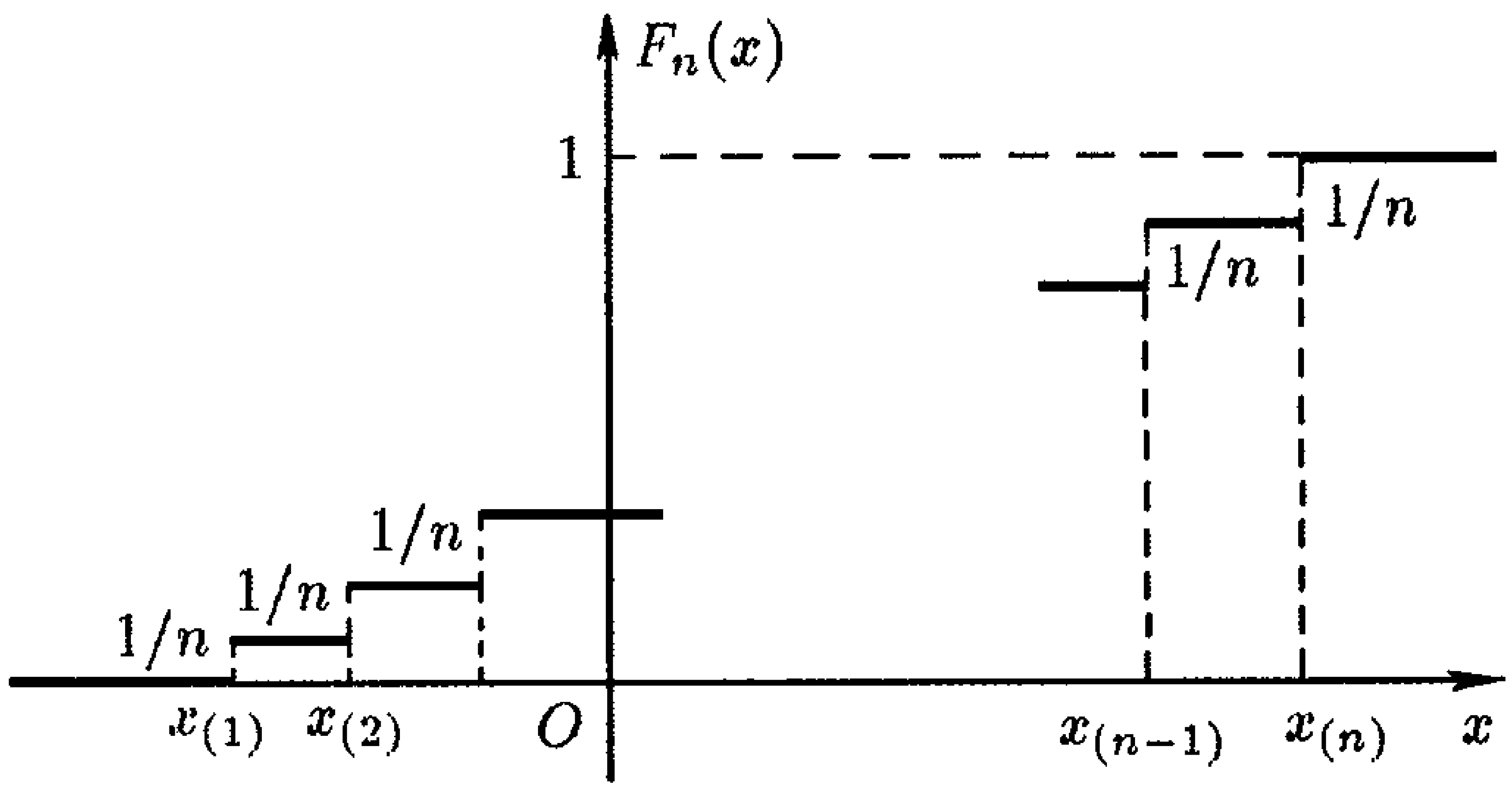
Вариационният ред е първична форма на записване на статистическия материал и може да бъде обработен по различни начини. Един от начините за такава обработка е построяването на емпиричната (статистическа) функция на разпределение на случайна величина ξ.

Функцията, определена с равенствата

$$F_n(x) = \begin{cases} 0 & \text{при} \quad x \leqq \xi_{(1)}, \\ \dfrac{k}{n} & \text{при} \quad \xi_{(k)} < x \leqq \xi_{(k+1)}, \\ 1 & \text{при} \quad x > \xi_{(n)}, \end{cases} \tag{3}$$

се нарича *емпирична функция на разпределение.* От (3) е ясно, че $F_n(x)$ се получава, като на всяка от n-те наблюдавани стойности на ξ се приписва вероятност $\dfrac{1}{n}$, т. е. $F_n(x)$ е разпределението на случайна величина, равномерно разпределена върху числата $\xi_1, \ldots, \xi_n$ или $(\xi_{(1)}, \xi_{(2)}, \ldots, \xi_{(n)})$.

Графиката на $F_n(x)$ е изобразена на фиг. 21.



Фиг. 21

Емпиричната функция на разпределение е монотонна, непрекъсната отляво и има точки на прекъсване само при стойности на аргумента, съвпадащи с членовете на вариационния ред. В точките на прекъсване големината на скока е кратна на $\frac{1}{n}$. Ще забележим, че при всяка стойност на x ординатата $F_n(x)$ е случайна величина, приемаща стойности $0, \frac{1}{n}, \frac{2}{n} \ldots, \frac{n-1}{n}, \frac{n}{n} = 1$.

Ако $F(x)$ е истинската функция на разпределение, то

$$\mathbf{P}\left\{F_n(x) = \frac{k}{n}\right\} = \binom{n}{k}[F(x)]^k\,[1 - F(x)]^{n-k}, \tag{4}$$

което показва, че $F_n(x)$ е разпределена по биномния закон с параметър $F(x)$.

Равенството (4) дава вероятността при n независими опита случайната величина ξ да приеме k стойности, по-малки от x, и $n - k$ стойности, по-големи от x.

Различието между емпиричната и теоретичната функция на разпределение се състои в това, че теоретичната функция $F(x)$ определя вероятността на събитието $\{\xi < x\}$, а емпиричната функция $F_n(x)$ — честотата на това събитие.

От теоремата на Бернули следва, че при неограниченото увеличаване на броя на опитите n честотата на събитието $\{\xi < x\}$ клони по вероятност към вероятността на това събитие. Това означава, че $F_n(x) \xrightarrow{p} F(x)$ при $n \to \infty$.

С други думи, $F_n(x)$ и $F(x)$ малко се отличават една от друга и $F_n(x)$ може да се използва за приближено представяне на $F(x)$.

Пример. Да се построи $F_n(x)$ по дадена извадка:

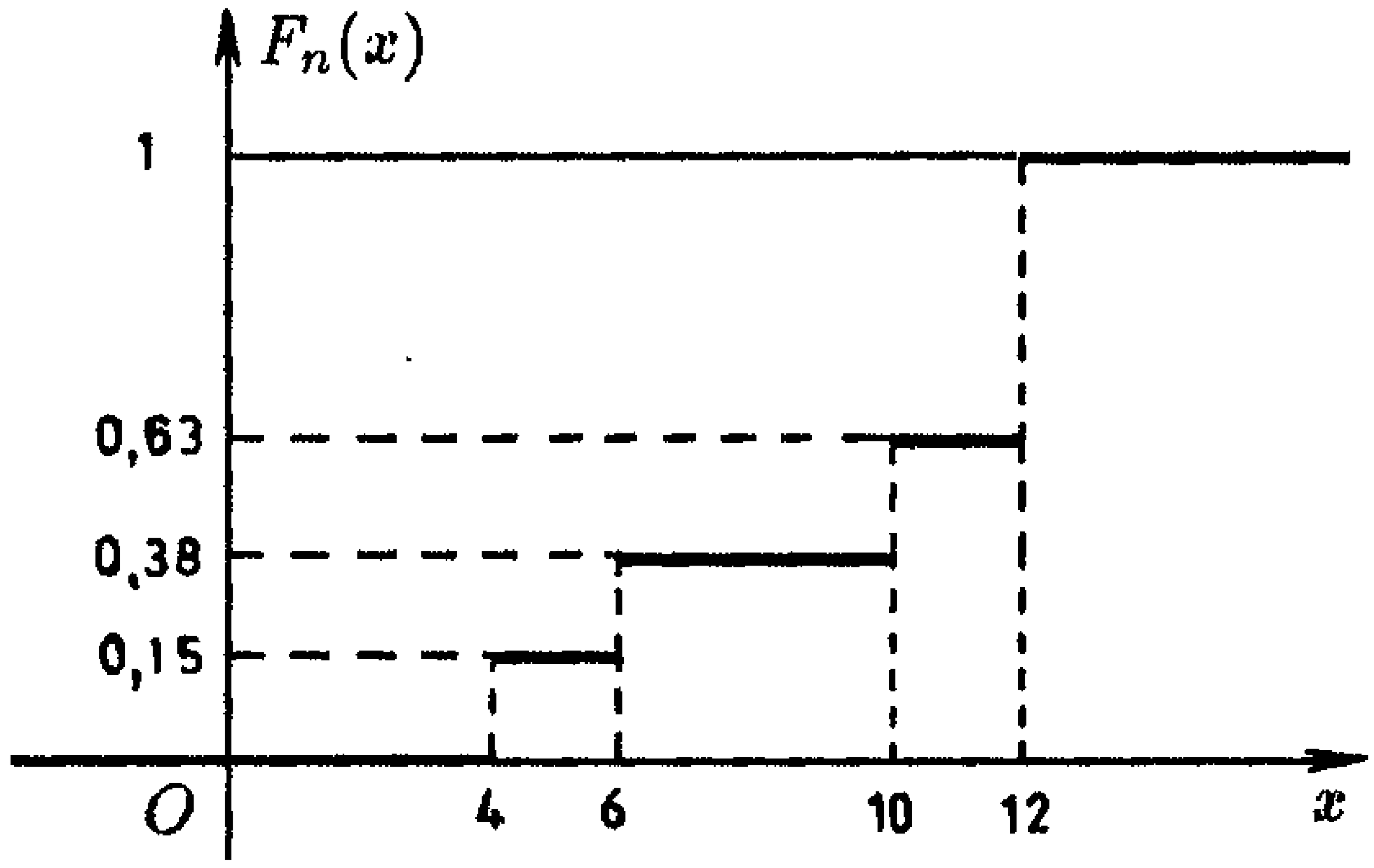
Наблюдения	4	6	10	12
Честота	12	18	20	30

Решение. Намираме обема на извадката: $12 + 18 + 20 + 30 = 80$. Най-малкото наблюдение е равно на 4. Следователно $F_n(x) = 0$ при $x \leqq 4$. Стойност $\xi < 6$, а именно $\xi_{(1)} = 4$, е наблюдавана 12 пъти и следователно $F_n(x) = \frac{12}{80} = 0,15$ при $4 \leqq x < 6$.

Като продължим така за следващите интервали между членовете на вариационния ред, ще видим, че търсената емпирична функция на разпределение е

$$F_n(x) = \begin{cases} 0 & \text{при} \quad x \leqq 4, \\ 0,15 & \text{при} \quad 4 < x \leqq 6, \\ 0,38 & \text{при} \quad 6 < x \leqq 10, \\ 0,63 & \text{при} \quad 10 < x \leqq 12, \\ 1 & \text{при} \quad x > 12. \end{cases}$$

Графиката на тази функция е дадена на фиг. 22.



Фиг. 22

Тук ще забележим, че при голям брой опити построяването на $F_n(x)$ е трудоемка задача, затова е удобно да се използват други характеристики на емпиричната функция на разпределение, аналогични не на $F(x)$, а на плътността на разпределение $f(x)$.

§ 2. Полигон и хистограма

Ако в реда (2) от § 1 се окаже, че някои случайни извадки са равни помежду си, то този ред може да се запише във вида

(1) $$\xi_{(1)}, \xi_{(2)}, \dots, \xi_{(k)}$$

$$n_1, n_2, \dots, n_k,$$

където n_i е броят на повторенията на стойностите $\xi_{(i)}$, $i = 1, \dots, k$.

Сумата $n_1 + \dots + n_k = n$ дава обема на вариационния ред.

Ако в (1) разделим всяко от числата на втората редица на n, то отношенията $\frac{n_i}{n}$ представляват относителните честоти на стойностите $\xi_{(i)}$ и сумата на относителните честоти е $\sum_{i=1}^{k} \frac{n_i}{n} = 1$.

Полигон (графика) на честотите наричаме начупената линия, отсечките на която съединяват точките $(\xi_{(1)}, n_1), (\xi_{(2)}, n_2), \ldots, (\xi_{(k)}, n_k)$. За построяването на полигона на честотите използваме правоъгълна координатна система, като по абсцисната ос нанасяме наблюденията $\xi_{(i)}$, а по ординатната ос — честотите n_i.

Ако съединим с отсечки точките

$$\left(\xi_{(1)}, \frac{n_1}{n}\right), \left(\xi_{(2)}, \frac{n_2}{n}\right), \ldots, \left(\xi_{(k)}, \frac{n_k}{n}\right),$$

получаваме ***полигон на относителните честоти***. Той дава първична представа за геометричната форма на графиката на плътността на разпределение на наблюдаваната случайна величина и затова служи за подсещане (насочване) към предполагаемия клас разпределения на ξ. За същите цели в практиката се използва и хистограмата на разпределението.

Нека $[a, b]$ е един интервал, съдържащ всички наблюдения на ξ, т. е. $a \leqq \xi_{(1)} \leqq \cdots \leqq \xi_{(n)} \leqq b$.

Да разделим интервала $[a, b]$ на части с точки на деление $a = c_0 < c_1 < c_2 < \cdots < c_r = b$. Да означим с Δ_i интервала $[c_{i-1}, c_i)$, $i = 1, \ldots, r$. Нека n_i е броят на наблюденията от реда $\xi_1, \ldots, \xi_n$, попаднали в интервала Δ_i. Функцията

$$\hat{f}(x) = \begin{cases} 0 & \text{при} \quad x \notin [a, b), \\ \dfrac{n_i}{n} \cdot \dfrac{1}{c_i - c_{i-1}} & \text{при} \quad x \in [c_{i-1}, c_i), \end{cases}$$

се нарича ***хистограма на наблюденията***. Най-често се използват хистограми, в които интервалите Δ_i са еднакви.

Хистограмата на наблюденията дава представа за плътността на разпределение на ξ. Нормираните интеграли от хистограмата се наричат ***сплайни***. За да даде сплайнът представа за плътността на разпределение, нормировката се прави по такъв начин, че интегралът от получената функция също да е равен на единица. Познати са сплайни от произволен ред. Хистограмата е сплайн от нулев ред. Сплайнът от първи ред също се нарича полигон на разпределението. Сплайнът от ред s се получава чрез интегриране (с подходящо нормиране) от сплайна от ред $s - 1$.

Пример 2. При изследване на процента на масленост в млякото на 50 крави са получени следните резултати:

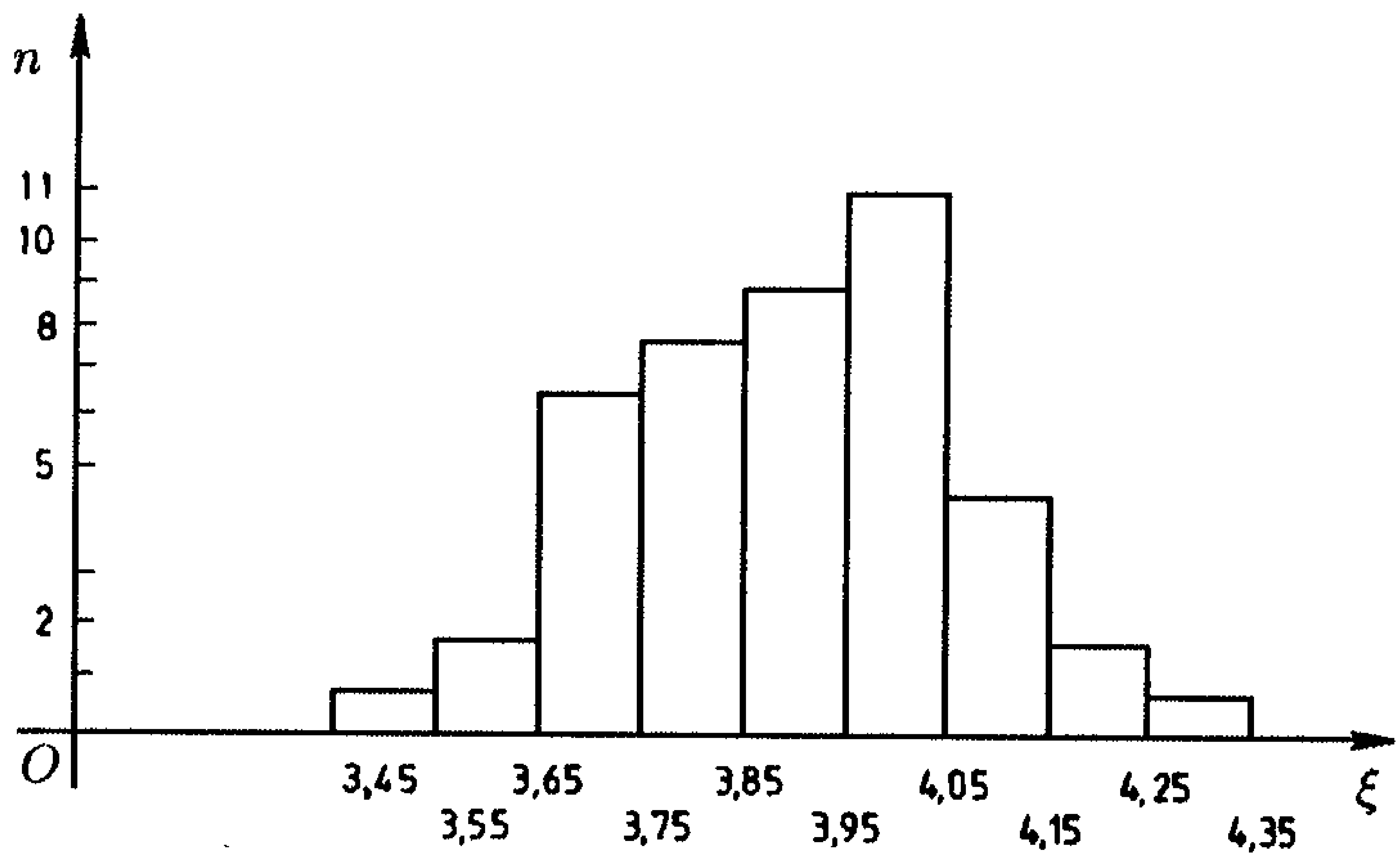
3,86	4,04	3,65	3,97	3,76	3,95	4,04	3,84	3,94	3,61
3,98	3,57	3,87	4,05	3,99	3,76	3,71	3,94	3,82	3,69
4,16	3,76	4,00	3,46	4,08	4,01	3,93	3,71	3,81	3,88
4,02	4,17	3,72	4,09	3,76	3,73	3,52	3,89	3,92	4,02
4,18	4,26	4,03	4,14	3,72	3,82	4,03	3,62	3,91	4,35

По тези данни постройте хистограмата на наблюденията в интервала $[3,45;\ 4,35]$.

Решение. Разделяме интервала $[3,45;\ 4,35]$ на равни части с разлика $0,10$ и съставяме таблица, в която в първия стълб са разположени интервалите, а във втория стълб са намерени честотите за всеки от интервалите.

Процент на маслеността (интервали)	Брой на кравите	Процент на маслеността (интервали)	Брой на кравите
[3,45;3,55]	2	[3,95;4,05]	11
[3,55;3,65]	3	[4,05;4,15]	5
[3,65;3,75]	7	[4,15;4,25]	3
[3,75;3,85]	8	[4,25;4,35]	2
[3,85;3,95]	9	сума	50

Хистограмата на това разпределение е дадена на фиг. 23.



Фиг. 23

§ 3. Числови характеристики на емпиричното разпределение

Законът за разпределение на случайните величини напълно характеризира (описва) случайните величини от вероятностна гледна точка. Обаче при решаването на много практически задачи не е необходима пълна характеристика на случайните величини, а е достатъчно да се дадат само отделни числови характеристики, които определят съществени страни на разпределението на случайните величини. Основни числови характеристики на случайните величини са математическото очакване, дисперсията, квантилите и моментите от различен ред.

Аналогични числови характеристики съществуват и за емпиричното разпределение. Аналогът на математическото очакване на случайна величина ξ е средното аритметично $\overline{\xi}_n$ на редицата $\xi_1, \ldots, \xi_n$ и то може да бъде получено от емпиричната функция на разпределение от § 1 с равенството $\overline{\xi}_n = \int\limits_{-\infty}^{+\infty} x dF_n(x)$.

Аналогът на дисперсията на случайна величина ξ е извадковата дисперсия σ_n^2, която може да се определи така:

$$\sigma_n^2 = \int\limits_{-\infty}^{+\infty} (x - \overline{\xi}_n)^2 dF_n(x) = \frac{1}{n} \sum_{i=1}^{n} (\xi_{(i)} - \overline{\xi}_n)^2.$$

Аналогично се определят емпиричните начални и централни моменти от произволен ред като интеграли от съответната емпирична функция на разпределение, получена по извадката $\xi_1, \ldots, \xi_n$.

Ще забележим, че с увеличаване на броя на наблюденията всички тези средни статистически (емпирични) характеристики клонят по вероятност към съответните числови характеристики на случайната величина ξ, която се наблюдава (ако те съществуват). Например за случайна величина ξ, имаща разпределение на Коши, чиято плътност се задава с формулата

$$f(x) = \frac{1}{\pi} \cdot \frac{b}{b^2 + (x-a)^2},$$

не съществуват $\mathbf{E}\xi$ и $\mathbf{D}\xi$ поради това, че несобствените интеграли са разходящи. Същевременно съответните емпирични моменти съществуват.

Нека $\xi_{(1)}, \ldots, \xi_{(n)}$ е вариационен ред от наблюдения на случайната величина ξ, имаща непрекъсната функция на разпределение $F(x)$. Интервалите $(\xi_{(i)}, \xi_{(i+1)}]$ се наричат *извадкови блокове*, а числата $p_i = F(\xi_{(i)}) - F(\xi_{(i-1)})$ — *части*. Полага се $\xi_{(0)} = -\infty$, $\xi_{(n+1)} = +\infty$. Така извадковите блокове и съответните части, определени от поредните статистики, са определени за всички $i = 0, 1, 2, \ldots, n+1$. При непрекъсната функция на разпределение $F(x)$ всички поредни случайни наблюдения са различни помежду си с вероятност 1. Важна роля в изучаването на общите свойства на емпиричните случайни извадки от непрекъснати случайни величини играе следното твърдение:

Лема 1. *Ако ξ е непрекъсната случайна величина с функция на разпределение $F(x)$, то случайната величина $\eta = F(\xi)$ е равномерно разпределена в $[0, 1]$.*

Доказателство. Имаме

$$\mathbf{P}\{\eta < y\} = \mathbf{P}\{F(\xi) < y\} = \begin{cases} 0 & \text{при} \quad y < 0, \\ y & \text{при} \quad y \in [0, 1], \\ 1 & \text{при} \quad y > 1, \end{cases}$$

защото $\mathbf{P}\{F(\xi) < y\} = \mathbf{P}\{\xi < F^{-1}(y)\} = F(F^{-1}(y)) = y$.

На вариационния ред $\xi_{(1)}, \ldots, \xi_{(n)}$ в такъв случай съответства вариационния ред $\eta_{(1)} \leqq \cdots \leqq \eta_{(n)}$ от n наблюдения на равномерната в $[0, 1]$ случайна величина $\eta = F(\xi)$.

Лема 2. *При условията на лема 1, ако $\xi_{(1)}, \ldots, \xi_{(n)}$ е вариационен ред, получен от проста случайна извадка $\xi_1, \ldots, \xi_n$, то векторът $(F(\xi_{(1)}), \ldots, F(\xi_{(n)}))$ има наредено n-мерно разпределение на Дирихле $\mathfrak{D}^*(1, \ldots, 1; 1)$.*

Доказателство. Ненаредено k-мерното разпределение на Дирихле се задава със своята вероятностна плътност на разпределение

$$p(x_1, \ldots, x_k) \tag{1}$$
$$= \frac{\Gamma(\lambda_1 + \cdots + \lambda_{k+1})}{\Gamma(\lambda_1) \ldots \Gamma(\lambda_{k+1})} x_1^{\lambda_1 - 1} \ldots x_k^{\lambda_k - 1} (1 - x_1 - \cdots - x_k)^{\lambda_{k+1} - 1}$$

във всяка точка на симплекса $S_k = \left\{ \vec{x} :\ x_i \geqq 0,\ \sum\limits_{i=1}^{k} x_i \leqq 1 \right\}$ и тя е равна на 0 в другите точки на $\mathbf{R}^k$. Тук параметрите λ_i са неотрицателни. Разпределението на Дирихле (1) е многомерен аналог на бета-разпределението: $\mathfrak{D}(\lambda_1;\lambda_2) = B(\lambda_1,\lambda_2)$. Неговите характеристики са:

$$\mathbf{E}\xi_i = \frac{\lambda_i}{\lambda_1+\cdots+\lambda_{k+1}}, \quad i = 1, \ldots, k;$$

$$\mathbf{D}\xi_i = \frac{\lambda_i(\lambda_1+\cdots+\lambda_{k+1}-\lambda_i)}{\left[\sum\limits_{i=1}^{k+1}\lambda_i\right]^2\left[\sum\limits_{i=1}^{k+1}\lambda_i+1\right]}, \quad i = 1, \ldots, k;$$

$$\operatorname{cov}(\xi_i,\xi_j) = \frac{\lambda_i\lambda_j}{\left[\sum\limits_{i=1}^{k+1}\lambda_i\right]^2\left[\sum\limits_{i=1}^{k+1}\lambda_i+1\right]}, \quad i \neq j.$$

Ако извадката $\xi_1, \ldots, \xi_k$ има k-мерно разпределение на Дирихле $\mathfrak{D}(\lambda_1,\ldots,\lambda_k;\lambda_{k+1})$, случайният вектор $(\eta_{(1)},\ldots,\eta_{(k)})$, получен от равенствата $\eta_{(1)} = \xi_1$, $\eta_{(2)} = \xi_1+\xi_2, \ldots, \eta_{(k)} = \xi_1+\cdots+\xi_k$ има наредено разпределение на Дирихле $\mathfrak{D}^*(\lambda_1,\ldots,\lambda_k;\lambda_{k+1})$, чиято плътност на разпределение е (покажете това!)

$$p^*(y_1, \ldots, y_k)$$

$$= \frac{\Gamma(\lambda_1+\cdots+\lambda_{k+1})}{\Gamma(\lambda_1)\cdots\Gamma(\lambda_{k+1})} y_1^{\lambda_1-1}(y_2-y_1)^{\lambda_2-1}\cdots(y_k-y_{k-1})^{\lambda_k-1}(1-y_k)^{\lambda_{k+1}-1},$$

определена за $0 < y_1 < y_2 < \cdots < y_k < 1$.

По-нататъшни сведения вижте в задачи 5 – 7 на упражненията към тази глава.

За да докажем твърдението на лема 2, е достатъчно да отбележим, че при $\lambda_1 = \cdots = \lambda_{k+1} = 1$ елементарната вероятност на n-мерното наредено разпределение на Дирихле е

$$n!dy_1dy_2\ldots dy_k \quad (0 \leqq y_1 \leqq \cdots \leqq y_k < 1),$$

и това съвпада с елементарната вероятност да се наблюдават точно стойностите $\eta_{(1)} = y_1 < \cdots < \eta_{(k)} = y_k$ на вариационен ред от k независими наблюдения $y_1, \ldots, y_k$ на случайната величина η, равномерно разпределена в $[0,1]$.

Теорема 1. *Случайните величини $F(\xi_{(k)})$ имат бета-разпределение $B(k,n-k+1)$, $k = 1, 2, \ldots, n$.*

Доказателство. k-тата величина в нареденото разпределение на Дирихле $\eta_{(k)} = F(\xi_{(k)})$ може да бъде представена (вж. доказателството на лема 2) във вида $\eta_{(k)} = \eta_1+\cdots+\eta_k$, $k = 1, 2, \ldots, n$, където $(\eta_1, \ldots, \eta_n) \in \mathfrak{D}(\lambda_1, \ldots, \lambda_n;\lambda_{n+1})$. Но съгласно зад. 5 и 6 сумата $\eta_{(k)} \in B(\lambda_1+\lambda_2+\cdots+\lambda_k; \lambda_{k+1}+\cdots+\lambda_{n+1})$. Съгласно лема 2 имаме $\lambda_1=\lambda_2=\cdots=\lambda_{n+1}=1$, откъдето следва твърдението.

Теорема 2. *Елементарната вероятност* $\mathbf{P}\{\xi_{(k)} \in (x, x+dx]\}$ *за k-тата поредна статистика на непрекъснатата случайна величина ξ е*

$$(2)\quad \mathbf{P}\{\xi_{(k)} \in (x, x+dx]\} = \frac{n!}{(k-1)!(n-k)!}[F(x)]^{k-1}[1-F(x)]^{n-k} f(x)dx.$$

Доказателство. Нека $G_{nk}(x) = \mathbf{P}\{\xi_{(k)} < x\}$. Изпълнено е

$$G_{nk}(x) = \mathbf{P}\{\xi_{(k)} < x\} = \mathbf{P}\left\{F_n(x) \geqq \frac{k}{n}\right\} = \sum_{r=k}^{n} \binom{n}{r}[F(x)]^r[1-F(x)]^{n-r},$$

където $F_n(x)$ е емпиричната функция на разпределение. Оттук също намираме, че

$$G_{nk}(x) = \frac{n!}{(k-1)!(n-k)!} \int_0^{F(x)} y^{k-1}(1-y)^{n-k}dy.$$

И двете представяния дават израза (2) за търсената елементарна вероятност $dG_{nk}(x) = G'_{nk}(x)dx$, а тя от своя страна определя и плътността на разпределение $G'(x)$ на k-тата поредна статистика $\xi_{(k)}$.

Теорема 3. *Нека $\xi_{(1)}, \ldots, \xi_{(n)}$ са поредните статистики от извадка с непрекъсната функция на разпределение $F(x)$. Тогава величините*

$$u_1 = F(\xi_{(1)}),\quad u_2 = F(\xi_{(2)}) - F(\xi_{(1)}),\ \ldots,\ u_n = F(\xi_{(n)}) - F(\xi_{(n-1)})$$

имат n-мерно разпределение на Дирихле $\mathfrak{D}(1, \ldots, 1; 1)$.

Доказателство. Използваме доказателството на лема 2. Величините $u_1, \ldots, u_n$, получени от трансформацията

$$u_{(1)} = u_1,\quad u_{(2)} = u_1 + u_2,\ \ldots,\quad u_{(n)} = u_1 + \cdots + u_n,$$

имат елементарна вероятност $n!du_1 \ldots du_n$ в симплекса

$$S_n = \left\{\vec{u}:\ u_i \geqq 0,\ \sum_{i=1}^{n} u_i \leqq 1\right\},$$

която е 0 вън от S_n. А това е n-мерното разпределение на Дирихле $\mathfrak{D}(1, \ldots, 1; 1)$. Освен това $u_1, \ldots, u_n$ и тяхното разпределение не зависят от $F(x)$.

Теорема 4. *Ако* $r_1 > 0, \dots, r_s > 0$, $r_1 + \dots + r_s \leqq n$ *са дадени числа, то случайните величини* $F(\xi_{(r_1)})$, $F(\xi_{(r_1+r_2)})$, $\dots$, $F(\xi_{(r_1+\dots+r_s)})$ *имат наредено s-мерно разпределение на Дирихле*

$$\mathfrak{D}^*(r_1, \dots, r_s; n - r_1 - \dots - r_s).$$

Доказателство. Следва от резултата на зад. 7 и лема 2.

Нека x_p е p-квантилът на непрекъснатото разпределение $F(x)$ и $\xi_1, \dots, \xi_n$ е проста извадка.

Теорема 5. *За произволни цели числа* $k > 0$, $m > 0$, $k + m \leqq n$ *вероятността интервалът* $(\xi_{(k)}, \xi_{(k+m)})$ *да съдържа квантила* x_p *не зависи от* $F(x)$ *и се задава с израза*

(3) $\mathbf{P}\{\xi_{(k)} < x_p < \xi_{(k+m)}\} = B_p(k, n-k+1) - B_p(k+m, n-k-m+1)$,

където $B_p(k,m) = \dfrac{\Gamma(k+m)}{\Gamma(k)\Gamma(m)} \int\limits_0^p x^{k-1}(1-x)^{m-1}dx$ *е непълната бета-функция.*

Доказателство. Неравенствата $\xi_{(k)} < x_p < \xi_{(k+m)}$ са изпълнени само когато $F(\xi_{(k)}) < p < F(\xi_{(k+m)})$. Следователно вероятността на лявата страна на (3) е равна на $\mathbf{P}\{\eta_{(k)} < p < \eta_{(k+m)}\}$, което не зависи от $F(x)$, тъй като η_i са наблюдения от равномерно $[0,1]$-разпределение. Обаче статистиките $\eta_{(k)}, \eta_{(k+m)}$ имат съгласно теорема 4 съвместно наредено двумерно разпределение на Дирихле $\mathfrak{D}^*(k, m, n-k-m+1)$, чиято плътност на разпределение е

$$f(u,v) = \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(m)\Gamma(n-k-m+1)} u^{k-1}(v-u)^{m-1}(1-v)^{n-k-m+1}$$

при $0 < u < v < 1$ и 0 в другите случаи за (u, v). Затова

$$\mathbf{P}\{\eta_{(k)} < p < \eta_{(k+m)}\} = \int\limits_p^1 \int\limits_0^p f(u,v)dudv$$

$$= \int\limits_0^p \int\limits_u^1 f(u,v)dvdu - \int\limits_0^p \int\limits_0^v f(u,v)dudv.$$

Като направим смяна на променливите $u = t$, $v = 1 - s(1-t)$ в първия интеграл и смяната $u = st$, $v = s$ във втория интеграл, последното равенство ще ни убеди във верността на твърдението.

Теорема 5 показва, че кои да са две от поредните статистики $\xi_{(k)}$ и $\xi_{(k+m)}$ могат да служат за доверителен интервал на квантила x_p на функцията $F(x)$, като доверителната му вероятност е определена с дясната страна на (3). Затова при дадено p трябва да се търсят най-подходящите k и m, които дават най-близки номера (т. е. m да е възможно най-малкото число). (Вж. зад. 9.)

§ 4. Теорема на Гливенко – Кантели

Съгласно усиления закон за големите числа (теоремата на Борел – Колмогоров) честотата $F_n(x)$ на събитието $\{\xi < x\}$ с вероятност 1 клони към теоретичната вероятност $F(x) = \mathbf{P}\{\xi < x\}$, т. е. когато обемът на извадката расте, изпълнено е

$$\mathbf{P}\left\{F_n(x) \underset{n\to\infty}{\longrightarrow} F(x)\right\} = 1$$

за всяко фиксирано x. Вярно е обаче по-силното твърдение:

Теорема на Гливенко – Кантели. *Ако $F(x)$ е истинската функция на разпределение на случайната величина ξ, а $F_n(x)$ — емпиричната функция на разпределение на ξ, получена по n независими наблюдения, то*

$$\mathbf{P}\left\{\sup_{-\infty<x<\infty} |F_n(x) - F(x)| \underset{n\to\infty}{\longrightarrow} 0\right\} = 1.$$

Доказателство. А. Да въведем означенията

$$x_{j,k} = \inf\left\{x:\ F(x-0) = F(x) \leqq \frac{j}{k} \leqq F(x+0)\right\}$$

за всяко цяло $k > 0$ и $j = 1, 2, \dots, k$. Тогава $F(x_{j,k}) \leqq \dfrac{j}{k} \leqq F(x_{j,k}+0)$, а по усиления закон за големите числа събитията

$$A_{j,k}^{(1)} = \left\{F_n(x_{j,k}) \underset{n\to\infty}{\longrightarrow} F(x_{j,k})\right\}$$

имат вероятност 1, т. е. $\mathbf{P}\{A_{j,k}^{(1)}\} = 1$. По същия начин и събитията

$$A_{j,k}^{(2)} = \left\{F_n(x_{j,k}+0) \underset{n\to\infty}{\longrightarrow} F(x_{j,k}+0)\right\}$$

имат вероятност 1, т. е. $\mathbf{P}\{A_{j,k}^{(2)}\} = 1$. Нека при $\varepsilon = \pm 0$

$$A_{j,k} = A_{j,k}^{(1)} \cap A_{j,k}^{(2)}$$

и

$$A_k = \bigcap_{j=1}^{k} A_{j,k} = \left\{ \sup_{1 \leqq j \leqq k} |F_n(x_{j,k} + \varepsilon) - F(x_{j,k} + \varepsilon)| \xrightarrow[n\to\infty]{} 0 \right\}.$$

Събитията $A_{j,k}$ и A_k имат вероятност 1 за всяко k. Наистина

$$\mathbf{P}\{A_{j,k}\} = \mathbf{P}\left\{A_{j,k}^{(1)}\right\} + \mathbf{P}\left\{A_{j,k}^{(2)}\right\} - \mathbf{P}\left\{A_{j,k}^{(1)} \cup A_{j,k}^{(2)}\right\} = 1,$$

следователно $\mathbf{P}\{\overline{A}_{j,k}\} = 0$. Тогава

$$\mathbf{P}\{\overline{A}_k\} = \mathbf{P}\left\{\bigcup_{j=1}^{k} \overline{A}_{j,k}\right\} \leqq \sum_{j=1}^{k} \mathbf{P}\{\overline{A}_{j,k}\} = 0,$$

т. е.

$$\mathbf{P}\{A_k\} = 1 - \mathbf{P}\{\overline{A}_k\} = 1.$$

Б. За всяко x, лежащо между $x_{j,k}$ и $x_{j+1,k}$ са изпълнени неравенствата

$$F(x_{j,k} + 0) \leqq F(x) \leqq F(x_{j+1,k});$$

$$F_n(x_{j,k} + 0) \leqq F_n(x) \leqq F_n(x_{j+1,k}).$$

При това $0 \leqq F(x_{j+1,k}) - F(x_{j,k} + 0) \leqq \frac{1}{k}$. Следователно

$$F_n(x) - F(x) \leqq F_n(x_{j+1,k}) - F(x_{j,k} + 0) \leqq F_n(x_{j+1,k}) - F(x_{j+1,k}) + \frac{1}{k};$$

$$F_n(x) - F(x) \geqq F_n(x_{j,k} + 0) - F(x_{j+1,k}) \geqq F_n(x_{j,k} + 0) - F(x_{j,k} + 0) - \frac{1}{k}.$$

Така заключаваме, че за всяко $k \geqq 0$, всяко x и $\varepsilon = \pm 0$ имаме

$$|F_n(x) - F(x)| \leqq \sup_{1 \leqq j \leqq k} |F_n(x_{j,k} + \varepsilon) - F(x_{j,k} + \varepsilon)| + \frac{1}{k}.$$

Следователно изпълнено е

$$D_n = \sup_{-\infty < x < \infty} |F_n(x) - F(x)| \leqq \sup_{1 \leqq j \leqq k} |F_n(x_{j,k} + \varepsilon) - F(x_{j,k} + \varepsilon)| + \frac{1}{k}.$$

Нека $B=\left\{D_n \xrightarrow[n\to\infty]{} 0\right\}$. Очевидно $A=\bigcap\limits_{k=1}^{\infty} A_k \subset B$ и понеже $\mathbf{P}(A) = 1$, то и $\mathbf{P}(B) = 1$.

§ 5. Критерий на Колмогоров

Теоремата на Гливенко – Кантели има всички недостатъци на усилен закон за големите числа — не дава точна оценка за вероятностите на възможните отклонения на $F_n(x)$ от $F(x)$, нито за скоростта на сближаване, а само установява тяхната близост. Затова възниква задачата за определяне на асимптотичното разпределение на случайната величина

$$D_n = \sup_{-\infty<x<\infty} |F_n(x) - F(x)|.$$

Задачата е решена от А. Н. Колмогоров в 1933 г. и решението се дава със следната:

Теорема. *Ако функцията на разпределение $F(x)$ на случайната величина ξ е непрекъсната, то при $n \to \infty$ е изпълнено*

$$\lim_{n\to\infty} \mathbf{P}\{\sqrt{n}D_n < z\} = K(z) = \begin{cases} \sum\limits_{k=-\infty}^{\infty} (-1)^k\, e^{-2k^2z^2} & \text{за} \quad z > 0; \\ 0 & \text{за} \quad z \leqq 0. \end{cases}$$

Понякога се записва

$$K(z) = 1 - 2\sum_{k=1}^{\infty}(-1)^{k-1}\, e^{-2k^2z^2}.$$

Няма да се спираме на доказателството на теоремата на Колмогоров. Ще покажем само как тя може да се използва за проверка на параметрични статистически хипотези.

Досега разглеждахме параметрични хипотези, при които множеството на допустимите разпределения, съответстващи на проверяемите хипотези, има вида

$$L(x) = \{L(x\,|\,\theta),\ \theta \in \Theta\}, \quad \Theta \subset \mathbf{R}^k.$$

Нека $\mathfrak{C}$ е множеството на всички непрекъснати функции на разпределение и $L_0 \in \mathfrak{C}$. Задачата за проверка на основната хипотеза $H_0 : L(x) = L_0(x)$ срещу алтернативите $H_1 : L(x) \neq L_0(x)$ с ниво на съгласие $\alpha \in (0,1)$ е задача за проверка на проста непараметрична хипотеза.

Задачата за проверка на сложни непараметрични статистически хипотези би могла да има и следния частен вариант: Ако $C_1 \subset \mathfrak{C}$, основната хипотеза е $H_0 : L(x) \in C_1$ срещу алтернативите $H_1 : L(x) \in \mathfrak{C} - C_1$.

Критерий на Колмогоров. Нека $\xi_1, \ldots, \xi_n$ са независими наблюдения над случайната величина ξ с функция на разпределение $F(x) \in \mathfrak{C}$. Да се построи критерий за проверка на хипотезата

$$H_0 : F(x) = F_0(x) \in \mathfrak{C}$$

срещу алтернативите $H_1 : F(x) \neq F_0(x)$ с ниво на съгласие $\alpha \in (0,1)$.

Решение. Построяваме емпиричната функция на разпределение $F_n(x)$ и пресмятаме

$$D_n^0 = \sup_{-\infty<x<\infty} |F_n(x) - F_0(x)| = \max_{1 \leqq k \leqq n} |F_n(\xi_{(k)} + \varepsilon) - F_0(\xi_{(k)})|, \quad \varepsilon = \pm 0.$$

При достатъчно големи n имаме

$$\mathbf{P}\{\sqrt{n} D_n^0 < z \,|\, H_0\} \approx K(z).$$

Определяме $z_0 = z_0(\alpha)$ от уравнението $1 - K(z_0) = \alpha$, т. е.

$$\mathbf{P}\{\sqrt{n} D_n^0 \geqq z_0(\alpha) \,|\, H_0\} \approx 1 - K(z_0) = \alpha.$$

Следователно, ако след извършване на наблюденията се осъществи събитието $\{\sqrt{n} D_n^0 > z_0(\alpha)\}$, което има малката (предварително избрана) вероятност α, трябва да отхвърлим хипотезата H_0, че истинската функция на разпределение на ξ е $F_0(x)$. Ако се окаже, че $\sqrt{n} D_n^0 < z_0(\alpha)$, няма основания да се отхвърли хипотезата.

Да отбележим, че $\sqrt{n} D_n^0 \to \infty$, ако хипотезата H_0 не е вярна.

§ 6. Критерий на Смирнов

Теоремата на Гливенко – Кантели показва, че при големи n емпиричната функция на разпределение $F_n(x)$ на наблюдаваната случайна величина ξ при независими наблюдения е много близка до истинската ѝ функция на разпределение $F(x)$. Ето защо редно е да се очаква, че при големи n и m ще бъдат близки емпиричните функции на разпределение $F_n(x)$ и $F_m(x)$, построени по две отделни групи от наблюдения $\xi_1', \ldots, \xi_n'$ и $\xi_1'', \ldots, \xi_m''$ над величината ξ.

Следователно, ако са дадени две независими помежду си серии от наблюдения

(1) $$\xi_1, \xi_2, \dots, \xi_n;$$

(2) $$\eta_1, \eta_2, \dots, \eta_m$$

над величините ξ и η, разликата

(3) $$D_{n,m} = \sup_{-\infty<x<\infty} |F_n(x) - G_m(x)|$$

между емпиричните функции на разпределение $F_n(x)$ и $G_m(x)$, получени от двата реда (1) и (2), може да служи като обективен показател за близостта на разпределенията $F(x) = \mathbf{P}\{\xi < x\}$ и $G(x) = \mathbf{P}\{\eta < x\}$.

Най-често величината $D_{n,m}$ служи за построяване на критерий за проверка на хипотезата за равенство на $F(x)$ и $G(x)$. В редица задачи такава хипотеза се откроява като особено съществена.

Пример 1. Две предприятия произвеждат едни и същи изделия и твърдят, че между тяхната продукция няма никакви различия и изделията им (телевизори, автомобили) трябва да се продават на еднакви цени. Вярно ли е такова твърдение или не?

Пример 2. Двама изследователи провеждат поотделно и независимо един от друг еднакви експерименти. Кога може да се твърди, че експериментите се провеждат при еднакви условия? Дори при провеждане на продължителен експеримент възможно ли е по измервания в началото му и по измервания на друг негов етап да се установи, че условията на експеримента са се променили?

Задачата е от съществено значение за непрекъснати технологични процеси, произвеждащи еднородна продукция, където условията на производство дават отражение върху качеството на продукцията.

В редица задачи за сравняване на еднородни обекти се използва показателят $D_{n,m}$ — например при сравняване на продуктивността на различни сортове от едно растение, на нова методология на преподаване, на нова организация на управлението, на лекуването с ново лекарство и др.

Формално задачата се поставя така:

Пример 3. Нека ξ и η са независими случайни величини с непрекъснати функции на разпределение съответно $F(x)$ и $G(x)$. Нека редът (1) е проста случайна извадка от стойности на ξ, а редът

(2) — проста случайна извадка от стойности на η. Да се построи критерий за проверка на основната хипотеза $H_0 : F(x) = G(x)$ за всяко x и $F, G \in \mathfrak{C}$, срещу алтернативата $H_1 : F(x) \neq G(x)$ с ниво на съгласие α.

За решаване на поставената задача се предлага следната процедура: По наблюденията (1) и (2) се строят емпиричните функции на разпределение $F_n(x)$ и $G_m(x)$. Полага се

(4) $$D_{n,m} = \sup_{-\infty<x<\infty} |F_n(x) - G_m(x)|,$$

(5) $$D^+_{n,m} = \sup_{-\infty<x<\infty} (F_n(x) - G_m(x))$$

и се използва следната теорема, обикновено свързана с името на Н. В. Смирнов.

Теорема. *Ако* (1) *и* (2) *са независими наблюдения с една и съща непрекъсната функция на разпределение* $F(x)$, *то при условието* $\lim\limits_{n\to\infty} \frac{m}{n} = \rho$, $0 < \rho < \infty$, *е изпълнено:*

а) $$\lim_{n\to\infty} \mathbf{P}\left\{\sqrt{\frac{mn}{m+n}}\, D^+_{n,m} < z\right\} = \begin{cases} 0 & \text{за} \quad z \leqq 0; \\ 1 - e^{-2z^2} & \text{за} \quad z > 0; \end{cases}$$

б) $$\lim_{n\to\infty} \mathbf{P}\left\{\sqrt{\frac{mn}{m+n}}\, D_{n,m} < z\right\} = \begin{cases} 0 & \text{за} \quad z \leqq 0; \\ K(z) & \text{за} \quad z > 0. \end{cases}$$

Разпределението $1 - e^{-2z^2}$, $z > 0$, се нарича разпределение на Смирнов, а $K(z)$ е разпределението на Колмогоров, дадено в § 5.

По аналогичен начин, както се използва теоремата на Колмогоров, се използват и резултатите на Смирнов. Строят се критични области от вида

$$\left\{\sqrt{\frac{mn}{m+n}}\, D^+_{n,m} < z^+_\alpha\right\}; \quad \left\{\sqrt{\frac{mn}{m+n}}\, D_{n,m} \geqq z_0(\alpha)\right\},$$

където числата z^+_α и $z_0(\alpha)$ се определят от условията

$$\mathbf{P}\left\{\sqrt{\frac{mn}{m+n}}\, D^+_{n,m} \geqq z^+_\alpha \mid H_0\right\} \approx 1 - \exp\{-2(z^+_\alpha)^2\} = \alpha,$$

$$\mathbf{P}\left\{\sqrt{\frac{mn}{m+n}}\, D_{n,m} \geqq z_0(\alpha) \mid H_0\right\} \approx 1 - K(z_0(\alpha)) = \alpha.$$

Ако в резултат на наблюденията се получат емпирични стойности на $D^+_{n,m}$ и $D_{n,m}$, които, нормирани с множителя $\sqrt{\dfrac{mn}{m+n}}$, се окажат по-големи съответно от z^+_α и $z_0(\alpha)$, хипотезата H_0 за равенство на $F(x)$ и $G(x)$ следва да се отхвърли като противоречаща (несъгласувана) с резултата от експеримента при ниво на съгласие α. В противен случай няма основания тя да бъде отхвърлена. При използване на посочената процедура в $\alpha\%$ от случаите ще бъде отхвърлена правилната хипотеза H_0, което при достатъчно малко α е допустимо и приемливо.

В случаите, когато се ползва едностранният критерий

$$\left\{\sqrt{\frac{mn}{m+n}}\,D^+_{n,m} \geqq z^+_\alpha\right\},$$

едновременно с отхвърлянето на хипотезата H_0 би могло да се направи заключение, че въобще $\sup\limits_{-\infty<x<\infty}(F(x)-G(x))>0$, което говори за възможност случайната величина ξ да е стохастически по-малка от случайната величина η.

§ 7. Теорема на Гнеденко – Королюк

Ще докажем теоремата на Смирнов в частния случай, когато $m=n$, като ще намерим за този случай даже точните разпределения на величините $D^+_{n,n}=D^+_{n,m}$ и $D_{n,n}$. Полученото частно твърдение е доказано от Б. В. Гнеденко и В. С. Королюк през 1952 г. и носи тяхното име.

Теорема на Гнеденко – Королюк. *Нека $F_n(x)$ и $G_n(x)$ са емпиричните функции на разпределение, получени от редиците наблюдения $\xi_1, \ldots, \xi_n$ и $\eta_1, \ldots, \eta_n$ над случайните величини ξ и η с непрекъснати функции на разпределение съответно $F(x)$ и $G(x)$. Нека $D_{n,m}$ и $D^+_{n,m}$ са дефинирани с равенствата* (6.4) *и* (6.5) *и нека*

$$\Phi_n(z)=\mathbf{P}\left\{\sqrt{\frac{n}{2}}\,D_{n,n}<z\right\}; \tag{1}$$

$$\Phi^+_n(z)=\mathbf{P}\left\{\sqrt{\frac{n}{2}}\,D^+_{n,n}<z\right\}. \tag{2}$$

Ако хипотезата $H_0 : F(x) = G(x)$ е вярна, функциите $\Phi_n(z)$ и $\Phi_n^+(z)$ се определят с изразите

$$\Phi_n(z) = \begin{cases} 0 & \text{за } z < \frac{1}{\sqrt{2n}}, \\ 1 - 2\sum_{k=1}^{[n/c]}(-1)^{k-1}\frac{\binom{2n}{n-kc}}{\binom{2n}{n}} & \text{за } z \in \left(\frac{1}{\sqrt{2n}}, \sqrt{\frac{n}{2}}\right), \\ 1 & \text{за } z > \sqrt{\frac{n}{2}}, \end{cases}$$

$$\Phi_n^+(z) = \begin{cases} 0 & \text{за } z \leqq 0, \\ 1 - \binom{2n}{n-c} & \text{за } 0 < z \leqq \sqrt{\frac{n}{2}}, \\ 1 & \text{за } z > \sqrt{\frac{n}{2}}, \end{cases}$$

където е положено $c =]z\sqrt{2n}[$ (с $]x[$ е означено най-малкото цяло число, не по-малко от числото x, а с $[x]$ — най-голямото цяло число, ненадминаващо x).

Доказателство. Ще означим двете извадки $(\xi) = (\xi_1, \ldots, \xi_n)$ и $(\eta) = (\eta_1, \ldots, \eta_n)$. Да образуваме общия вариационен ред на $n + n = 2n$ наблюдения на (ξ) и (η). Нека това е редът

$$(\zeta) = (\zeta_{(1)} < \zeta_{(2)} < \cdots < \zeta_{(2n)}).$$

Върху реда (ζ) дефинираме случайните величини $z_1, z_2, \ldots, z_{2n}$ с равенството

$$z_k = \begin{cases} 1, & \text{ако } \zeta_{(k)} \text{ е наблюдение от } (\xi), \\ -1, & \text{ако } \zeta_{(k)} \text{ е наблюдение от } (\eta). \end{cases}$$

Когато е вярна хипотезата H_0, никои две числа от (ξ) и (η) не могат да бъдат равни, а случайните величини z_k са независими, еднакво разпределени и

$$\mathbf{P}\{z_k = +1\} = \mathbf{P}\{z_k = -1\} = \frac{1}{2}.$$

Да положим

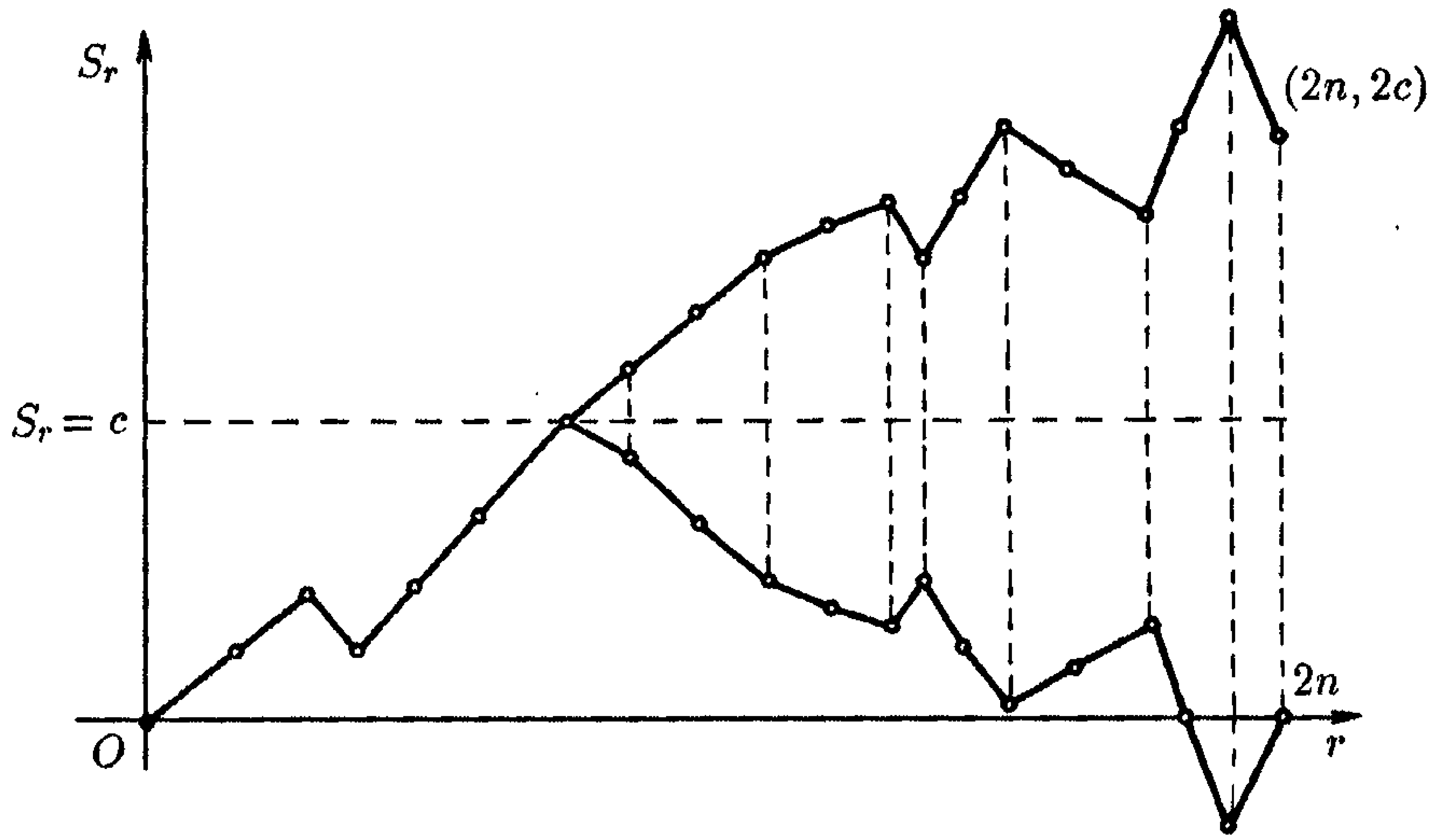
$$S_0 = 0; \quad S_r = z_1 + z_2 + \cdots + z_r.$$

Тогава очевидно е изпълнено

$$nD_{n,n}^+ = \sup_{0 \leqq r \leqq 2n} S_r; \quad nD_{n,n} = \sup_{0 \leqq r \leqq 2n} |S_r|, \tag{3}$$

които равенства ще ни послужат при определяне разпределенията на величините $D^+_{n,n}$ и $D_{n,n}$. За целта ще ползваме геометричната интерпретация на редицата числа (r, S_r), $r = 0, 1, \ldots, 2n$, като случайно лутане в равнината.

Да нанесем върху абсцисната ос на правоъгълна координатна система в равнината числата r, а върху ординатната ос — сумите S_r. Тогава редицата (r, S_r) при изменението на r от 0 до $2n$ ще описва съвкупност от точки, които съединени последователно с отсечки, ще образуват начупена линия, излизаща от точката $(0,0)$ и завършваща в точката $(2n, 0)$. Всяка такава начупена линия ще наричаме траектория на случайното лутане (фиг. 24).



Фиг. 24

Броят на всички траектории с единични скокове, съединяващи точките $(0,0)$ и $(2n, 0)$, е равен на броя на всички редици от $2n$ символа (по n единици и n минус единици), които са $\binom{2n}{n}$. Въобще броят на траекториите с единични скокове, съединяващи точката $(0,0)$ с точката (m, k), е

$$\binom{m}{(m-k)/2} = \binom{m}{(m+k)/2}.$$

Такава траектория съдържа m, символа "1" и "-1" и ако броят на единиците е l, то броят x на минус единиците трябва да удовлетворява условията

$$\left| \begin{array}{l} 1 \cdot l + x \cdot (-1) = k, \\ \quad l + x = m, \end{array} \right.$$

откъдето определяме, че $l = \dfrac{m+k}{2}$, а $x = \dfrac{m-k}{2}$.

При търсене вероятностите на неравенства от вида

$$\left\{ \sup_{0 \leqq r \leqq 2n} S_r < c \right\} \quad \text{и} \quad \left\{ \sup_{0 \leqq r \leqq 2n} |S_r| < c \right\}$$

особено полезен е т. нар.

Принцип на отражението. Броят на всички траектории, излизащи от точката $(0,0)$ и завършващи в точката $(2n, 0)$, които пресичат или допират правата $S_r = c$, е равен на броя на всички траектории, които водят от $(0,0)$ в $(2n, 2c)$, и е равен на $\binom{2n}{n-c}$.

Доказателството на принципа на отражението е просто, като се използва точката на първо допиране на траекторията (r, S_r) с правата $S_r = c$ и нататък се замени траекторията (r, S_r) с огледалния ѝ образ спрямо правата $S_r = c$, който ще води винаги в точката $(2n, 2c)$, ако оригиналът завършва пътя си в $(2n, 0)$ (вж. фиг. 24).

Неравенствата (1) и (2), дефиниращи $\Phi_n(z)$ и $\Phi_n^+(z)$ изглеждат, с оглед на (3), така:

$$(4) \quad \Phi_n^+(z) = \mathbf{P}\left\{\sqrt{\frac{n}{2}} D_{n,n}^+ < z\right\} = \mathbf{P}\left\{n D_{n,n}^+ < z\sqrt{2n}\right\} = \mathbf{P}\left\{\sup_{0 \leqq r \leqq 2n} S_r < c\right\},$$

където $c =]z\sqrt{2n}[$, и

$$(5) \quad \Phi_n(z) = \mathbf{P}\left\{\sqrt{\frac{n}{2}} D_{n,n} < z\right\} = \mathbf{P}\left\{n D_{n,n} < z\sqrt{2n}\right\} = \mathbf{P}\left\{\sup_{0 \leqq r \leqq 2n} |S_r| < c\right\}.$$

Равенство (4) означава, че $\Phi_n^+(z)$ е вероятността траекторията (r, S_r) на случайното лутане да бъде под правата $S_r = c$. Но такива траектории са на брой $\binom{2n}{n} - \binom{2n}{n-c}$ — разликата от всички възможни траектории, водещи от $(0,0)$ до $(2n, 0)$, и тези от тях, които се допират или пресичат правата $S_r = c$. Следователно

$$\Phi_n^+(z) = \frac{\binom{2n}{n} - \binom{2n}{n-c}}{\binom{2n}{n}} = 1 - \frac{\binom{2n}{n-c}}{\binom{2n}{n}},$$

където $c =]z\sqrt{2n}[$. Ако $z \leqq 0$, то и $c \leqq 0$, откъдето $\Phi_n^+(z) = 0$ (винаги е изпълнено $n \geqq \sup\limits_r S_r = nD_{n,n}^+ \geqq 0$). Събитието $\sup\limits_r S_r = n$ е изпълнено, когато получим вариационен ред (ζ), първите n члена на който са наблюденията (ξ), а следващите n — наблюденията (η). Освен това $\Phi_n^+(z) = 1$, ако $]z\sqrt{2n}[\geqq n$, т. е. ако $z > \sqrt{\dfrac{n}{2}}$.

Равенството (5) означава, че $\Phi_n(z)$ е вероятността траекторията (r, S_r) на случайното лутане да бъде винаги заключена строго между правите $S_r = -c$ и $S_r = c$.

Нека M е множеството от всички траектории от $(0,0)$ до $(2n, 0)$. Техният брой, както видяхме, е $\binom{2n}{n}$. Нека N е множеството на траекториите, заключени строго между правите $\alpha : S_r = -c$ и $\beta : S_r = c$. Ще определим броя $\#(N)$ на различните елементи в множеството N.

Нека:

U_1 е множеството от траекториите, достигащи правата α, но недостигащи правата β;

V_1 — множеството от траекториите, достигащи правата β, но недостигащи правата α;

U_2 — множеството от траекториите, достигащи първо α , после β и повече недостигащи α;

V_2 — множеството от траекториите, достигащи първо β, после α и повече недостигащи β;

Въобще нека:

U_{2i} е множеството от траекториите на M, при които точно i пъти се достигат първо α, а после β, после отново α, β и т. н., като след последното i-то пресичане на двойката $(\alpha\beta)$ повече не достигат правата α;

V_{2i} — множеството от траекториите на M, при които точно i пъти се достигат първо β, а после α, после отново β, α и т. н., като след последното i-то пресичане на двойката $(\beta\alpha)$ повече не достигат правата β;

U_{2i+1} — множеството от траекториите на M, при които точно i пъти се достигат първо α, а после β, после отново α, β и т. н., като след последното $(i+1)$-во достигане на α повече не достигат правата β;

V_{2i+1} — множеството от траекториите на M, при които точно i пъти се достигат първо β, а после α, после отново β, α и т. н.

които след последното $(i+1)$-во достигане на β повече не достигат правата α.

Така множеството M се представя като обединение на непресичащите се множества N, U_i, V_i, т. е.

$$M = N + \sum_{i \geqq 1} (U_i + V_i). \tag{6}$$

Очевидно е, че от известно място нататък множествата U_i, V_i ще бъдат празни.

Да въведем нови множества A_i, B_i по следния начин:

A_1 е множеството от траекториите на M, достигащи правата α поне веднъж;

B_1 — множеството от траекториите на M, достигащи правата β поне веднъж;

A_2 — множеството от траекториите на M, достигащи поне веднъж правите α и β в посочения ред;

B_2 — множеството от траекториите на M, достигащи поне веднъж правите β и α в посочения ред;

A_i — множеството от траекториите на M, достигащи първо α, после β, после отново α и т. н. алтернативно поне веднъж до i-тото поредно достигане;

B_i — множеството от траекториите на M, достигащи първо β, после α, после отново β, α и т. н. алтернативно поне веднъж до i-тото поредно достигане.

Изпълнени са

$$A_1 = U_1 + \sum_{j=2}^{n} (U_j + V_j), \quad B_1 = V_1 + \sum_{j=2}^{n} (U_j + V_j),$$

$$A_i = U_i + \sum_{j=i+1}^{n} (U_j + V_j), \quad B_i = V_i + \sum_{j=i+1}^{n} (U_j + V_j).$$

Следователно, за всяко $i \geqq 1$ са верни равенствата

$$\begin{aligned}(A_{2i-1} - A_{2i}) + (B_{2i-1} - B_{2i}) &= (U_{2i-1} + V_{2i}) + (V_{2i-1} + U_{2i}) \\ &= (U_{2i-1} + V_{2i-1}) + (U_{2i} + V_{2i}),\end{aligned}$$

откъдето намираме, че

$$\sum_{i=1}^{n} (U_i + V_i) = \sum_{i=1}^{n} [(A_{2i-1} - A_{2i}) + (B_{2i-1} - B_{2i})].$$

Оттук и от равенство (6) стигаме до представянето

$$N = M - \sum_{i=1}^{n} [(A_{2i-1} - A_{2i}) + (B_{2i-1} - B_{2i})],$$

от което следва

$$\#(N) = \#(M) - \sum_{i \geqq 1} [\#(A_{2i-1}) - \#(A_{2i}) + \#(B_{2i-1}) - \#(B_{2i})]. \tag{7}$$

Обаче всяка траектория от A_i достига поне i-пъти правите $\alpha : S_r = c$ и $\beta : S_r = -c$. Нейният огледален образ спрямо правата α след първото ѝ достигане ще стигне до правата $\alpha_2 : S_r = 3c$ при първото достигане на β. По-нататък успоредно пренесеният остатък от траекторията на S_r, до следващото ѝ достигане на правата α ще стигне до правата $\alpha_3 : S_r = 5c$, след което отново образуваният огледален образ на остатъка от разглежданата траектория до следващото ѝ достигане на правата β ще стигне до правата $\alpha_4 : S_r = 7c$ и т. н., след последното i-то достигане на някоя от правите α или β на разглежданата траектория, конструираната съпътстваща (и помощна) траектория, пренесена успоредно на остатъка от разглежданата, ще достигне от $(x, (2i+1)\,c)$ до точката $(2n, 2ic)$, докато истинската траектория от A_i стигне до точката $(2n, 0)$. Така, по този обобщен принцип на отражението откриваме, че всяка от траекториите на A_i е еквивалентна на някоя от траекториите, водещи от точката $(0,0)$ до точката $(2n, 2ic)$. Аналогично, всяка траектория от B_i е еквивалентна на траектория, водеща от точката$(0,0)$ до точката $(2n, -2ic)$. Следователно

$$\#(A_i) = \#(B_i) = \binom{2n}{n - ic}.$$

Тогава от (7) получаваме, че

$$\begin{aligned} \#(N) &= \binom{2n}{n} - 2 \sum_{i \geqq 1} \left[\binom{2n}{n - (2i-1)c} - \binom{2n}{n - 2ic} \right] \\ &= \binom{2n}{n} - 2 \sum_{i=1}^{[n/c]} (-1)^{i-1} \binom{2n}{n - ic}, \end{aligned}$$

което се твърди в теоремата, тъй като $\Phi_n(z) = \dfrac{\#(N)}{\binom{2n}{n}}$.

§ 8. Доказателство на теоремата на Смирнов

След като имаме резултата от теоремата на Гнеденко – Королюк, сме в състояние да докажем теоремата на Смирнов за случая $n = m$.

Теорема 1. *Нека са изпълнени условията на теоремата на Гнеденко – Королюк. Тогава:*

а) $\lim_{n\to\infty} \Phi_n^+(z) = 1 - e^{-2z^2}$ *за* $z > 0$;

б) $\lim_{n\to\infty} \Phi_n(z) = K(z) = 1 - 2 \sum_{k=1}^{\infty} (-1)^{k-1} e^{-2k^2z^2}$ *за* $z > 0$.

Доказателство. Разглеждаме събираемото

$$I_{k,n}(c) = \frac{\binom{2n}{n-kc}}{\binom{2n}{n}} = \frac{(n!)^2}{(n-kc)!(n+kc)!}, \quad c =]z\sqrt{2n}[,$$

в израза за $\Phi_n(z)$. По формулата на Стирлинг

$$n! \sim \sqrt{2\pi n}\, e^{-n} n^n (1 + o(1))$$

получаваме

$$I_{k,n}(c) = \frac{2\pi n e^{-2n} n^{2n} (1 + o(1))}{\sqrt{2\pi(n-kc)}\, e^{-n+kc} \sqrt{2\pi(n+kc)}\, e^{-n-kc} n^{2n}}$$

$$= \frac{1}{\sqrt{1 - \frac{k^2c^2}{n^2}}} \cdot \frac{1}{\left(1 - \frac{k^2c^2}{n^2}\right)^n} \cdot \left(\frac{n-kc}{n+kc}\right)^{kc} \cdot (1 + o(1)).$$

Като вземем предвид, че k и z са фиксирани, а $c =]z\sqrt{2n}[\to \infty$ заедно с n, за отделните множители ще получим

$$\sqrt{1 - \frac{k^2c^2}{n^2}} \sim \sqrt{1 - \frac{k^2z^22n}{n^2}} \longrightarrow 1 \quad \text{при} \quad n \to \infty,$$

$$\left(1 - \frac{k^2c^2}{n^2}\right)^n \sim \left(1 - \frac{k^2z^22n}{n^2}\right)^n \longrightarrow e^{-2k^2z^2} \quad \text{при} \quad n \to \infty,$$

$$\frac{(n-kc)^{kc}}{(n+kc)^{kc}} = \frac{\left(1 - \frac{kc}{n}\right)^{kc}}{\left(1 + \frac{kc}{n}\right)^{kc}} \sim \frac{\left[\left(1 - \frac{kz\sqrt{2}}{\sqrt{n}}\right)^{\sqrt{n}}\right]^{kz\sqrt{2}}}{\left[\left(1 + \frac{kz\sqrt{2}}{\sqrt{n}}\right)^{\sqrt{n}}\right]^{kz\sqrt{2}}}$$

$$\longrightarrow \frac{\left(e^{-kz\sqrt{2}}\right)^{kz\sqrt{2}}}{\left(e^{kz\sqrt{2}}\right)^{kz\sqrt{2}}} = e^{-4k^2z^2} \quad \text{при} \quad n \to \infty.$$

Следователно изпълнено е

$$\lim_{n\to\infty} I_{k,n}(c) = \frac{e^{-4k^2z^2}}{e^{-2k^2z^2}} = e^{-2k^2z^2}, \quad k = 1, 2, \ldots \tag{1}$$

Ще покажем, че за всяко $\varepsilon > 0$ и $z > 0$ при всички достатъчно големи n е изпълнено

$$\delta_n = \left| \mathbf{P}\left\{ \sqrt{\frac{n}{2}}\, D_{n,n} < z \right\} - K(z) \right| < \varepsilon, \tag{2}$$

където първият член е функцията $\Phi_n(z)$, определена в условието на теоремата на Гнеденко – Королюк от § 7, а $K(z)$ е функцията на разпределение на Колмогоров от условие б) на теоремата на Смирнов.

При дадено $\varepsilon > 0$ съществува такова N, че $\exp(-2N^2z^2) < \varepsilon$ за всяко фиксирано $z > 0$. Да фиксираме N. Съгласно (1) съществува такова n_0, че при всички $n > n_0$ е изпълнено

$$\delta_{1,n} = \left| 2\sum_{k=1}^{N} (-1)^{k-1} I_{k,n}(c) - 2\sum_{k=1}^{N} (-1)^{k-1} e^{-2k^2z^2} \right| < \varepsilon.$$

Очевидно (2) може да се представи във вида

$$\delta_n \leqq \delta_{1,n} + \delta_{2,n}, \tag{3}$$

където

$$\begin{aligned}
\delta_{2,n} &= 2\left| \sum_{N<k\leqq[n/2]} (-1)^{k-1} I_{k,n}(c) - \sum_{k>N} (-1)^{k-1} e^{-2k^2z^2} \right| \\
&\leqq 2\left| \sum_{N<k\leqq[n/2]} (-1)^{k-1} I_{k,n}(c) \right| + 2\sum_{k>N} e^{-2k^2z^2} \\
&= S_{1,n} + S_{2,n}.
\end{aligned}$$

Първата сума $S_{1,n}$ в последния израз може да се оцени след замяна на всички събираеми с максималния първи член и отчитане на техния брой $(n/2 - N)$ с израза

$$S_{1,n} < 4\,\frac{\binom{2n}{n-Nc}}{\binom{2n}{n}} = 4e^{-2N^2z^2}(1 + o(1)) < 5\varepsilon. \tag{4}$$

Втората сума $S_{2,n}$ се оценява аналогично, като се изнесе най-големият ѝ член $\exp(-2N^2z^2)$ пред скоби и се отчете, че сумата на оставащите членове в безкрайния ред се мажорира от геометрична прогресия с частно $\frac{1}{2}$. Така

$$S_{2,n} < 4e^{-2N^2z^2} < 4\varepsilon. \tag{5}$$

От (3) – (5) стигаме до извода, че за всяко $\varepsilon > 0$ и за всяко $z > 0$ при достатъчно големи $n\,(n > n_0)$ е изпълнено $\delta_n < 10\varepsilon$.

Доказаните теореми ни убеждават, че разпределенията на Смирнов и Колмогоров, които се използват за проверка на хипотезите за съвпадане на функциите на разпределение в две извадки, не зависят от вида на разпределението им. Единственото изискване то да е непрекъснато може да следва от логически и чисто физически съображения относно характера на изследваното явление. Затова критериите на Смирнов и на Колмогоров се отнасят към непараметрични критерии за проверка на хипотези. Такива критерии играят важна роля в приложната статистика.

§ 9. χ^2–критерий за съгласие

Нека $\xi_1, \ldots, \xi_n$ е проста извадка от наблюдения над случайната величина ξ. Искаме да узнаем разумно ли е да разглеждаме случайната величина ξ като имаща някакъв конкретен закон на разпределение.

Издигаме хипотезата H_0, състояща се в това, че опитните данни са стойности на случайната величина ξ, имаща функция на разпределение $F(x)$. Тук се предполага, че $F(x)$ е напълно определена, така че аналитичният ѝ вид не съдържа вече никакви неизвестни параметри и стойността на $F(x)$ може да бъде изчислена за всяко x. Трябва да изберем метод за проверка на това, може ли да се счита, че опитните данни са съгласувани с хипотезата H_0. Един от най-често използваните за тези условия е χ^2-критерият за съгласие на Пирсън. Да разгледаме този критерий. Както при построяването на хистограми, разделяме цялата област на изменение на случайната величина ξ на l интервала $\Delta_1, \ldots, \Delta_l$ и пресмятаме броя $\widehat{m}_j$, $j = 1, 2, \ldots, l$ на величините ξ_i, $i = 1, \ldots, n$, принадлежащи на Δ_j. Предполагайки, че теоретичният закон $F(x)$ е известен, можем да определим вероятността

p_j случайната величина ξ да попадне в интервала Δ_j:

$$p_j = \mathrm{P}\{\xi_i \in \Delta_j\}, \quad j = 1, \ldots, l.$$

Тогава теоретичният брой на стойностите на случайната величина ξ, попаднали в интервала Δ_j, може да се пресметне по формулата $np_j = m_j$.

Резултатите от приведените пресмятания обикновено се дават в таблица от вида

Интервали	Δ_1	...	Δ_j	...	Δ_l
Емпирична честота	$\widehat{m}_1$	...	$\widehat{m}_j$	...	$\widehat{m}_l$
Теоретична честота	np_1	...	np_j	...	np_l

Тук $\widehat{m}_1 + \cdots + \widehat{m}_l = n$ и $p_1 + \cdots + p_l = 1$.

Ако емпиричните честоти силно се отличават от теоретичните, то проверяваната хипотеза H_0 следва да се отхвърли, в противен случай — да се приеме.

Ще формулираме критерия, който би характеризирал степента на близост между емпиричните и теоретичните честоти:

Ако проверяваната хипотеза H_0 е вярна, то като мярка за близост между $\widehat{m}_j$ и np_j за $j = 1, \ldots, l$ се използва критерият

$$\widehat{\chi}^2 = \sum_{j=1}^{l} \frac{(\widehat{m}_j - np_j)^2}{np_j}.$$

При вярна хипотеза H_0 величината $\widehat{\chi}^2$ има асимптотично разпределение както $\chi^2(k)$ с $k = l - r - 1$ степени на свобода. Тук r е броят на параметрите на $F(x)$, пресметнати с помощта на извадката. Правилото за прилагане на χ^2-критерия се свежда към следното: По зададено ниво на съгласие α от таблицата на разпределенията на $\chi^2(k)$ се определя квантилът χ^2_α така, че

$$P(\chi^2(k) > \chi^2_\alpha) = \int_{\chi^2_\alpha}^{\infty} f_{\chi^2(k)}(x)dx = \alpha.$$

Ако $\widehat{\chi}^2 < \chi^2_\alpha$, хипотезата H_0 се приема, а ако $\widehat{\chi}^2 > \chi^2_\alpha$, хипотезата H_0 се отхвърля.

Да отбележим, че χ^2–критерият е незаменим в случаите, когато директно получаваме емпиричните честоти $\widehat{m}_j$, $j = 1, 2, \ldots, l$, без да знаем точните наблюдения над случайната величина ξ.

Пример. В телеграфна станция са направени наблюдения за броя ξ на неправилните свързвания в минута. В продължение на един час са получени следните резултати: 3, 1, 3, 1, 4, 2, 2, 4, 0, 3, 0, 2, 2, 0, 2, 1, 4, 3, 3, 1, 4, 2, 2, 1, 1, 2, 1, 0, 3, 4, 1, 3, 2, 7, 2, 0, 0, 1, 3, 3, 1, 2, 4, 2, 0, 2, 3, 1, 7, 5, 1, 1, 0, 1, 1, 2, 2, 1, 1, 5. Да се намерят $\overline{\xi}_n$ и σ_n^2 на неправилните свързвания в минута. Проверете изпълнено ли е основното условие за разпределението на Поасон $\mathbf{E}\xi = \mathbf{D}\xi$. Намерете теоретичното разпределение на Поасон, съответно на дадената извадка, и проверете степента на съгласие на теоретичното и емпиричното разпределение по χ^2-критерия при ниво на съгласие $\alpha = 0,05$.

Решение. Съставяме таблицата

Δ_j	0	1	2	3	4	5	6	7
$\widehat{m}_j$	8	17	16	10	6	2	0	1

Тогава

$$\mathbf{E}\xi \approx \overline{\xi}_n = \frac{1}{60}\sum_{j=0}^{7}\xi_j\widehat{m}_j = 2,$$

$$\mathbf{D}\xi \approx \sigma_n^2 = \frac{1}{60}\sum_{j=0}^{7}\widehat{m}_j(\xi_j - \overline{\xi}_n)^2 \approx 2,1.$$

Необходимото условие за разпределението на Поасон практически е изпълнено с $\lambda = 2$. Теоретичният закон на Поасон е

$$p_j = \mathbf{P}\{\xi = j\} = \frac{2^j}{j!}e^{-2},$$

където сме приели, че $\lambda = \mathbf{E}\xi = \overline{\xi}_n = 2$. Тогава намираме

$$P(0) = \frac{2^0}{0!}e^{-2} = 0,1353 = p_0; \quad P(1) = \frac{2}{1!}e^{-2} = 0,2707 = p_1;$$

$$P(2) = \frac{2^2}{2!}e^{-2} = 0,2707 = p_2; \quad P(3) = \frac{2^3}{3!}e^{-2} = 0,1804 = p_3;$$

$$P(4) = \frac{2^4}{4!}e^{-2} = 0,0402 = p_4; \quad P(5) = \frac{2^5}{5!}e^{-2} = 0,0361 = p_5;$$

$$P(6) = \frac{2^6}{6!}e^{-2} = 0,0120 = p_6; \quad P(7) = \frac{2^7}{7!}e^{-2} = 0,0034 = p_7.$$

Получените резултати записваме в следната таблица:

j	0	1	2	3	4	5	6	7
ξ_j	0	1	2	3	4	5	6	7
$\widehat{m}_j$	8	17	16	10	6	2	0	1
np_j	8,1	16,24	16,24	10,82	5,41	2,166	0,72	0,204

Обединяваме петия, шестия и седмия стълб поради това, че във всеки от тях имаме малко наблюдения. Тогава получаваме

j	0	1	2	3	4	5	6	7
ξ_j	0	1	2	3	4	5 и повече		
$\widehat{m}_j$	8	17	16	10	6	3		
np_j	8,1	16,24	16,24	10,82	5,41	3,09		

Изчисляваме

$$\widehat{\chi}^2 = \frac{(8-8,1)^2}{8} + \frac{(17-16,24)^2}{16,24} + \frac{(16-16,24)^2}{16,24}$$
$$+ \frac{(10-10,82)^2}{10,82} + \frac{(6-5,41)^2}{5,41} + \frac{(3-3,09)^2}{3,09} \approx 0,2.$$

От това, че $l = 6$, $r = 1$, то $k = l - r - 1 = 4$ степени на свобода. Тогава от таблицата за χ^2 при $\alpha = 0,05$ и $k = 4$ намираме $\chi^2_\alpha = 9,5$. **Тъй като $9,5 > 0,2$, нямаме основание да отхвърлим** проверяваната хипотеза.

Задачи

1. Намерете емпиричните характеристики $\overline{\xi}_n$ и σ_n^2 по членовете на вариационния ред и техните честоти:

а)

ξ_i	10,3	10,5	10,7	10,9	11,1	11,3	11,5	11,7	11,9	12,1
n_i	4	7	8	10	25	15	12	10	4	5

б)

ξ_i	83	85	87	89	91	93	95	97	99	101
n_i	6	7	12	15	30	10	8	0	4	2

Отг. а) $\overline{\xi}_n = 11,19$, $\sigma_n^2 = 0,19$; б) $\overline{\xi}_n = 90,72$, $\sigma_n^2 = 17,20$.

2. Определете $F_n(x)$ и постройте графиката ѝ по данните:

ξ_i	2	6	10
n_i	12	18	30

$$\text{Отг. } F_n(x) = \begin{cases} 0 & \text{при} \quad x \leqq 2, \\ 0,2 & \text{при} \quad 2 < x \leqq 6, \\ 0,5 & \text{при} \quad 6 < x \leqq 10, \\ 1 & \text{при} \quad x > 10. \end{cases}$$

3. Постройте полигона на честотите на разпределението

ξ_i	1	3	5	7	9
n_i	10	15	30	33	12

4. Постройте хистограмата по таблицата

ξ_i	$0 \div 1,5$	$1,5 \div 3$	$3 \div 4,5$	$4,5 \div 6$
n_i	0,3	0,2	0,4	0,1

5. Покажете, че ако $(\xi_1, \ldots, \xi_n)$ има n-мерно разпределение на Дирихле $\mathfrak{D}(\lambda_1, \ldots, \lambda_k; \lambda_{k+1})$, то сумата $\eta = \xi_1 + \cdots + \xi_k$ има бета-разпределение $\mathrm{B}(\lambda_1 + \ldots + \lambda_k; \lambda_{k+1})$.

6. Покажете, че ако $(\xi_1, \ldots, \xi_k) \in \mathfrak{D}(\lambda_1, \ldots, \lambda_k; \lambda_{k+1})$, то за всяко $r < k$ е изпълнено $(\xi_1, \ldots, \xi_r) \in \mathfrak{D}(\lambda_1, \ldots, \lambda_r; \lambda_{r+1} + \cdots + \lambda_{k+1})$.

7. Покажете, че ако $(\xi_1, \ldots, \xi_k) \in \mathfrak{D}^*(\lambda_1, \ldots, \lambda_k; \lambda_{k+1})$, то при всеки избор на числата $r_1 > 0, \ldots, r_s > 0$, $r_1 + \ldots + r_s \leqq k$ е изпълнено

$$(\xi_{r_1}, \xi_{r_1+r_2}, \ldots, \xi_{r_1+\cdots+r_s}) \in \mathfrak{D}^*(\lambda_{(1)}, \ldots, \lambda_{(s)}; \lambda_{(s+1)}),$$

където

$$\lambda_{(1)} = \lambda_1 + \cdots + \lambda_{r_1},$$

$$\ldots\ldots\ldots\ldots\ldots$$

$$\lambda_{(s)} = \lambda_{r_1+\cdots+r_{s-1}+1} + \cdots + \lambda_{r_1+\cdots r_s},$$

$$\lambda_{(s+1)} = \lambda_{r_1+\cdots+r_s+1} + \cdots + \lambda_{k+1}.$$

8. Покажете, че вероятностните плътности на разпределение на първия и на последния член на вариационен ред от наблюдения над непрекъснатата случайна величина ξ с функция на разпределение $F(x)$ са съответно

$$G'_{n1}(x) = nf(x)[1 - F(x)]^{n-1}, \quad G'_{nn}(x) = nf(x)[F(x)]^{n-1}.$$

9. Нека при $p \in (0,1)$ броят ν_p на компонентите на извадката $\xi_1, \dots, \xi_n$, по-малки от p-квантила x_p, е асимптотично нормален $\mathcal{N}(np, np(p-1))$ при $n \to \infty$. Покажете въз основа на този факт, че ако t_γ е решение на уравнението $\Phi(t) = \dfrac{1+\gamma}{2}$, то при големи n наредените статистики $\xi_{([np_\gamma])}$, $\xi_{([n\overline{p}_\gamma])}$, където p_γ и $\overline{p}_\gamma$ са решенията на уравнението

$$\frac{(\nu_p - np)^2}{[np(1-p)]} = t_\gamma^2,$$

образуват γ-доверителен интервал за x_p.

10. Ръдърфорд и Гайгер в течение на 2608 периода, всеки с продължителност 7,5 s, са отчитали броя на α-частиците, излъчени от радиоактивен обект. Броят ν_i на периодите, в течение на които са наблюдавани n_i α-частици, са приведени в таблицата:

n_i	0	1	2	3	4	5	6	7	8	9	10	11	12
ν_i	57	203	383	525	532	408	273	139	45	27	10	4	2

Проверете с помощта на χ^2-критерия хипотезата, че броят на α-частиците, излъчени за интервала от време с продължителност 7,5 s, се подчинява на закона на Поасон с параметър λ, за статистическата оценка на който се взема средната на извадката, при ниво на съгласие $\varepsilon = 0,05$.

Отг. $\overline{\xi}_n = 3,87$, $\widehat{\chi_n^2} = 12,9$, $\widehat{\chi_\varepsilon^2} = 18,3$. Хипотезата се приема.

11. В една от работите на В. Борткевич се дава пример за прилагане на закона на Поасон, като се изучава броят на смъртните случаи на войници от старата пруска армия, ударени от копитата на конете. Този брой е даден в таблицата:

Брой на случаите в корпусите	Брой на корпусите
0	109
1	65
2	22
3	3
4	1

Намерете теоретичното разпределение на Поасон и проверете степента на съгласие между теоретичното и емпиричното разпределение по χ^2-критерия на Пирсън с ниво на съгласие $\alpha = 0,05$.

Отг. $\widehat{\chi}^2 = 0,32, \quad \chi^2_\alpha(3) = 7,8.$

12. В таблицата, дадена по-долу, са приведени данни за наблюденията над случайната величина ξ. Проверете хипотезата, че случайната величина ξ има нормално разпределение при ниво на съгласие $\alpha = 0,05$.

Δ_j	$3,0 \div 3,6$	$3,6 \div 4,2$	$4,2 \div 4,8$	$4,8 \div 5,4$	$5,4 \div 6,0$	$6,0 \div 6,6$	$6,6 \div 7,2$
$\widehat{m}_j$	2	8	35	43	22	15	5

Отг. $\widehat{\chi}^2 = 6,22, \quad \chi^2_\alpha(5) = 9,5.$

13. При изследване чувствителността на един от каналите на 40 телевизора са получени записаните в следващата таблица резултати в микроволтове:

Интервали	Честоти	Интервали	Честоти
75-125	0	425-475	2
125-175	1	475-525	2
175-225	5	525-575	0
225-275	9	575-625	0
275-325	6	625-675	1
325-375	8	675-725	0
375-425	6		

С помощта на χ^2-критерия и критерия на Колмогоров проверете хипотезата, че разпределението на чувствителността е нормално с параметри $\overline{\xi}_n$ и σ^2_n при ниво на съгласие $\alpha = 0,05$.

Отг. $\overline{\xi}_n = 349, \sigma^2_n = 106;$

а) $\widehat{\chi}^2 = 5,49, \quad \chi^2_\alpha(10) = 11,1$. Хипотезата се приема.

б) $D_n = 0,234, \quad \sqrt{n}D_n = 1,48, \quad Z_\alpha = 1,358$. Хипотезата се отхвърля.

14. Имаме две групи по 60 детайла от един и същ вид, произведени на два струга. Резултатите от измерванията са дадени в таблицата:

№ i	I група	II група	№ i	I група	II група	№ i	I група	II група
1	72,58	72,50	21	72,50	72,35	41	72,30	72,31
2	72,35	72,35	22	72,69	72,16	42	72,28	72,46
3	72,33	72,69	23	72,54	72,51	43	72,51	72,36
4	72,54	72,60	24	72,48	72,50	44	72,37	72,39
5	72,24	72,54	25	72,36	72,50	45	72,14	72,30
6	72,42	72,42	26	72,50	72,48	46	72,42	72,30
7	72,58	72,68	27	72,43	72,53	47	72,36	72,38
8	72,47	72,54	28	72,46	72,25	48	72,28	72,55
9	72,54	72,55	29	72,56	72,48	49	72,20	72,36
10	72,24	72,33	30	72,48	72,36	50	72,48	72,24
11	72,38	72,56	31	72,43	72,53	51	72,66	72,23
12	72,70	72,36	32	72,56	72,23	52	72,64	72,16
13	72,47	72,36	33	72,34	72,55	53	72,73	72,17
14	72,49	72,15	34	72,38	72,51	54	72,43	72,37
15	72,28	72,48	35	72,56	72,25	55	72,28	72,38
16	72,47	72,46	36	72,32	72,11	56	72,64	72,46
17	72,95	72,36	37	72,41	72,44	57	72,72	72,12
18	72,18	72,38	38	72,14	72,51	58	72,35	72,28
19	72,66	72,40	39	72,29	72,55	59	72,60	72,23
20	72,35	72,38	40	72,31	72,24	60	72,46	72,38

Проверете, използвайки критерия на Колмогоров, хипотезата, че двете извадки принадлежат на една и съща генерална съвкупност с ниво на съгласие $\alpha = 0,05$.

Тринадесета глава

Непараметрични критерии

В тази глава ще бъдат разгледани група критерии за вземане на решения относно свойствата на наблюденията в извадката или на наблюденията в няколко различни извадки, при което не се правят никакви предположения относно разпределението на наблюдаваната случайна величина. Непараметричните критерии служат за установяване на някои непараметрични факти — свойства на извадките или на техни трансформации, които не зависят от разпределенията на наблюдаваните величини. Получените въз основа на тези критерии изводи най-често имат спомагателен характер и не са крайна цел на статистическото изследване. Те са начален или междинен етап в поредицата от методи за извличане на полезна информация от проведените наблюдения. Ние се спираме на най-популярните критерии от тази група. В повечето случаи строгите теоретични обосновки на критериите са изпуснати и са приведени само интуитивните им мотиви.

§ 1. Изключване на груби грешки от извадката

Нека са направени n независими наблюдения $\xi_1, \ldots, \xi_n$ и нека ξ^* е онова наблюдение, което се отличава силно от останалите. Във всички случаи ξ^* е или минималният, или максималният член на общия вариационен ред на извадката.

Нека $\overline{\xi}_n = \dfrac{1}{n}(\xi_1 + \cdots + \xi_n)$. За пълнота ще разгледаме два случая:

а) $\mathbf{D}\xi = \sigma^2$ е известна; б) $\mathbf{D}\xi$ е неизвестна.

Формално случаят а) следва да бъде отнесен към параметричните случаи.

Случай а). Величината

$$t^* = \frac{\xi^* - \overline{\xi}_n}{\sigma\sqrt{n+1}}\sqrt{n}$$

би трябвало да има (поне приблизително при големи n) стандартно нормално разпределение, когато ξ_n^* е наблюдение от същата извадка (хипотеза H_0), т. е. когато не е груба грешка. Нека t е абсолютната стойност на t^*. При вярна хипотеза H_0 вероятността на събитието $\{|t^*| \geqq z\}$ е $2 - 2\Phi(z)$. Ако пресметнатата по този начин вероятност се окаже твърде малка, счита се, че стойността ξ^* съдържа груба грешка и трябва да се изключи от по-нататъшна обработка на резултатите.

Пример 1. Извършени са 41 независими наблюдения над величината ξ с дисперсия $\sigma^2 = (0,133)^2$. Едно от измерванията ξ^* е $6,866$ и силно се отличава от останалите числа. Може ли да се счита, че наблюдението ξ^* съдържа груба грешка и да се отстрани от по-нататъшните обработки на наблюденията, ако средното аритметично на останалите 40 наблюдения е $\overline{\xi}_n = 6,5$?

Решение. Пресмятаме

$$t^* = \frac{6,866 - 6,500}{0,133\sqrt{41}}\sqrt{40} = 2,72.$$

От таблицата на нормалната функция на разпределение намираме за вероятността $p = 2 - 2\Phi(2,72) = 0,0036$. Това е твърде малка вероятност в случай, че ξ^* не съдържа груба грешка. Следователно с вероятност $1 - p = 0,9964$ можем да считаме, че ξ^* е груба грешка и можем да го изключим от по-нататъшна обработка.

Случай б). Нека $s^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(\xi_i - \overline{\xi}_n)^2$. При големи n величината

$$t^* = \frac{|\xi^* - \overline{\xi}_n|}{s}\sqrt{\frac{n-1}{2}} \tag{1}$$

би трябвало да бъде разпределена приблизително по закона на Стюдънт с $n-1$ степени на свобода, когато е вярна хипотезата H_0, че ξ^* не съдържа груба грешка и също е наблюдение над ξ.

От таблицата за разпределението на Стюдънт за даденото n се пресмята квантилът t_α за определено, предварително избрано α (обикновено $\alpha = 0,9; 0,95$) и се сравняват t^* и t_α. След това се постъпва така: При $t^* \geqq t_\alpha$ може да се счита с вероятност за извода, по-голяма от α, че ξ^* съдържа груба грешка и трябва да се изключи от по-нататъшни пресмятания. При $t^* < t_\alpha$ това не означава все още, че няма груба грешка. Просто няма достатъчно основания за изключване на силно отличаващото се наблюдение.

З а б е л е ж к а. Приведените критерии работят най-добре, когато наблюдаваната случайна величина ξ има нормално разпределение.

Пример 2. В пример 1 нека $s = 0,133$ е получената в резултат на измерванията оценка на стандартното отклонение σ. Може ли да се счита, че наблюдението $\xi^* = 6,866$ съдържа груба грешка?

Решение. За отношението (1) получаваме

$$t = \frac{0,366}{0,133}\sqrt{\frac{39}{2}} = 12,65.$$

От таблицата за разпределението на Стюдънт с 39 степени на свобода при $\alpha \doteq 0,99$ намираме $t_\alpha = 2,71$. Тъй като очевидно $t^* = 12,75 > 2,71$, то явно ξ^* трябва да се изключи от следващите обработки на статистическите данни.

§ 2. Критерий на знаците

В експерименталните изследвания често трябва да се сравняват два реда статистически наблюдения над обекти, принадлежащи на две различни разновидности, типове, сортове, получени по разни методики, технологии и пр. Математическата постановка на задачата е следната:

Разполагаме с две независими редици от еднакъв брой независими наблюдения:

$$(\xi): \ \xi_1, \xi_2, \ldots, \xi_n;$$
$$(\eta): \ \eta_1, \eta_2, \ldots, \eta_n$$

над случайните величини ξ и η. Да се построи критерий за проверка на хипотезата

H_0 : двете редици (ξ) и (η) са наблюдения над случайни величини с един и същ закон на разпределение

срещу алтернативата

H_1 : наблюдаваните редици (ξ) и (η) са от съвкупности с различни помежду си разпределения.

По този начин може да се отговори на въпроса, дали различията между наблюденията в редиците (ξ) и (η) са значими, съществени, т. е. дали те са свързани с различни качества на двете разглеждани разновидности от обекти, или разликите $d_i = \xi_i - \eta_i$ следва да бъдат отнесени за сметка на случайни отклонения в стойностите на измерваните показатели? Последното твърдение е същността на нулевата хипотеза H_0. Аналогично на случая от критерия на Смирнов тя може да се запише с равенството

$$H_0: \ \mathbf{P}\{\xi_i < x\} = \mathbf{P}\{\eta_i < x\} = F_i(x) \quad \text{за всяко } x.$$

З а б е л е ж к а. Законите на разпределение $F_i(x)$ могат да бъдат различни в различните опити ($i = 1, \ldots, n$), тъй като разликите $d_i = \xi_i - \eta_i$ са симетрично разпределени около нулата, т. е. отклоненията към положителния и отрицателния знак са равновероятни:

$$\mathbf{P}\{d_i \leqq 0\} = \mathbf{P}\{d_i \geqq 0\} = \frac{1}{2}.$$

Обикновено разликите, равни на нула, се изключват от разглеждане. Но при непрекъснати разпределения $F_i(x)$ на наблюденията нито една от разликите d_i не ще е равна на нула.

От редиците (ξ) и (η) дефинираме редицата от n знаци

$$\operatorname{sgn}(\xi - \eta): \ +1, +1, -1, \ldots, +1,$$

в която на i-та позиция стои $+1$, ако $d_i > 0$, или -1, ако $d_i < 0$. Нека ν_n е броят на $+1$ в редицата $\{\operatorname{sgn}(\xi - \eta)\}$. Проверката на нулевата хипотеза H_0 се свежда до проверка на еквивалентната ѝ хипотеза

H_0' : разпределението на ν_n е биномно с параметри n и $p = \dfrac{1}{2}$

срещу алтернативата

$H_1' : p \neq \dfrac{1}{2}$.

Правилата за проверка на такива сложни хипотези са добре известни от трета глава. Критичната област при зададено ниво на съгласие α е двустранна и се определя с число C_α така, че точно или приблизително да е изпълнено равенството

$$\alpha = \mathbf{P}\left\{\left|\nu_n - \frac{n}{2}\right| \geqq C_\alpha\right\} = \sum_{k=0}^{\frac{n}{2}-C_\alpha} \binom{n}{k} \frac{1}{2^n} + \sum_{k=\frac{n}{2}+C_\alpha}^{n} \binom{n}{k} \frac{1}{2^n}.$$

При неголеми n (за $n \leqq 50$) съществуват таблици за $\frac{\alpha}{2}$-квантила на биномното разпределение с параметър $p = \frac{1}{2}$, с чиято помощ по-просто се търси този квантил. При големи n (например $n > 50$) константата C_α може да се определи с използване на нормалното приближение за биномното разпределение

$$\mathbf{P}\left\{\left|\nu_n - \frac{n}{2}\right| \geqq C_\alpha\right\} = \mathbf{P}\left\{\frac{|\nu_n - \frac{n}{2}|}{\frac{1}{2}\sqrt{n}} > \frac{C_\alpha}{\frac{1}{2}\sqrt{n}}\right\} \approx 1 - \Phi\left(\frac{2C_\alpha}{\sqrt{n}}\right) + \Phi\left(-\frac{2C_\alpha}{\sqrt{n}}\right),$$

т. е. от уравнението

$$\Phi\left(\frac{2C_\alpha}{\sqrt{n}}\right) = 1 - \frac{\alpha}{2}.$$

Ако t_α е коренът на уравнението $\Phi(t) = \alpha$, тогава ще имаме $C_\alpha \approx \frac{1}{2}\sqrt{n}\, t_{1-\frac{\alpha}{2}}$. Хипотезата H_0 за еднородност на двете извадки следва да се приеме като непротиворечаща на резултатите от наблюденията, когато за броя ν_n на положителните знаци в редицата $\{\operatorname{sgn}(\xi - \eta)\}$ е изпълнено

$$\frac{n}{2} - \frac{1}{2} t_{1-\frac{\alpha}{2}} \sqrt{n} \leqq \nu_n \leqq \frac{n}{2} + \frac{1}{2} t_{1-\frac{\alpha}{2}} \sqrt{n}.$$

Когато тези неравенства са нарушени, хипотезата H_0 се отхвърля, като вероятността да се допусне грешка от първи род е α.

Пример 1. Извършените 100 паралелни наблюдения над образци от продукцията на два цеха за електрически лампи с мощност 60 W показали, че в 36 от случаите лампите от първия цех светят по-дълго от тези, произведени във втория цех. При ниво на съгласие $\alpha = 0,05$ може ли да се твърди, че двата цеха произвеждат продукция с еднакво качество?

Решение. Използваме нормалното приближение на споменатия критерий. Решението на уравнението $\Phi(t) = 1 - \frac{0,05}{2}$ е $t = 1,96$. Областта на приемане на хипотезата H_0 е

$$50 - 0,98\sqrt{100} \leqq \nu_n \leqq 50 + 0,98\sqrt{n},$$

т. е. $40 \leqq \nu_n \leqq 60$. Тъй като в случая $\nu_n = 36$ е вън от тази област, заключаваме, че продукцията на двата цеха не е еднородна.

Нещо повече, можем да твърдим, че продукцията на втория цех е с по-добри качествени показатели от гледна точка на срока на използване на лампите.

Критерият на знаците е широко разпространен благодарение на това, че процедурата на прилагането му е изключително проста. Измерванията на сравняемия показател могат да се провеждат с груби средства, достатъчно е да се определи в кой от двата обекта този показател има превес. Трябва да се има предвид обаче, че при малки n критерият на знаците ще отхвърля нулевата хипотеза само в случаите, когато p е близо до 0 или до 1. Затова брой на наблюденията 10 – 15 може да се използва само за предварителни и твърде груби съждения. Когато стойностите на конкуриращи се хипотези с $p = \frac{1}{2}$ са близки до $\frac{1}{2}$, това различие може да се долови само при големи n.

§ 3. Критерий на Уилкоксън

Критерият на знаците е неприложим, когато броят на наблюденията в двете извадки е различен. Тогава проверката за принадлежност на извадките към една и съща генерална съвкупност се извършва по критерия на Уилкоксън.

Задачата е почти същата както при наблюдения

$$(\xi): \ \xi_1, \dots, \xi_n;$$

$$(\eta): \ \eta_1, \dots, \eta_m,$$

за които предполагаме, че са извадки от случайни величини с непрекъснати функции на разпределение $F(x)$ и $G(x)$ (за да няма съвпадащи стойности между тях). Образуваме общия вариационен ред на двете редици от наблюдения

$$\eta_{i_1} < \xi_{j_1} < \xi_{j_2} < \eta_{i_2} \dots < \eta_{i_m} < \xi_{j_n}.$$

Нека величината $U_{m,n}$ е общият брой инверсии в този вариационен ред. Казваме, че наблюдението ξ_i образува k инверсии, ако в общия вариационен ред числото ξ_i се предхожда от точно k величини от реда (η). Доказва се, че когато е вярна хипотезата $H_0 : F(x) = G(x)$, т. е. когато редиците (ξ) и (η) имат една и съща функция на разпределение, величината $U_{m,n}$ е случайна съответно със средно и дисперсия

$$a = \mathbf{E}U_{m,n} = \frac{mn}{2}, \quad \sigma^2 = \mathbf{D}U_{m,n} = \frac{mn}{12}(m+n+1).$$

При големи m и n (на практика такива се считат $m, n \geqq 10$) случайната величина $U_{m,n}$ при вярна хипотеза H_0 е асимптотично $\mathcal{N}(a, \sigma^2)$-разпределена. Следователно критичната област W с ниво на съгласие α може да се дефинира с неравенство от вида

$$\left|\frac{U_{m,n} - a}{\sigma}\right| > K_\alpha,$$

където константата K_α се определя от условието

$$\alpha = \mathbf{P}\left\{\left|\frac{U_{m,n} - a}{\sigma}\right| > K_\alpha \,|\, H_0\right\} \approx 1 - \Phi(K_\alpha) + \Phi(-K_\alpha) = 2[1 - \Phi(K_\alpha)].$$

Константата K_α е решение на уравнението $\Phi(t) = 1 - \frac{\alpha}{2}$, а критичната област може да бъде записана още с неравенствата

$$W = \begin{cases} U_{m,n} > a_{m,n} + \sigma^2_{m,n} t_{1-\frac{\alpha}{2}}, \\ U_{m,n} < a_{m,n} - \sigma^2_{m,n} t_{1-\frac{\alpha}{2}}. \end{cases}$$

Когато $U_{m,n} \subset W$, хипотезата H_0 се отхвърля, в противен случай H_0 не противоречи на резултатите от наблюденията.

Пример. За един и същ образец на чиста повърхност с правилен профил са проведени по 12 измервания с два двойни микроскопа с номера 61 и 263. Измерванията са дадени в табл. 1.

Таблица 1

Уред	Резултати от измерванията												Означения
№	1	2	3	4	5	6	7	8	9	10	11	12	i
61	0,8	1,9	3,0	3,5	3,8	2,5	1,7	0,9	1,0	2,3	3,3	3,4	(ξ)
263	1,4	2,1	3,1	3,6	2,7	1,8	1,1	0,2	1,6	2,8	4,0	4,7	(η)

При ниво на съгласие $\alpha = 0,05$ може ли да се счита, че между показанията на уредите няма систематични различия и че те имат еднаква разделителна способност?

Необходимо е да се провери хипотезата H_0 : разпределенията на двата реда измервания имат еднакви функции на разпределение.

Решение. Разполагаме данните в обща редица в растящ ред, като наблюденията над ξ са подчертани:

$$0{,}2\ \ \underline{0{,}8}\ \ \underline{0{,}9}\ \ \underline{1{,}0}\ \ 1{,}1\ \ 1{,}4\ \ 1{,}6\ \ \underline{1{,}7}\ \ 1{,}8\ \ \underline{1{,}9}\ \ 2{,}1\ \ \underline{2{,}3}$$

$$\underline{2{,}5}\ \ 2{,}7\ \ 2{,}8\ \ \underline{3{,}0}\ \ 3{,}1\ \ \underline{3{,}3}\ \ \underline{3{,}4}\ \ \underline{3{,}5}\ \ 3{,}6\ \ \underline{3{,}8}\ \ 4{,}0\ \ 4{,}7$$

Броят на инверсиите за (ξ) е

$$U = 1+1+1+4+5+6+6+8+9+9+9+10 = 69.$$

По горните формули намираме

$$\mathbf{E}U = \frac{12.12}{2} = 72;$$

$$\mathbf{D}U = \frac{12.12}{12}\cdot(12+12+1) = 300,$$

$$\sigma_n = \sqrt{300} \approx 17,3.$$

Критичната област за хипотезата H_0 при ниво на съгласие $\alpha = 0,05$ има вида (според нормалната апроксимация)

$$W = \begin{cases} U < 72 - 1,96.17,3 \approx 38; \\ U > 72 + 1,96.17,3 \approx 106. \end{cases}$$

Получената стойност за броя на инверсиите е $U = 69$ и тя не е в критичната област. Затова хипотезата H_0 не се отхвърля, т. е. няма основания да се счита, че двата уреда имат съществено различаващи се точности.

З а б е л е ж к а. При малък брой инверсии (или при твърде голямо U) наред с отхвърлянето на хипотезата H_0 може да се направи заключението, че случайната величина ξ е стохастично по-малка (стохастично по-голяма) от величината η. Означава се $\xi \prec \eta$ (или $\xi \succ \eta$) и се разбира, че за всяко число x е изпълнено $\mathbf{P}\{\xi < x\} \geqq \mathbf{P}\{\eta < x\}$, т. е. $F_\xi(x) \geqq F_\eta(x)$ при $\xi \prec \eta$.

§ 4. χ^2–критерий за еднородност

Когато се проверява хипотезата за еднаква разпределеност на случайните величини ξ и η по независими наблюдения $\xi_1, \dots, \xi_n$ и $\eta_1, \dots, \eta_m$ както в предишния параграф, вместо критерия на Уилкоксън може да се използва χ^2-критерият. Проверяваме основната хипотеза

$$H_0 : \mathbf{P}\{\xi < x\} = \mathbf{P}\{\eta < x\}, \quad x \in (-\infty, \infty),$$

независимо от точния вид на функцията на разпределение. За разлика от предишните случаи тук се допуска разпределенията да не са непрекъснати.

Процедурата на проверка на хипотезата H_0 срещу алтернативата $H_1 \neq H_0$ (отрицанието на H_0) е следната. Интервалът $[a, b]$, който съдържа извадките, се разделя на r подинтервала с поне по 5 наблюдения във всеки интервал. Нека това деление е

$$a = z_0 < z_1 < \cdots < z_{r-1} < z_r = b.$$

Да означим съответно с n_i и m_i броя на наблюденията от извадките (ξ) и (η), попаднали в интервала $[z_{i-1}, z_i]$. При вярна хипотеза H_0 величината

$$\widehat{\chi}^2 = nm \sum_{i=1}^{r} \frac{\left(\frac{n_i}{n} - \frac{m_i}{m}\right)^2}{n_i + m_i}$$

има асимптотично (при големи n и m) χ^2-разпределение с $r - 1$ степени на свобода. Затова при дадено ниво на съгласие α и при избрано r се определя критичната стойност $\chi^2_\alpha(r-1)$, с която по-нататък ще се сравнява пресметнатата от наблюденията статистика $\widehat{\chi}^2$.

Ако $\widehat{\chi}^2 \geqq \chi^2_\alpha(r-1)$, хипотезата H_0 се отхвърля, тъй като извадките не могат да се разглеждат като наблюдения от една и съща генерална съвкупност. Когато $\widehat{\chi}^2 < \chi^2_\alpha(r-1)$, хипотезата H_0 може да се приеме за вярна.

Пример. По статистически данни разпределението на доходите по възрастовите групи на 40-годишните (на възраст от 40 до 50 г.) и на 50-годишните (от 50 до 60 г.) работници е показано на табл. 2.

Таблица 2

Доходи (в крони)	До 1000	От 1000 до 2000	От 2000 до 3000	От 3000 до 4000	От 4000 до 6000	Над 6000	n, m
40 г.	7831	26740	35572	20009	11527	6919	108598
50 г.	7558	20685	24185	12280	6776	4222	75707

Може ли да се счита за значителна разликата в разпределението на доходите между 40-годишните и 50-годишните работници?

Решение. Ако положим $\widetilde{\omega}_i = \dfrac{n_i}{n_i + m_i}$, $\widetilde{\omega} = \dfrac{n}{n+m}$, то изразът за $\widehat{\chi}^2$ може да се препише в по-удобната за пресмятане форма

$$\widehat{\chi}^2 = \frac{1}{\widetilde{\omega}(1-\widetilde{\omega})}\left(\sum_{i=1}^{r} n_i\widetilde{\omega}_i - n\widetilde{\omega}\right).$$

Съответните пресмятания показват, че в случая $\widehat{\chi}^2 = 840,62$ при 5 степени на свобода. Дори при ниво на съгласие $\alpha = 0,001$ съответният квантил $\chi^2_{0,001}(5)$ е по-малък от $\widehat{\chi}^2$. Това е признак за силно значима разлика между разпределенията.

§ 5. χ^2–критерий за независимост и еднородност на дисперсиите

А. Често възникващ въпрос е дали два вида измервания на един и същ индивид на генералната съвкупност са зависими помежду си величини или не. Така например може да се постави въпроса, зависи ли ръстът на човека от неговото тегло или зависи ли кръвното му налягане от възрастта и пр.

Решаваме следната математическа задача: Нека ξ и η са две случайни величини, които се наблюдават в един и същ експеримент и образуват случайния вектор (ξ, η). Да се провери дали те са независими случайни величини, т. е. да се построи критерий за проверка на основната хипотеза

$$H_0 : \mathbf{P}\{\xi < x,\, \eta < y\} = \mathbf{P}\{\xi < x\}\mathbf{P}\{\eta < y\}.$$

Процедурата на решение е следната. Прави се извадка от едновременни наблюдения (ξ_1, η_1), ..., (ξ_n, η_n). Нека стойностите $\xi_1, \dots, \xi_n$ се съдържат в интервала $[a, b]$, а $\eta_1, \dots, \eta_n$ — в интервала $[c, d]$. Да разделим $[a, b]$ и $[c, d]$ съответно на r и на s подинтервала $a = z_0 < z_1 < \cdots < z_r = b$, $c = t_0 < t_1 < \cdots < t_s = d$. Нека n_{ij} е броят на двойките наблюдения (ξ_k, η_k), за които $\xi_k \in [z_{i-1}, z_i)$ и $\eta_k \in (t_{j-1}, t_j]$. Да положим

$$n_{\bullet j} = \sum_{i=1}^{r} n_{ij}, \quad n_{i\bullet} = \sum_{j=1}^{s} n_{ij}.$$

Представянето на данните (ξ_k, η_k) във вид на двумерна таблица от съвместни стойности на честоти по двойки подинтервали се нарича *корелационна таблица*, а представянето им в графичен вид в координатна система (ξ, η) — ***корелационно поле*** (вж. § 14.3). Очевидно

$$\sum_{j=1}^{s} n_{\bullet j} = \sum_{i=1}^{r} n_{i\bullet} = \sum_{i=1}^{r}\sum_{j=1}^{s} n_{ij} = n.$$

При вярна хипотеза H_0 величината

$$\widehat{\chi}^2 = n\sum_{i=1}^{r}\sum_{j=1}^{s} \frac{\left(n_{ij} - \frac{n_{i\bullet}n_{\bullet j}}{n}\right)^2}{n_{i\bullet}n_{\bullet j}} = n\sum_{i,j}\left(\frac{n_{ij}}{n_{i\bullet}n_{\bullet j}} - \frac{1}{n}\right)^2 n_{i\bullet}n_{\bullet j}$$

има асимптотично χ^2-разпределение с $(r-1)(s-1)$ степени на свобода.

Приемането или отхвърлянето на хипотезата H_0 става по познатия начин. При избрано ниво на съгласие α се определя квантилът $\chi^2_\alpha((r-1)(s-1))$ и се сравнява с пресметнатата от наблюденията стойност на $\widehat{\chi}^2$. Когато $\widehat{\chi}^2 \geqq \chi^2_\alpha$, хипотезата за независими разпределения на двете величини се отхвърля, в останалите случаи тя се приема като непротиворечаща на резултата от наблюденията.

Пример 1. Проверяват се $n = 500$ цилиндрични детайла относно два размера: диаметър ξ и дължина η. Всяко отклонение от нормите се счита за дефект. Резултатите са дадени в табл. 3.

Таблица 3

Дължина η	*Диаметър* ξ		
	Качествени	Дефектни	Всичко
Качествени	464	4	468
Дефектни	6	26	32
Всичко	470	30	500

Независими ли са признаците ξ и η?

Решение. В случая $r = s = 2$,

$$n_{11} = 464, \quad n_{12} = 4, \quad n_{1\bullet} = 468,$$
$$n_{21} = 6, \quad n_{22} = 26, \quad n_{2\bullet} = 32,$$
$$n_{\bullet 1} = 470, \quad n_{\bullet 2} = 30, \quad n = 500.$$

Статистиката $\widehat{\chi}^2$ има вида

$$\widehat{\chi}^2 = n\,\frac{(n_{11}n_{22} - n_{12}n_{21})^2}{n_{1\bullet}n_{2\bullet}n_{\bullet 1}n_{\bullet 2}} = 343,20.$$

От таблиците за $\chi^2(k)$ с $k = (r-1)(s-1) = 1$ степени на свобода (вж. табл. 2 от приложението) при ниво на съгласие $\alpha = 0,05$ отчитаме $\chi^2_{0,05}(1) = 3,84$.

Тъй като $\widehat{\chi}^2 = 343,2 > 3,84 = \chi^2_\alpha$, хипотезата H_0 се отхвърля. Двата размера не може да се приемат за независими случайни величини.

Б. Хипотезите относно дисперсиите заемат видно място в статистиката, тъй като само познаването на параметрите на разсейването на резултатите позволява да съдим за надеждността на оценките.

Възможни са няколко постановки:

Дадена е извадката $\xi_1, \ldots, \xi_n$ от независими наблюдения над случайната величина ξ с неизвестна дисперсия σ^2. Как да проверим основната хипотеза $H_0 : \sigma^2 = \sigma_0^2$ срещу алтернативата $H_1 : \sigma^2 > \sigma_0^2$? Задачата се среща на практика в случаите, когато по установените традиции и опит се знае, че дисперсията на някоя важна характеристика е σ_0^2. Но очевидно са се променили условията, при които се е работило с тази характеристика. Повлияли ли са те върху нейната дисперсия? Ако дисперсията не се е променила, то статистиката

$$(n-1)\,\frac{s_n^2}{\sigma_0^2} = \sum_{i=1}^{n} \frac{(\xi_i - \overline{\xi}_n)}{\sigma_0^2} = \widehat{\chi}^2$$

асимптотично ще има χ^2-разпределение с $n-1$ степени на свобода. Критерият за проверка на H_0 с ниво на съгласие α срещу алтернативата H_1 изглежда така:

$$(n-1)\,\frac{s_n^2}{\sigma_0^2} > \chi^2_\alpha(n-1),$$

където $\chi^2_\alpha(n-1)$ се определя по таблиците за χ^2-разпределението с $n-1$ степени на свобода от уравнението

$$\mathbf{P}\{\chi^2(n-1) > \chi^2_\alpha(n-1)\} = \alpha.$$

Когато $\widehat{\chi}^2 < \chi^2_\alpha$, хипотезата H_0 се приема. При $\widehat{\chi}^2 > \chi^2_\alpha$ хипотезата, че наблюдаваната величина ξ има предполагаемата дисперсия σ_0, се отхвърля.

Пример 2. Произведените на автоматичен струг цилиндрични шайби имат дължина, която е случайна величина ξ с нормално разпределение $\mathcal{N}(a, \sigma^2)$. Известно е, че точността на настройката на струга е ± 20 μm, т. е. $\sigma_0 = 20$ μm, или че $\mathbf{D}\xi = 400$. Направени са пробни измервания на 20 шайби, от които са сметнати $\bar{\xi}_n = 18$ μm и $s_n^2 = 784$. При ниво на съгласие $\alpha = 0,02$ да се провери хипотезата $H_0 : \mathbf{D}\xi = 400$ срещу алтернативата $H_1 : \mathbf{D}\xi \neq 400$ (което съответства на разстройка на струга).

Решение. По таблиците за χ^2-разпределение с 19 степени на свобода при ниво на съгласие $\alpha = 0,02$ определяме квантила $\chi^2_\alpha = 34,9$. Емпиричната статистика дава

$$\widehat{\chi}^2 = 19 \cdot \frac{784}{400} = 37,24.$$

Тъй като $\widehat{\chi}^2 > \chi^2_{0,02}(19)$, трябва да се приеме, че точността на струга се е изменила (т. е. $\mathbf{D}\xi \neq 400$).

§ 6. F-критерий на Фишер

F-критерият на Фишер най-често се използва в статистиката за проверка на основната хипотеза H_0: в две генерални съвкупности дисперсиите на даден количествено измерим признак са равни. При това от двете съвкупности са получени извадки

$$(\xi) : \ \xi_1, \ldots, \xi_n;$$

$$(\eta) : \ \eta_1, \ldots, \eta_m,$$

които са независими наблюдения в съвкупност. Хипотезата H_0 може да се изрази математически с равенството $\mathbf{D}\xi = \mathbf{D}\eta$ и трябва да се провери срещу алтернативата $H_1 : \ \mathbf{D}\xi \neq \mathbf{D}\eta$.

Хипотезите за дисперсиите играят голяма роля в науката и практиката. От измерваната чрез дисперсията величина на разсейването се характеризират такива важни конструкторски и технологични показатели като точност на машините и уредите, грешки в показанията на измервателните уреди, точност на технологичния процес и т. н. От резултатите при проверката на хипотезата за равенство на дисперсиите зависят и някои по-нататъшни статистически процедури на обработка на двата реда от наблюдения за проверка на други хипотези (вж. § 7).

Проверката на хипотезата H_0 използва статистическия показател

$$\widehat{F} = \frac{\widehat{s}^2_\xi}{\widehat{s}^2_\eta},$$

където $\widehat{s}_\xi^2$ и $\widehat{s}_\eta^2$ са неизместени оценки на дисперсиите $\mathbf{D}\xi$ и $\mathbf{D}\eta$ (вж. десета глава). Когато е вярна хипотезата $H_0 : \mathbf{D}\xi = \mathbf{D}\eta$ и случайните величини ξ и η са нормално разпределени, показателят $\widehat{F}$ има F-разпределение на Фишер с $(n-1, m-1)$ степени на свобода. Ако хипотезата H_0 не е вярна, $\widehat{F}$ няма F-разпределение.

Критичната област за критерия $\widehat{F}$ при дадено ниво на съгласие α се строи така: Хипотезата H_0 се отхвърля, когато $\widehat{F} > F_{1-\frac{\alpha}{2}}$ или $\widehat{F} < F_{\frac{\alpha}{2}}$.

Тук числата $F_{1-\frac{\alpha}{2}}$ и $F_{\frac{\alpha}{2}}$ се избират съответно като $\left(1-\dfrac{\alpha}{2}\right)$-квантил и $\dfrac{\alpha}{2}$-квантил на $F(n-1, m-1)$–разпределението, т. е. като решения на уравненията

$$\mathbf{P}\left\{F(n-1, m-1) > F_{1-\frac{\alpha}{2}}\right\} = \frac{\alpha}{2};$$

$$\mathbf{P}\left\{F(n-1, m-1) < F_{\frac{\alpha}{2}}\right\} = \frac{\alpha}{2}.$$

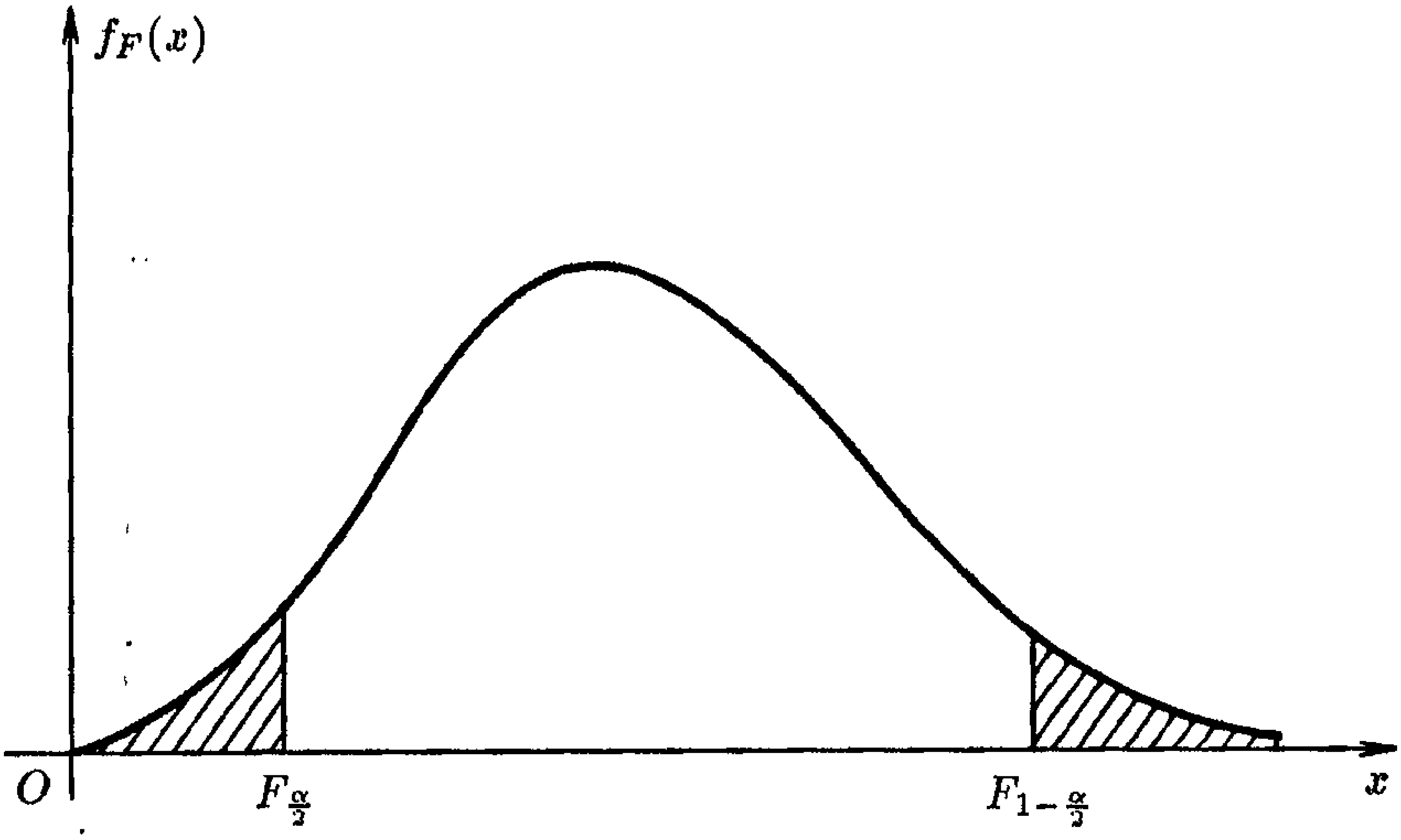
Този избор на критичната област осигурява най-голяма чувствителност на F-критерия — отхвърлят се едновременно много големите и много малките стойности на $\widehat{F}$.

На фиг. 25 е представена кривата на F-разпределението. Всяко от защрихованите лица е равно на $\dfrac{\alpha}{2}$. Фигуративно казано, когато извадковата стойност $\widehat{F}$ се окаже в защрихованата критична област, хипотезата $\mathbf{D}\xi = \mathbf{D}\eta$ трябва да бъде отхвърлена.

Нека $\widehat{F}' = (\widehat{F})^{-1} = s_\eta^2/s_\xi^2$. При вярна хипотеза H_0 и нормално разпределени ξ и η величината $\widehat{F}'$ има F-разпределение на Фишер с $(m-1, n-1)$ степени на свобода. С помощта на функцията $\widehat{F}'$ вероятността $\mathbf{P}\left\{\widehat{F} < F_{\frac{\alpha}{2}}\right\}$ може да бъде изразена така:

$$\frac{\alpha}{2} = \mathbf{P}\left\{\widehat{F} < F_{\frac{\alpha}{2}}\right\} = \mathbf{P}\left\{\frac{1}{\widehat{F}} > \frac{1}{F_{\frac{\alpha}{2}}}\right\} = \mathbf{P}\left\{\widehat{F}' > \frac{1}{F_{\frac{\alpha}{2}}}\right\}.$$

Този резултат показва, че лявата критична граница на критерия $\widehat{F}$ съответства на дясната критична граница на критерия $\widehat{F}'$. Затова е достатъчно да се намерят само десните критични граници



Фиг. 25

за $\widehat{F}$ и $\widehat{F}'$. Тази е причината в съществуващите таблици за F-разпределението да се дават само десните критични граници, т. е. квантилите $F_{1-\frac{\alpha}{2}}$, табулирани за различни съчетания на степените на свобода $k_1 = n - 1$ и $k_2 = m - 1$, и то за $k_1 < k_2$.

Забележка. Обикновено при пресмятането на $\widehat{F}$ се взема отношението на по-голямата по стойност от двете дисперсии $\widehat{s}^2_\xi$ и $\widehat{s}^2_\eta$ към по-малката. Тогава можем да се условим отхвърлянето на хипотезата за равенство на дисперсиите да става, когато емпирично определеният критерий $\widehat{F}$ има стойност, по-голяма от критичната граница $F_{1-\frac{\alpha}{2}}$. Така ще се налага само едно търсене в таблиците за F-разпределението.

Пример. При обработка на втулки (по еднакъв образец) на два автоматични струга са отбрани две проби. От първия струг са взети 10 втулки, а от втория — 15. По данните от тези проби са пресметнати емпиричните дисперсии $\widehat{s}^2_1 = 9,6\ \mu$k и $\widehat{s}^2_2 = 5,7\ \mu$k. При ниво на съгласие $\alpha = 0,05$ може ли да се твърди, че двата струга произвеждат детайлите с еднаква точност?

Решение. Критерият $\widehat{F}$ има стойност

$$\widehat{F} = \frac{\widehat{s}^2_1}{\widehat{s}^2_2} = \frac{9,6}{5,7} = 1,68.$$

Ако хипотезата за еднаква точност на произвежданите детайли е вярна, $\widehat{F}$ трябва да има разпределение $F(9,14)$ с $(9,14)$ степени на свобода. От таблиците за F-разпределението виждаме (с интерполация на междинните стойности, когато липсват данни за съответното ниво на съгласие или съответните степени на свобода), че $F_{0,975}(9,14) = 2,65$. Тъй като $\widehat{F} = 1,68 < 2,65$, то предположението за еднаква точност на произвежданите детайли на двата струга не противоречи на резултатите от извършените проби.

§ 7. Критерий на Стюдънт

Една от централните задачи в математическата статистика е проверката на хипотези за математическите очаквания. Вярното определяне на средното за даден параметър в генералната съвкупност по емпирични данни от извадката и сравнението на две и повече съвкупности по средните стойности на дадена тяхна характеристика е обичайна задача за повечето експерименти. За пълнота ще разгледаме и параметричния случай на използване на t-критерия, като изрично ще отбележим, че неговата коректност е безспорна само за случая на нормално разпределена проста извадка.

А. Нека $\xi_1, \ldots, \xi_n$ са независими наблюдения над случайна величина със средно a и дисперсия σ^2. Емпиричните оценки за a и σ^2 са величините

$$\overline{\xi}_n = \frac{1}{n}(\xi_1 + \cdots + \xi_n); \quad \widehat{s}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\xi_i - \overline{\xi}_n)^2.$$

Обикновено не знаем нищо за a и σ^2. Проверката на хипотезата $\mathbf{E}\xi = a_0$ се извършва с помощта на едностранни и двустранни критерии в зависимост от характера на алтернативната хипотеза H_1.

Ако алтернативата H_1 е, че $\mathbf{E}\xi > a_0$, то критичната област за проверка на хипотезата H_0 срещу алтернативата H_1 с ниво на съгласие α изглежда така:

$$\widehat{t}_n = \frac{\widehat{\xi}_n - a_0}{\widehat{s}_n\sqrt{n}} > t_\alpha.$$

Тук $\widehat{\xi}_n$ и $\widehat{s}_n^2$ са емпиричните оценки за средното и дисперсията на ξ, а числото t_α се определя от условието $\mathbf{P}\{t_n > t_\alpha\} = \alpha$. Величината $\widehat{t}_n$ при големи n има приблизително разпределение на

Стюдънт (t-разпределение) с $n-1$ степени на свобода. Числото t_α се определя по дадени n и α от таблиците за t-разпределението като негов α-квантил.

Ако алтернативната хипотеза е $H_1 : \mathbf{E}\xi \neq a_0$, критичната област за проверка на хипотезата $\mathbf{E}\xi = a_0$ срещу алтернативата H_1 (двустранен критерий) изглежда така:

$$|\overline{\xi}_n - a_0| > t_{\frac{\alpha}{2}} \widehat{s}_n \sqrt{n}. \tag{1}$$

Тук $t_{\frac{\alpha}{2}}$ се определя от условието $\mathbf{P}\left\{t_n > t_{\frac{\alpha}{2}}\right\} = \frac{\alpha}{2}$.

Пример 1. В един завод се изпробва нова технология за производство на радиолампи. Проверени са 9 лампи, които показват, че средният им срок за безотказна работа е 590 часа и дисперсията е $\widehat{s}^2 = 100$. Известно е, че средният срок за безотказна работа на радиолампите, произвеждани по старата технология, е 500 часа. Може ли да се твърди при ниво на съгласие $\alpha = 0,05$, че новата технология е изменила средния срок за безотказна работа на радиолампите?

Решение. Построяваме критерий за проверка на хипотезата $H_0 : \mathbf{E}\xi = 500$ срещу алтернативата $H_1 : \mathbf{E}\xi > 500$, където ξ е срокът за безотказна работа на радиолампите при новата технология. Емпиричните резултати дават

$$\widehat{t}_n = \frac{590-500}{\sqrt{100}\sqrt{9}} = \frac{90}{30} = 3.$$

Ако е вярна H_0 (и освен това, ако ξ е нормално разпределена случайна величина), $\widehat{t}_n$ трябва да има t-разпределение с 8 степени на свобода. По таблицата на t-разпределението намираме, че решението на уравнението

$$\mathbf{P}\{t(8) > t_{0,05}\} = 0,05$$

е $t_{0,05} = 2,306$. Тъй като $\widehat{t}_n = 3 > 2,306$, хипотезата H_0 следва да се отхвърли, т. е. новата технология съществено променя средния срок за безотказна работа на радиолампите, като го увеличава.

Б. Нека са дадени две независими извадки

$$(\xi): \ \xi_1, \ldots, \xi_n;$$
$$(\eta): \ \eta_1, \ldots, \eta_m.$$

Критерият на Стюдънт може да се използва за проверка на хипотезата $H_0: \ \mathbf{E}\xi = \mathbf{E}\eta$ срещу алтернативата $H_1: \ \mathbf{E}\xi \neq \mathbf{E}\eta$.

Тази задача е важна в практиката, когато възниква въпрос за сравняване на средните резултати от две технологии за производство на едни и същи изделия, за еднотипност на производството от две различни предприятия и пр.

За построяването на критичната област е важно да се знаят дисперсиите σ_ξ^2, σ_η^2 на двете величини ξ и η. Процедурата е различна за случаите $\sigma_\xi^2 = \sigma_\eta^2$ и $\sigma_\xi^2 \neq \sigma_\eta^2$. Ето защо задачата за сравняване на две средни величини по емпирични данни се предхожда обикновено от задачата за сравняване на дисперсиите в двете извадки (по този въпрос вж. § 6).

По данните от наблюденията (ξ) и (η) се определят емпиричните средни $\widehat{\xi}_n$ и $\widehat{\eta}_n$ и емпиричните дисперсии s_ξ^2 и s_η^2. След това се сравняват дисперсиите σ_ξ^2 и σ_η^2 по F-критерия, описан в § 6.

Ако е потвърдено предположението, че $\sigma_\xi^2 = \sigma_\eta^2$, за общата им дисперсия се полага

$$s^2 = \frac{s_\xi^2 \cdot f_1 + s_\eta^2 \cdot f_2}{f_1 + f_2},$$

където $f_1 = n - 1$, $f_2 = m - 1$. Критичната област с ниво на съгласие α изглежда така:

$$\widehat{t} = \frac{|\overline{\xi}_n - \overline{\eta}_m|}{s\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}} > t_{\frac{\alpha}{2}}. \tag{2}$$

Когато е вярна хипотезата H_0, величината $\widehat{t}$ има приблизително разпределение на Стюдънт с $n + m - 2$ степени на свобода. Числото $t_{\frac{\alpha}{2}}$ се определя от условието $\mathbf{P}\left\{t(n+m-2) > t_{\frac{\alpha}{2}}\right\} = \dfrac{\alpha}{2}$ по таблиците за разпределението на Стюдънт при даденото α.

Пример 2. Въз основа на две серии от наблюдения за ξ с $n = 7$ наблюдения и за η с $m = 11$ наблюдения се оказало, че $\overline{\xi}_n = 5$; $\widehat{s}_\xi^2 = 0,8$; $\overline{\eta}_m = 4,6$; $\widehat{s}_\eta^2 = 0,5$. Да проверим хипотезата $\mathbf{E}\xi = \mathbf{E}\eta$ при $\alpha = 0,05$.

Решение. Да сравним отначало дисперсиите. Тяхната разлика е незначителна. Наистина при $\alpha = 0,05$ по таблиците за разпределението на Фишер със степени на свобода $(6,10)$ може да се види, че $F_{\frac{\alpha}{2}} = 0,12$, а $F_{1-\frac{\alpha}{2}} = 3,2$. Отношението на емпиричните дисперсии е

$$\widehat{F} = \frac{0,8}{0,5} = 1,6.$$

Обединената оценка за дисперсията на ξ и η е

$$\widehat{s}^2 = \frac{6.0,8 + 10.0,5}{6+10} = 0,61.$$

Сега да построим t-критерия за проверка на равенството на средните $\mathbf{E}\xi = \mathbf{E}\eta$. По (2) пресмятаме $\widehat{t} = 1,36$. При ниво на съгласие $\alpha = 0,05$ и t-разпределение с 16 степени на свобода $(16 = 6 + 10)$ определяме $t_{\frac{\alpha}{2}} = 3,3$. Тъй като $1,36 < 3,3$, хипотезата, че $\mathbf{E}\xi = \mathbf{E}\eta$, трябва да се приеме.

Ако сравняването на дисперсиите в двете извадки е довело до извода, че $\sigma_\xi^2 \neq \sigma_\eta^2$, критичната област изглежда така:

$$\widehat{T} = \frac{|\overline{\xi}_n - \overline{\eta}_m|}{\sqrt{\dfrac{\widehat{s}_\xi^2}{n} + \dfrac{\widehat{s}_\eta^2}{m}}} > t_{\frac{\alpha}{2}}. \tag{3}$$

Тук $\widehat{T}$ е също разпределена (при големи m и n) по закона на Стюдънт с $m + n - 2$ степени на свобода.

Едностранните t-критерии се прилагат, когато алтернативната хипотеза е $H_1 : \mathbf{E}\xi > \mathbf{E}\eta$ (или $\mathbf{E}\xi < \mathbf{E}\eta$). Тогава в числителите на посочените критерии се взема просто разликата $\overline{\xi}_n - \overline{\eta}_n$, а критичната граница при ниво на съгласие α вместо с квантила $t_{\frac{\alpha}{2}}$ се определя с квантила t_α.

Пример 3. При пробни изпитания за разкъсване на две извадки с по 50 намотки от кабели, произведени в два завода, се оказало, че

$$\overline{\xi}_n = 120,8\ \text{kg/mm}^2 \quad (\text{за завод А}),$$

$$\overline{\eta}_m = 128,2\ \text{kg/mm}^2 \quad (\text{за завод Б}).$$

От предишни наблюдения и експерименти е доказано, че ξ и η са нормално разпределени величини със средно-квадратични отклонения $\widehat{s}_\xi = 8,0\ \text{kg/mm}^2$ и $\widehat{s}_\eta = 9,4\ \text{kg/mm}^2$. Да се определи при ниво на съгласие $\alpha = 0,05$ дали има съществено различие в механичните качества на кабелите, произвеждани в заводите А и Б.

Решение. Определяме $\widehat{T} = 4,23$ съгласно (3). При вярна хипотеза $\mathbf{E}\xi = \mathbf{E}\eta$ величината $\widehat{T}$ следва да има приблизително t-разпределение с $50 + 50 - 2 = 98$ степени на свобода. При голямо n обаче разпределението на $t(n)$ е много близко до стандартно нормално $\mathcal{N}(0,1)$. Затова в таблиците за квантилите на $t(n)$ се срещат стойности на n до 30 – 40. За $n > 40$ се ползва таблицата за $\Phi(x)$. Коренът на уравнението $\Phi(x) = 1 - 0,0025$ е $x = 1,96$.

Тъй като $\widehat{T} = 4,23 > 1,96$, следва да заключим, че разликите в качеството на продукцията на заводите А и Б са съществени.

Критерият на Стюдънт може да се използва и в случаите, когато наблюденията в двете извадки (ξ) и (η) не са нормално разпределени и се проверява хипотезата $H_0 : \mathbf{E}\xi = \mathbf{E}\eta$ срещу алтернативата $H_1 : \mathbf{E}\xi \neq \mathbf{E}\eta$.

Основанията за това се съдържат в следните разсъждения: При големи n и m извадковите средни $\overline{\xi}_n$ и $\overline{\eta}_m$ по централната гранична теорема могат да се разглеждат като приблизително нормално разпределени — съответно $\mathcal{N}\left(\overline{\xi}_n, \dfrac{\widehat{s_\xi^2}}{n}\right)$ и $\mathcal{N}\left(\overline{\eta}_m, \dfrac{\widehat{s_\eta^2}}{m}\right)$.

Тогава разликата ще бъде (по законите за разпределение на суми от независими нормални случайни величини) приблизително нормално разпределена $\mathcal{N}\left(\overline{\xi}_n - \overline{\eta}_m; \dfrac{\widehat{s_\xi^2}}{n} + \dfrac{\widehat{s_\eta^2}}{m}\right)$. Всяка хипотеза относно разликата $\mathbf{E}\xi - \mathbf{E}\eta$ може да се проверява, като се използва този факт. Специално критичната област за хипотезата H_0 при ниво на съгласие α ще бъде от вида

$$\frac{|\overline{\xi}_n - \overline{\eta}_m|}{\sqrt{\dfrac{\widehat{s_\xi^2}}{n} + \dfrac{\widehat{s_\eta^2}}{m}}} > K_\alpha, \tag{4}$$

където K_α се определя като решение на уравнението $\Phi(K_\alpha) = 1 - \dfrac{\alpha}{2}$.

Взето е предвид, че величината отляво на (4) е $\mathcal{N}(0,1)$-разпределена (да се запомни! — при големи n и m, $n \geqq 30$, $m \geqq 30$).

§ 8. Дисперсионен анализ

Дисперсионният анализ обединява група методи, свързани с изследването на съвкупности от нормално разпределени случайни величини. Тези методи съществено използват условието за нормалност на разпределението на изучаваните величини. Обикновено дисперсионният анализ е свързан с критериите за проверка дали са равни средните стойности в няколко групи извадки. Тези критерии ще изложим тук като продължение на задачите от § 7. Но дисперсионният анализ е свързан и с редица други задачи като проверките за адекватност на регресионни и други модели от многомерния статистически анализ, които в нашия курс не изучаваме.

Нека $x_{j1}, \ldots, x_{jn_j}$ представляват проста извадка с обем n_j от случайната величина $\xi_i \in \mathcal{N}(a_j, \sigma^2)$, $j = 1, \ldots, r$, с еднакви дисперсии. Да означим $\vec{x} = (x_{11}, \ldots, x_{1n_1}, \ldots, x_{r1}, \ldots, x_{rn_r})$. Вярна е следната

Теорема 1. *Критерият W^* за проверка на сложната хипотеза*

$$H_0 : a_1 = a_2 = \cdots = a_r = a, \ \sigma^2 > 0,$$

срещу сложната алтернатива H_1 : не всички a_j са равни, $\sigma^2 > 0$, с ниво на съгласие α изглежда така:

$$(1) \qquad W^* = \left\{ \vec{x} : \ \frac{\frac{1}{r-1} \sum\limits_{j=1}^{r} n_j (\overline{x}_j - \overline{\overline{x}})^2}{\frac{1}{n-r} \sum\limits_{j=1}^{r} \sum\limits_{k=1}^{n_j} (x_{ij} - \overline{x}_j)^2} > F_{1-\alpha}(r-1, n-r) \right\},$$

където $n = n_1 + \cdots n_r$; $F_{1-\alpha}(r-1, n-r)$ е $(1-\alpha)$-квантил на F-разпределението на Фишер с $(r-1, n-r)$ степени на свобода и е положено

$$\overline{x}_j = \frac{1}{n_j} \sum_{k=1}^{n_j} x_{jk}; \qquad \overline{\overline{x}} = \frac{1}{n} \sum_{j=1}^{r} \sum_{k=1}^{n_j} x_{jk}.$$

Доказателство. Съгласно критерия с частно на правдоподобията от § 11.5 критичната област W^* се определя с неравенството

$$(2) \qquad \lambda(\vec{x}) = \frac{\sup\limits_{a,\sigma} L(\vec{x}; a, \sigma^2)}{\sup\limits_{a_1,\ldots,a_r,\sigma} L(\vec{x}; a_1, \ldots, a_r, \sigma^2)} \leq K_\alpha,$$

където

$$L(\vec{x}; a_1, \ldots, a_r, \sigma^2) = \prod_{j=1}^{r}\prod_{k=1}^{n_j} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x_{jk}-a_j)^2}{2\sigma^2}\right]$$

и е положено $L(\vec{x}; a, \sigma^2) = L(\vec{x}; a, \ldots, a, \sigma^2)$. Числителят и знаменателят на (2) достигат своите екстремални (максимални) стойности съответно (проверете!) за

$$a = \overline{\overline{x}}, \quad \sigma^2 = \sum_{j=1}^{r}\sum_{k=1}^{n_j}(x_{jk} - \overline{\overline{x}})^2 \quad \text{в числителя и}$$

$$a_j = \overline{x}_j, \quad \sigma^2 = \frac{1}{n}\sum_{j=1}^{r}\sum_{k=1}^{n_j}(x_{jk} - \overline{x}_j)^2 \quad \text{в знаменателя.}$$

След прости преобразувания (добавяне и изваждане на $\overline{x}_j$ в събираемите на числителя) се убеждаваме, че неравенството (2) е еквивалентно на

$$(3) \qquad \mathfrak{F}(\vec{x}) = \frac{1}{r-1}\sum_{j=1}^{r}\left(\frac{n_j(\overline{x}_j - \overline{\overline{x}})^2}{\frac{1}{n-r}\sum\limits_{i=1}^{r}\sum\limits_{k=1}^{n_j}(x_{jk} - \overline{x}_j)^2}\right) \geqq C.$$

Лесно е да се убедим, че ако хипотезата H_0 е вярна и означим $y_{jk} = \dfrac{x_{jk} - a}{\sigma}$ за всички $k = 1, \ldots, n_j$, $j = 1, \ldots, r$, то статистиките $\mathfrak{F}(\vec{x})$ и $\mathfrak{F}(\vec{y})$ са еднакво разпределени. Ще покажем, че $\mathfrak{F}(\vec{y})$ има F-разпределение на Фишер с $(r-1, n-r)$ степени на свобода, с което доказателството на теоремата ще бъде завършено.

Нека

$$\overline{y}_j = \frac{1}{n_j}\sum_{k=1}^{n_j} y_{jk}; \qquad \overline{\overline{y}} = \frac{1}{n}\sum_{j=1}^{r}\sum_{k=1}^{n_j} y_{jk}.$$

Имаме предвид, че $y_{jk} \in \mathcal{N}(0,1)$ и са независими. Лесно е да проверим, че

$$(4) \qquad \begin{aligned} \sum_{j=1}^{r}\sum_{k=1}^{n_j} y_{jk}^2 &= \sum_{j=1}^{r}\sum_{k=1}^{n_j}(y_{jk} - \overline{y}_j)^2 + \sum_{j=1}^{r} n_j \overline{y}_j \\ &= \sum_{j=1}^{r}\sum_{k=1}^{n_j}(y_{jk} - \overline{y}_j)^2 + \sum_{j=1}^{r} n_j(\overline{y} - \overline{\overline{y}})^2 - n\overline{\overline{y}}^2, \end{aligned}$$

което може да се запише като сума на три квадратични форми ($\mathbf{Y}'$ е транспонираният вектор–стълб на n-мерния вектор–ред $\mathbf{Y} = (y_{11}, \ldots, y_{1n_1}, \ldots, y_{r1}, \ldots, y_{rn_r})$):

$$\mathbf{Y}'\mathbf{Y} = \mathbf{Y}'\mathbf{A}\mathbf{Y} + \mathbf{Y}'\mathbf{B}\mathbf{Y} + \mathbf{Y}'\mathbf{C}\mathbf{Y}. \tag{5}$$

Като забележим, че $\sum_{k=1}^{n_j} (y_{jk} - \overline{y}_j)^2 = \sum_{k=1}^{n_j} y_{jk}^2 - n_j\overline{y}_j^2$, лесно намираме матрицата на тази квадратична форма:

$$\mathbf{D}_j = \begin{pmatrix} 1 - \frac{1}{n_j}, & -\frac{1}{n_j}, & \ldots, & -\frac{1}{n_j} \\ -\frac{1}{n_j}, & 1 - \frac{1}{n_j}, & \ldots, & -\frac{1}{n_j} \\ \ldots\ldots\ldots & & & \\ -\frac{1}{n_j}, & -\frac{1}{n_j} & \ldots, & 1 - \frac{1}{n_j}, \end{pmatrix}.$$

Стълбовете на $\mathbf{D}_j$ са линейно зависими, тъй като сумата им е 0. Значи $\operatorname{rank}(\mathbf{D}_j) \leqq n_j - 1$. По-нататък, ако извадим първия стълб на $\mathbf{D}_j$ от всеки от останалите, ще получим матрица със същия ранг като $\mathbf{D}_j$. Но получената матрица може да се запише във вида

$$\mathbf{D}_j^* = \begin{pmatrix} 1 - \frac{1}{n_j} & \mathbf{F} \\ \mathbf{E} & \mathbf{I} \end{pmatrix},$$

където $\mathbf{E} = \left(-\frac{1}{n_j}, \ldots, -\frac{1}{n_j}\right)$, $\mathbf{F} = (-1, -1, \ldots, -1)$, а $\mathbf{I}$ е единичната матрица от тип $(n_j - 1) \times (n_j - 1)$. Това показва, че е изпълнено $\operatorname{rank}(\mathbf{D}_j) = \operatorname{rank}(\mathbf{D}_j^*) = n_j - 1$. Тъй като за различни стойности на j квадратичната форма в първото събираемо на (4) и (5) включва различни групи ($j = 1, 2, \ldots, r$) от случайната величина, то рангът на $\mathbf{Y}'\mathbf{A}\mathbf{Y}$, т. е. рангът на $\mathbf{A}$, е $\sum_{j=1}^{r} (n_j - 1) = n - r$.

Рангът на третия член $\mathbf{Y}'\mathbf{C}\mathbf{Y}$ в (5) се пресмята лесно, тъй като всички членове на $\mathbf{C}$ са равни на $\frac{1}{n}$, т. е. $\operatorname{rank}(\mathbf{C}) = 1$. Остава да пресметнем $\operatorname{rank}(\mathbf{B})$. Понеже $\mathbf{I} = \mathbf{A} + \mathbf{B} + \mathbf{C}$, имаме

$$\begin{aligned} n = \operatorname{rank}(\mathbf{I}) &= \operatorname{rank}(\mathbf{A} + \mathbf{B} + \mathbf{C}) \leqq \operatorname{rank}(\mathbf{A}) + \operatorname{rank}(\mathbf{B}) + \operatorname{rank}(\mathbf{C}) \\ &= n - r + \operatorname{rank}(\mathbf{B}) + 1, \end{aligned}$$

или

(6) $$\operatorname{rank}(\mathbf{B}) \geqq r-1.$$

Ако положим $L_j = \sqrt{n_j}(\overline{y}_j - \overline{\overline{y}})$, $j = 1, \ldots, r$, ще е налице представянето

$$\mathbf{Y}'\mathbf{B}\mathbf{Y} = \sum_{j=1}^{r} n_j(\overline{y}_j - \overline{\overline{y}})^2 = \sum_{j=1}^{r} L_j^2.$$

Обаче линейните форми L_j, $j = 1, \ldots, r$, са линейно зависими, тъй като

$$\sum_{j=1}^{r} \sqrt{n_j} L_j = \sum_{j=1}^{r} n_j(\overline{y}_j - \overline{\overline{y}}) = \sum_{j=1}^{r} n_j \overline{y}_j - n\overline{\overline{y}} = 0.$$

Следователно $\operatorname{rank}(\mathbf{B}) \leqq r-1$, което съгласно (6) показва, че $\operatorname{rank}(\mathbf{B}) = r-1$.

Ще отбележим, че сумата от квадратичните форми (5) удовлетворява следната

Теорема на Кокрън. *Нека $\vec{\eta} = (\eta_1, \ldots, \eta_n)$ са n независими $\mathcal{N}(0,1)$-разпределени случайни величини и $Q_1(\vec{\eta}), \ldots, Q_m(\vec{\eta})$ са m квадратични форми, такива че $\vec{\eta}\vec{\eta}' = \sum_{i=1}^{m} Q_i(\vec{\eta})$. Нека $r_i = \operatorname{rank}(Q_i)$, а $r_1 + \cdots + r_m = n$. Тогава $Q_1(\vec{\eta}), \ldots, Q_m(\vec{\eta})$ са m независими случайни величини и $Q_i(\vec{\eta})$ имат χ^2-разпределение с r_i степени на свобода.*

Доказателството на теоремата на Кокрън може да се види в [27, п. 3, гл. 10] и ние няма да го привеждаме тук.

Въз основа на теоремата на Кокрън при вярна хипотеза H_0 събираемите $\vec{y}'A\vec{y}$, $\vec{y}'B\vec{y}$ и $\vec{y}'C\vec{y}$ в (5) са независими и всяко от тях има χ^2-разпределение съответно с $n-s$, $s-1$ и 1 степени на свобода. Но критичната област (3) е дефинирана с отношението

$$\mathfrak{F}(\vec{y}) = \frac{1}{r-1} \cdot \frac{\vec{y}'B\vec{y}}{\left[\frac{1}{n-r}\vec{y}'A\vec{y}\right]}$$

и съгласно дефиницията на F-разпределението то има разпределение на Фишер с $(r-1, n-r)$ степени на свобода. Следователно хипотезата H_0 следва да се отхвърли, когато настъпи събитието

$$\{\mathfrak{F}(\vec{x}) \geqq F_{1-\alpha}(r-1, n-r)\},$$

както се твърди в теоремата.

Задачи

1. Две групи по 80 животни на възраст 6 месеца се отглеждат при два различни режима. Предимствата на отглеждане се оценяват по едномесечния прираст на всяко от животните. Оказало се, че прирастът в първата група преобладава в 47 случая, а във втората — в 32 от случаите. В една от двойките предварително номерирани съответни животни прирастът бил оценен като еднакъв. При ниво на съгласие $\alpha = 0,05$ може ли да се твърди, че двата режима на отглеждане дават еднакъв прираст?

Упътване: Покажете, че критичната област има приблизително следния вид: $|\nu_n - 39,5| \geqq 0,92\sqrt{79}$.

2. Измервателен уред при правилна изходна центровка дава измервания с грешки, които са нормално разпределени със средно 0 и дисперсия $\sigma^2 = 0,25$. При 6 измервания с този уред са получени резултатите (подредени по големина): -1; $-0,5$; 0; $0,5$; 1 и 2. Покажете, че при ниво на съгласие $\alpha = 0,05$ резултатът $x_{\max} = 2$ съдържа груба грешка и трябва да бъде отстранен от по-нататъшни разглеждания.

3. При изследване на точността на технологична операция в една от пробите за отклонения в размерите на произвежданите детайли от стандартните размери са получени следните резултати (в mm, подредени по големина): $0,07$; $0,09$; $0,10$; $0,12$; $0,13$; $0,16$; $0,17$; $0,25$.

Покажете, че при ниво на съгласие $\alpha = 0,05$ резултатът не съдържа груба грешка.

4. При проучване на разпределението на месечните доходи според възрастта на работниците в един завод са получени следните резултати:

Възраст (год.)	Доход (лв.)					
	До 100	100–150	150–200	250–300	300–400	Над 400
40–50	71	430	1072	1609	1178	158
50–60	54	324	894	1202	903	112

Покажете, че при ниво на съгласие $\alpha = 0,05$ няма разлика в заплащането на двете възрастови групи (двете извадки са извършени от една и съща генерална съвкупност през 1987 г.).

5. Два завода произвеждат стомана. Извършени са измервания за процентното съдържание на сяра в отливки, изработени в двата завода. Данните са нанесени в следващата таблица:

Завод	% на сярата			
	0–0,2	0,2–0,4	0,4–0,6	0,6–0,8
А	82	535	1173	1714
Б	63	429	995	1307

Покажете, че при ниво на съгласие $\alpha = 0,05$ няма основание да се счита, че процентното съдържание на сяра в стоманата, изливана в двата завода, е различно.

6. При социологическо проучване през 1986 г. били анкетирани 25 263 семейства. След обработка на резултатите се оказало, че могат да бъдат отделени следните групи според разпределението на годишните доходи и броя на децата в семейството:

Брой на децата	Годишен доход (в лв.) на човек от семейството			
	до 600	600–1000	1000–1500	над 1500
0	2161	3577	2184	1636
1	2755	5081	2222	1052
2	936	1753	640	306
3	225	419	96	38
4	39	98	31	14

Покажете, че при ниво на съгласие $\alpha = 0,01$ хипотезата за зависимост на броя на децата в семейството от годишния доход на човек от семейството е невярна.

7. При социологическо проучване на цвета на косите и цвета на очите у младежи били получени следните резултати:

	Светли очи	Тъмни очи
Светли коси	9468	3364
Тъмни коси	3238	30472

Покажете, че съществува забележима зависимост между цвета на косите и цвета на очите ($\alpha = 0,01$).

8. В две групи болни е измерен с помощта на кардиограми интервалът между два удара на сърцето. Получени са резултатите:

$$\xi_1 = \ 9,6; \quad \xi_2 = 10,0; \quad \xi_3 = \ 9,8; \quad \xi_4 = 10,2; \quad \xi_5 = 10,6;$$
$$\eta_1 = 10,4; \quad \eta_2 = \ 9,7; \quad \eta_3 = 10,0; \quad \eta_4 = 10,3;$$

Като се предположи, че измерванията са нормално разпределени, покажете, че при ниво на съгласие $\alpha = 0,01$:

а) хипотезата $\mathbf{D}\xi = \mathbf{D}\eta$ може да се приеме;

б) хипотезата $\mathbf{E}\xi = \mathbf{E}\eta$ също може да се приеме.

9. Точността на работа на един автомат по стандарт е $\sigma^2 = 0,1$. Направена е контролна проверка с измервания, които дали следните резултати:

Намерена стойност	3,0	3,5	3,8	4,4	4,5
Честота	2	6	9	7	1

Покажете, че при ниво на съгласие $\alpha = 0,04$ автоматът не осигурява желаната точност и се нуждае от регулировка.

10. От две партиди изделия, изготвени на две еднакво настроени еднотипни машини A и B, са взети за проба съответно $n = 10$ и $m = 12$ изделия. Контролните резултати са дадени в следната таблица:

Измервания за машина А	ξ_i	3,4	3,5	3,7	3,9
Брой на изделията	n_i	2	3	4	1
Измервания за машина В	η_j	3,2	3,4	3,6	—
Брой на изделията	m_j	2	2	8	—

Предполага се, че ξ и η са нормално разпределени. При ниво на съгласие $\alpha = 0,02$ покажете, че:

а) хипотезата $\mathbf{D}\xi = \mathbf{D}\eta$ не може да се приеме;

б) хипотезата $\mathbf{E}\xi = \mathbf{E}\eta$ (при алтернатива $\mathbf{E}\xi \neq \mathbf{E}\eta$) не противоречи на резултатите от контролните измервания.

11. Контролният размер на произвежданите от един струг детайли е $a_0 = 35$ mm. От продукцията са взети по случаен начин 20 детайла, чиито измервания са:

Размер ξ_i	34,8	34,9	35,0	35,1	35,3
Честота n_i	2	3	4	6	5

При ниво на съгласие $\alpha = 0,05$ покажете, че стругът не работи с желаната точност.

Упътване. Хипотезата $H_0 : \mathbf{E}\xi = 35$ срещу алтернативата $H_1 : \mathbf{E}\xi \neq 35$ се отхвърля.

12. В две редици от $n = 10$ и $m = 18$ наблюдения над нормално разпределени случайни величини ξ и η са пресметнати дисперсиите $\widehat{s_\xi^2} = 1,23$; $\widehat{s_\eta^2} = 0,41$ на извадките. Покажете, че основната хипотеза $\mathbf{D}\xi = \mathbf{D}\eta$ при съответна конкурираща хипотеза H_1 и ниво на съгласие α:

а) се отхвърля, ако $\alpha = 0,1$;

б) се отхвърля, ако $\alpha = 0,02$;

в) се отхвърля, ако $\alpha = 0,05$;

13. При условията на зад. 12 и данни:

а) $n = 14$, $m = 10$; $\widehat{s}^2_\xi = 0,84$; $\widehat{s}^2_\eta = 2,52$; $\alpha = 0,1$ покажете, че H_0 се отхвърля;

б) при $n = 9$, $m = 6$; $\widehat{s}^2_\xi = 14,4$; $\widehat{s}^2_\eta = 20,5$; $\alpha = 0,1$; $H_1 : \mathbf{D}\xi \neq \mathbf{D}\eta$ хипотезата H_0 не се отхвърля;

в) при $\alpha = 0,1$; $H_1 : \mathbf{D}\xi \neq \mathbf{D}\eta$ и резултати

ξ_i	1,08	1,10	1,12	1,14	1,15	1,25	1,36	1,38	1,40	1,42
η_j	1,11	1,12	1,18	1,22	1,33	1,35	1,36	1,38	—	—

няма основание да се отхвърли H_0.

14. Случайните величини ξ и η са нормално разпределени. От тях са получени независими прости извадки с обеми n и m и са пресметнати емпиричните средни $\overline{\xi}_n$, $\overline{\eta}_m$ и дисперсиите $\widehat{s}^2_\xi$ и $\widehat{s}^2_\eta$. Покажете, че при проверката на основната хипотеза $H_0 : \mathbf{E}\xi = \mathbf{E}\eta$ срещу алтернативата $H_1 : \mathbf{E}\xi \neq \mathbf{E}\eta$ и при ниво на съгласие α са валидни следните твърдения:

а) ако $n = 30$, $m = 40$, $\overline{\xi}_n = 130$, $\overline{\eta}_m = 125$ и $\mathbf{D}\xi = 60$, $\mathbf{D}\eta = 80$, $\alpha = 0,05$, то H_0 се отхвърля;

б) ако $n = m = 50$, $\alpha = 0,05$, то H_0 не се отхвърля;

в) ако $n = 10$, $m = 8$, $\overline{\xi}_n = 142$, $\overline{\eta}_m = 145,3$; $\widehat{s}^2_\xi = 2,7$; $\widehat{s}^2_\eta = 3,2$; $\alpha = 0,01$, може да се покаже, че $\mathbf{D}\xi = \mathbf{D}\eta$, но хипотезата H_0 се отхвърля;

г) ако $n = 10$, $m = 16$, $\overline{\xi}_n = 12,35$; $\overline{\eta}_m = 12,35$, $\widehat{s}^2_\xi = 0,11$; $\widehat{s}^2_\eta = 0,07$; $\alpha = 0,05$, може да се докаже, че $\mathbf{D}\xi = \mathbf{D}\eta$ и H_0 се приема.

15. Като използвате критерия с отношение на правдоподобията (§ 11.5), докажете, че формула (1) от § 7 се получава при проверка на хипотезата $H_0 : a = a_0$, $\sigma^2 > 0$, срещу алтернативата $H_1 : a \neq a_0$ с ниво на съгласие α за извадка от нормално разпределена случайна величина $\xi \in \mathcal{N}(a, \sigma^2)$.

16. Докажете, че ако хипотезата H_0 в схемата на дисперсионния анализ от § 8 е вярна, то $\mathfrak{F}(\vec{\xi}) \xrightarrow{p} 1$ при $\min_j(n_j) \to \infty$.

Четиринадесета глава

Корелация и регресия

§ 1. Функционална, стохастическа, статистическа и корелационна зависимост

Връзките между различните явления в природата са сложни и многообразни, обаче те могат да се класифицират. В много практически задачи трябва да се установи и оцени зависимостта на изучаваната случайна величина от една или няколко други величини.

Зависимостта между ξ и η се нарича *функционална*, когато на всяка възможна стойност на ξ е съпоставена еднозначно определена стойност на η, т. е. $\eta = f(\xi)$. Такава е например зависимостта между налягането и обема на газа, изразена в закона на Бойл – Мариот.

В реалния свят много природни явления протичат под въздействие на голям брой фактори (причини), влиянието на всеки от които е нищожно. В този случай връзките губят своята строга функционалност. Под въздействието на такива фактори изучаваните процеси и системи преминават не в определено състояние, а в някое от възможните за тях състояния, което не може предварително да се установи. Вече стана дума за т. нар. стохастична връзка, т. е. такава връзка, при която една случайна величина реагира на изменение на друга (или на други) с изменение на своя закон на разпределение. В статистическите изследвания често се разглежда частен случай на такава връзка, наречена статистическа. Най-често за такава връзка може да се говори, когато условното математическо очакване на една случайна величина е функция на стойностите на друга случайна величина (§ 5.3).

Пример 1. Нека случайната величина ξ зависи от случайните фактори a_1, a_2, b_1, b_2, а случайната величина η — от случайните фактори a_1, a_2, c_1, c_2.

Тогава между ξ и η съществува *статистическа зависимост* поради това, че случайните фактори a_1 и a_2 са общи за ξ и η.

Статистическата зависимост между случайните величини (променливи) има голямо практическо значение: с нейна помощ могат да се прогнозират стойностите на зависимата случайна променлива при предположение, че независимата случайна променлива приема определени стойности.

За изучаването на статистическата зависимост трябва да се знае условното математическо очакване на предполагаемата зависима случайна променлива от всички останали величини. За неговото оценяване е необходимо да се познава аналитичният вид на съвместното разпределение на всички участващи величини. Най-често аналитичният вид на това разпределение трудно се открива и затова се прибягва до опростена схема на преход от условно математическо очакване на случайната променлива към условна средна стойност при фиксирани стойности на величините в условието. Например при две случайни величини това изглежда така:

$$\mathbf{E}(\eta \mid \xi = x) = \overline{y}(x), \tag{1}$$

където лявата страна на (1) е математическото очакване на случайната величина η при условие, че случайната величина ξ приема стойност x.

Ако при статистическата зависимост е важен само фактът, че изменението на една от случайните величини води до изменение на средната стойност на друга случайна величина, то в този случай се говори за *корелационна зависимост*.

Пример 2. Нека с ξ означим количеството употребени торове, а с η — добива на зърно. От еднакви земни участъци при едни и същи количества торове се събира различен добив, т. е. η не е функция от ξ. Това се обяснява с влиянието на случайните фактори: дъждове, температура на въздуха и др. Но заедно с това, както показват опитите, средният добив се явява функция на количеството употребени торове, т. е. η е свързана с ξ в корелационна зависимост.

Корелационната зависимост има широко приложение в изследователската работа. Тя характеризира вида и силата на връзките между случайните величини.

§ 2. Регресия. Регресионен анализ

А. Случай на две променливи. Ако между две случайни величини съществува някаква връзка, то една от най-важните общи задачи на статистиката е да се определят видът и силата на тази връзка. При изучаване на статистическата зависимост видът на връзката може да се характеризира с функцията на регресия (линейна, квадратична, показателна и т. н.).

Условното математическо очакване на случайната величина η при условие, че случайната величина ξ приема стойност x, разглеждано като функция на x, е

$$\mathbf{E}(\eta \,|\, \xi = x) = u(x) = \int_{-\infty}^{\infty} y dF(y \,|\, x), \tag{1}$$

и се нарича ***функция на регресията на η относно ξ*** (или функция на регресията на η по ξ). Аналогично

$$\mathbf{E}(\xi \,|\, \eta = y) = v(y) = \int_{-\infty}^{\infty} x dF(x \,|\, y)$$

се нарича функция на регресията на случайната величина ξ относно η.

Пример 1. Нека $f(x,y) = 3x+y$ е плътността на случайния вектор (ξ, η) при $0 \leqq x \leqq 1$ и $0 \leqq y \leqq 1$. Намерете функцията на регресия на η относно ξ.

Решение. За функцията на регресия $u(x)$ от (1) намираме

$$u(x) = \int_0^1 y f(y \,|\, \xi = x) dy,$$

където $f(y \,|\, \xi = x)$ е условната плътност на η при $\xi = x$, която се определя по следния начин:

$$f(y \,|\, \xi = x) = \frac{f(x,y)}{\int_0^1 f(x,y) dy} = \frac{3x + y}{\int_0^1 (3x + y) dy} = \frac{3x + y}{3x + \frac{1}{2}}.$$

Окончателно

$$u(x) = \int_0^1 y \frac{3x + y}{3x + \frac{1}{2}} dy = \frac{\frac{3}{2} + \frac{1}{3}}{3x + \frac{1}{2}}.$$

Намирането на $v(y)$ предоставяме на читателя.

Уравнението $y = u(x)$, в което x играе ролята на „независима променлива“, се нарича *уравнение на регресията*, а съответната му графика — *линия на регресията* на величината η относно ξ.

От начина, по който определихме $u(x)$ (или $v(y)$), става ясно, че функцията на регресия изразява математическото очакване на случайната величина η (или ξ) за случая, когато случайната величина ξ (или η) приема определена числена стойност, или иначе казано, $u(x)$ (или $v(y)$) показва каква е средната стойност на случайната величина η (или ξ), когато случайната величина ξ (или η) приема стойност x (или y).

Регресионните линии притежават следното забележително свойство: Измежду всички реални функции $g(x)$ минимумът на математическото очакване $\mathrm{E}[\eta - g(\xi)]^2$ се достига за функцията $g(x) = u(x)$. Ето защо регресията на η относно ξ дава най-доброто представяне на величината η чрез величината ξ. Това свойство се използва за прогнозиране на η относно ξ: ако стойностите на η непосредствено не се наблюдават (и не могат да се управляват), но експериментите (възможностите ни) позволяват да се наблюдава (управлява) величината ξ, то в качеството на прогнозирана стойност (очаквано значение) на η може да се използва величината $u(x)$, когато е известно че $\xi = x$.

Важно място при това прогнозиране заема условната дисперсия на величината η, пресметната за всяка фиксирана стойност $\xi = x$ (като мярка за точност):

$$\mathrm{D}(\eta \mid \xi = x) = \mathrm{E}(\eta^2 \mid \xi = x) - [\mathrm{E}(\eta \mid \xi = x)]^2 = \sigma^2(x).$$

Ако $\sigma^2(x) = 0$ при всяка допустима стойност на x, то може със сигурност да се твърди, че η и ξ са свързани със строга функционална зависимост $\eta = u(x)$. Ако $\sigma^2(x) \neq 0$ при всяко x и $u(x)$ не зависи от x, казва се, че няма регресия на η относно ξ.

Пример 2. Като използвате пример 1, изчислете дисперсията на условното разпределение

$$\begin{aligned}\sigma^2(x) &= \mathrm{D}(\eta \mid \xi = x) = \mathrm{E}(\eta^2 \mid \xi = x) - [\mathrm{E}(\eta \mid \xi = x)]^2 \\ &= \int_0^1 y^2 \frac{3x + y}{3x + 0,5}\, dy - \left(\frac{1,5x + 0,5}{3x + 0,5}\right)^2 \\ &= \frac{x + 0,25}{3x + 0,5} - \frac{2,25x^2 + x + \frac{1}{9}}{(3x + 0,5)^2} = \frac{0,75x^2 + 0,25x + \frac{1}{72}}{(3x + 0,5)^2}.\end{aligned}$$

Решение. При $\xi = x = 1$ получаваме

$$\sigma^2(1) = \mathbf{D}(\eta \mid \xi = 1) = \frac{73}{882} = 0,083.$$

Последният резултат показва, че ако $\xi = 1$, то наблюдаваните стойности на η се отклоняват от своето математическо очакване (т. е. от функцията на регресия) в средно–квадратичен смисъл с 0,083.

В дадения случай $\sigma^2(x)$ зависи от стойностите на ξ, но това не може да бъде обща закономерност. Така например, ако ξ и η имат двумерно нормално разпределение, то $\sigma^2(x)$ е константа.

Тук ще отбележим, че практическото приложение на функцията на регресия е твърде ограничено. Това произтича от факта, че обикновено не се познава аналитичният вид на разпределението на случайния вектор (ξ, η). Само ако познаваме вида на това разпределение, можем точно да определим вида на функцията на регресия и след това да оценим неговите параметри. Обаче за подобни оценки в практиката най-често разполагаме само с извадка с ограничен обем, по която трябва де се намери видът на двумерното разпределение $f(x, y)$, а след това видът на функцията на регресия. Това може да доведе до значителни грешки, защото една и съща съвкупност от точки (x_i, y_i) в равнината може еднакво добре да се опише с помощта на различни функции. Именно затова са и ограничени възможностите за практическото приложение на функцията на регресия.

Най-прост е случаят, когато регресията на η относно ξ е линейна, т. е.

$$\mathbf{E}(\eta \mid \xi = x) = \beta_0 + \beta_1 x, \tag{2}$$

където β_0 и β_1 са ***коефициенти на регресията***. Коефициентите на регресия се определят с равенствата

$$\beta_0 = \overline{y} - \rho \frac{\sigma_y}{\sigma_x} \overline{x}; \quad \beta_1 = \rho \frac{\sigma_y}{\sigma_x},$$

където $\overline{x}$ и $\overline{y}$ са математическите очаквания на ξ и η; σ_x^2, σ_y^2 — техните дисперсии, а ρ е коефициентът на корелация между ξ и η.

Уравнението на регресията се записва с израза

$$y = \beta_0 + \beta_1 x = \overline{y} + \rho \frac{\sigma_y}{\sigma_x} (x - \overline{x}).$$

В случай, че съвместното разпределение на ξ и η е нормално, двете линии $y = u(x)$ и $x = v(y)$ са прави. На това се дължи голямата популярност на линейните регресионни модели в практиката.

Ако регресията на η относно ξ не е линейна, то представянето (2) е само една линейна апроксимация на истинското уравнение на регресията: Математическото очакване $\mathbf{E}(\eta - b_0 - b_1\xi)^2$ достига минимум по b_0 и b_1 само когато $b_0 = \beta_0$; $b_1 = \beta_1$.

Много често се срещат случаи на регресионни уравнения, изразени като линейни комбинации от зададени функции на x:

$$y = h(x) = \beta_0\varphi_0(x) + \beta_1\varphi_1(x) + \cdots + \beta_m\varphi_m(x). \tag{3}$$

Важно значение има полиномната регресия, при която

$$\varphi_0(x) = 1; \quad \varphi_1(x) = x, \ \ldots, \ \varphi_m(x) = x^m.$$

Б. Понятието регресия е приложимо не само към случайни величини, но и към случайни вектори. В частност, ако η е случайна величина, а $\vec{\xi} = (\xi_1, \ldots, \xi_k)$ — случаен вектор, то регресията на η относно $\vec{\xi}$ се определя с уравнението

$$y = u(x_1, \ldots, x_k) = \mathbf{E}(\eta \,|\, \xi_1 = x_1, \ldots, \xi_k = x_k).$$

Ако имаме представянето

$$u(x_1, \ldots, x_k) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k, \tag{4}$$

то регресията се нарича *линейна*. Такова представяне има например, когато $\xi_1, \ldots, \xi_k, \eta$ имат съвместно нормално разпределение.

Прост пример за регресия на η относно ξ е зависимостта, изразена с равенството $\eta = u(\xi) + \delta$, където $u(\xi) = \mathbf{E}(\eta \,|\, \xi = x)$, а величините ξ и δ са независими. Това представяне е полезно, когато се планира експеримент за изучаване функционалната връзка между неслучайни величини y и x, а измерванията на y и x не са точни.

На практика обикновено коефициентите на регресията са неизвестни и те биват оценявани по експериментални данни. Техниката на това оценяване, изчислителните процедури, изводите и следствията от получените резултати са обект на изследване на един обширен раздел на статистическия анализ, наречен *регресионен анализ.*

Целта на регресионния анализ се състои в определяне на общия вид на регресионните уравнения от даден клас функции, в построяване на оценки за неизвестните параметри в тях, участващи в регресионното уравнение, и в проверка на статистически хипотези относно адекватността на предлаганата регресия. Изборът на регресионния модел се определя от предположенията относно вида на зависимостта $y = g(x_1, \ldots, x_k)$ на зависимата величина y от променливите $x_1, \ldots, x_k$.

Най-естествени от гледна точка на методологията за определяне на неизвестните регресионни коефициенти са линейните модели (2), (3), (4).

Да предположим, че регресионната зависимост на величината η от величините $\xi_1, \ldots, \xi_k$ е от вида

$$\eta = u(\xi_1, \ldots, \xi_k) = \beta_0 + \beta_1\xi_1 + \cdots + \beta_k\xi_k.$$

Нека в резултат на проведените експерименти са получени наблюденията

$$\left(y_i, x_1^{(i)}, x_2^{(i)}, \ldots, x_k^{(i)}\right); \quad i = 1, \ldots, n.$$

Оценките $\widehat{\beta}_0, \widehat{\beta}_1, \ldots, \widehat{\beta}_k$ на регресионните коефициенти по метода на най-малките квадрати се получават като решения на системата уравнения (известни като „система нормални уравнения за регресионните коефициенти“)

$$\sum_{i=1}^{n}\left(y_i - \sum_{m=0}^{k}\widehat{\beta}_m x_m^{(i)}\right)x_l^{(i)} = 0, \quad l = 0, 1, \ldots, k; \; x_0^{(i)} = 1.$$

Ако положим

$$C_{ml=l} = \sum_{i=1}^{n} x_l^{(i)} x_m^{(i)}; \quad Z_l = \sum_{i=1}^{n} y_i x_l^{(i)},$$

то оценките $\widehat{\beta}_m$ се пресмятат по формулите

$$\widehat{\beta}_m = \sum_{l=0}^{k} C_{ml}^{-1} Z_l.$$

Тук C_{ml}^{-1} са елементите на обратната матрица $\mathbf{C}^{-1}$ на матрицата $\mathbf{C} = (C_{ml})$. Матрицата $\mathbf{C}$ е известна под името „информационна матрица на Фишер“.

Оценките $\widehat{\beta}_m$ са неизместени оценки за регресионните коефициенти β_m. Когато величините $(\eta, \xi_1, \ldots, \xi_k)$ са нормално разпределени, могат да се построят доверителни интервали за истинските коефициенти на регресията $\beta_0, \beta_1, \ldots, \beta_k$. Ако се окаже, че доверителният интервал за някой коефициент β_m съдържа нулата, този коефициент β_m се приема за *незначим* (величината ξ_m не влияе върху η) и членът $\beta_m\xi_m$ в записа на регресията може да бъде изпуснат.

Проверката на хипотезата за наличие на регресионна връзка

$$\mathbf{E}(\eta \mid \vec{\xi} = \vec{x}) = \widehat{\beta}_0 + \widehat{\beta}_1 x_1 + \cdots + \widehat{\beta}_k x_k$$

се провежда с помощта на критерия на Стюдънт. Пресмятат се числата

$$\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_1^{(i)} + \cdots + \widehat{\beta}_k x_k^{(i)}$$

и се проверява дали наблюденията $y_1, \ldots, y_k$ и числата $\widehat{y}_1, \ldots, \widehat{y}_k$ са от една и съща генерална съвкупност. Съществуват и по-силни методи, използващи също критерия на Стюдънт.

Регресионният анализ е един от най-разпространените методи за обработка на резултати от наблюдения при изучаване на зависимости между величини в биологията, физиката, икономиката, техниката и други области. На неговите модели са основани и други раздели от математическата статистика като дисперсионния анализ, планирането на експеримента, многомерния статистически анализ.

Пример 1. Намерете емпиричното уравнение на регресия по наблюденията

$\xi = x$	1,00	2,00	3,00	4,00	5,00
$\eta = y$	1,25	1,50	1,75	2,00	2,50

Решение. От (1) за $\widehat{\beta}_0$, $\widehat{\beta}_1$ намираме $\widehat{\beta}_0 = 0,90$; $\widehat{\beta}_1 = 0,30$ и следователно $y = 0,90 + 0,30x$.

§ 3. Корелация. Корелационен анализ

Корелацията е термин, използван в различни области на науката и техниката за означаване на взаимна зависимост, взаимно съответствие, съотношения между понятия, величини, предмети, процеси, събития. В математическата статистика корелацията е вероятностна или статистическа зависимост, нямаща строго функционален характер. За разлика от функционалната корелационната зависимост възниква тогава, когато една от величините зависи не само от дадена величина, но и от ред други, неотчетени случайни фактори или когато сред условията, от които зависят едната и другата величина, има общи условия и за двете.

Статистическата корелация е интересна, когато е признак за съществуване на закономерна обективна връзка между изучаваните явления.

В основата на теорията на корелацията е предположението, че изучаваните явления са подчинени на определени вероятностни закономерности. Зависимостта между две случайни събития се проявява в това, че условната вероятност на едното събитие при настъпване на другото се различава от безусловната. Например в две кооперирани предприятия вероятността едното да си изпълни годишния план зависи от условието, дали другото предприятие си е изпълнило плана.

А. Мерки за стохастична зависимост. Едни от важните и прости характеристики на степента на зависимост между две случайни величини са *корелационният момент* (ковариацията на тези величини) и техният *коефициент на корелация*. Ще припомним накратко определенията и свойствата на тези характеристики.

За случайните величини ξ и η смесеният момент от втори ред се нарича ковариация и се бележи така:

$$\operatorname{cov}(\xi,\eta) = \mathbf{E}[(\xi - \mathbf{E}\xi)(\eta - \mathbf{E}\eta)].$$

Коефициентът на корелация $\rho(\xi,\eta)$ на случайните величини ξ и η се определя с израза

$$\rho(\xi,\eta) = \frac{\operatorname{cov}(\xi,\eta)}{\sqrt{\mathbf{D}\xi}\sqrt{\mathbf{D}\eta}} = \mathbf{E}\left(\frac{\xi - \mathbf{E}\xi}{\sqrt{\mathbf{D}\xi}} \cdot \frac{\eta - \mathbf{E}\eta}{\sqrt{\mathbf{D}\eta}}\right), \tag{1}$$

където $\mathbf{D}\xi > 0$, $\mathbf{D}\eta > 0$.

За $\operatorname{cov}(\xi,\eta)$ е вярно неравенството

$$|\operatorname{cov}(\xi,\eta)| \leqq \sqrt{\mathbf{D}\xi}\sqrt{\mathbf{D}\eta}, \tag{2}$$

което е частен случай на неравенството на Коши – Шварц.

Съответно за коефициента на корелация е валидно неравенството

$$|\rho(\xi,\eta)| \leqq 1. \tag{3}$$

Коефициентът на корелация ρ е ± 1 само когато между ξ и η има строго функционална линейна зависимост. Ако $\rho = \pm 1$, т. е. $|\rho| = 1$, в (2) има равенство, а това означава, че $\eta - \mathbf{E}\eta = k(\xi - \mathbf{E}\xi)$, т. е. η е линейна функция на ξ от вида $\eta = a\xi + b$. При това: $a < 0$

при $\rho(\xi,\eta) = -1$ и $a > 0$ при $\rho(\xi,\eta) = 1$. Знакът на $\rho(\xi,\eta)$ показва посоката на зависимостта. Положителният знак показва, че на по-голяма стойност на ξ отговарят по-големи стойности на η. При отрицателен знак на големите стойности на ξ отговарят по-малки стойности на η, т. е. когато ξ расте, η намалява и обратно.

Основните свойства на $\rho(\xi,\eta)$ са тясно свързани с най-добрата линейна в средно–квадратичен смисъл регресия на η по ξ и на ξ по η. Ако положим

$$l_{10} : \frac{y - \mathbf{E}\eta}{\sigma_\eta} = \rho \frac{x - \mathbf{E}\xi}{\sigma_\xi}; \qquad l_1 : y(x) = ax + b,$$

$$l_{20} : \frac{y - \mathbf{E}\eta}{\sigma_\eta} = \frac{1}{\rho} \cdot \frac{x - \mathbf{E}\xi}{\sigma_\xi}; \quad l_2 : x(y) = cy + d$$

за двете линии на регресия, то минимумите на изразите

$$\min_{l_1} \mathbf{E}[\eta - y(\xi)]^2, \quad \min_{l_2} \mathbf{E}[\xi - x(\eta)]^2 \tag{4}$$

се достигат за правите $l_1 = l_{10}$, $l_2 = l_{20}$ и са съответно

$$\sigma_\eta^2(1 - \rho^2), \quad \sigma_\xi^2(1 - \rho^2). \tag{5}$$

Затова изразите (5) се наричат *остатъчни дисперсии от регресията*.

В зависимост от ρ двете случайни величини ξ, η биват:

1) при $\rho = 0$ — некорелирани;

2) при $\rho \neq 0$ — корелирани, като при $\rho > 0$ корелацията е положителна, а при $\rho < 0$ тя е отрицателна;

3) ако $\rho = \pm 1$, то между тях има строго функционална зависимост $\eta = a\xi + b$ и остатъчната дисперсия е нула.

Става ясно, че коефициента на корелация ρ можем да разглеждаме като мярка за линейност на връзката между координатите на случайния вектор (ξ,η).

Тук ще отбележим, че две независими случайни величини са винаги некорелирани, но обратното не е вярно: две некорелирани случайни величини не е задължително да бъдат независими.

За всеки две случайни величини ξ и η дисперсията се представя така:

$$\sigma_\eta^2 = \mathbf{E}[\eta - \mathbf{E}\eta]^2 = \mathbf{E}[\eta - m(\xi)]^2 + \mathbf{E}[m(\xi) - \mathbf{E}\eta]^2, \tag{6}$$

където $m(\xi) = \mathbf{E}(\eta \,|\, \xi)$. Тогава величината

$$\theta_{\eta\xi}^2 = \frac{\mathbf{E}[m(\xi) - \mathbf{E}\eta]^2}{\sigma_\eta^2}$$

се нарича *корелационно отношение* на η по ξ, въведено от К. Пирсън. От (6) получаваме

$$1 - \theta^2_{\eta\xi} = \frac{1}{\sigma^2_\eta} \mathbf{E}[\eta - m(\xi)]^2 \geqq 0$$

и следователно винаги $0 \leqq \theta^2_{\eta\xi} \leqq 1$.

Съществува представянето

$$\theta^2_{\eta\xi} = \rho^2 + \frac{1}{\sigma^2_\eta} \mathbf{E}[m(\xi) - \alpha - \beta\xi]^2.$$

Оттук $\theta^2_{\eta\xi} = 0$ само когато $m(x)$ не зависи от x. Наистина, ако $m(x)$ е константа, кривата на регресия $y = m(x)$ е хоризонтална права и следователно $\rho = \beta_1 = 0$ и $\theta^2_{\eta\xi} = 0$.

Корелационното отношение $\theta^2_{\eta\xi}$ е 1 само когато съществува тясна функционална зависимост между ξ и η, т. е. $\eta = m(\xi) = \mathbf{E}(\eta \mid \xi)$. Корелационното отношение може да се разглежда като мярка за стремежа на масата на разпределението на (ξ, η) да се натрупа около линията на регресия. Корелационното отношение $\theta^2_{\xi\eta}$ е ρ^2 само когато $m(\xi)$ е права линия $a\xi + b$. В такъв случай пресмятането на корелационното отношение не дава нищо ново, ако вече е известно ρ^2.

При нелинейна регресия винаги е изпълнено $\theta^2_{\eta\xi} > \rho^2_{\xi\eta}$. Разликата $\delta^2 = \theta^2_{\eta\xi} - \rho^2_{\xi\eta}$ характеризира отклонението на кривата $y = m(x)$ от правата, даваща най-добра линейна регресия.

За дискретни разпределения корелационното отношение може да се пресметне с израза

$$\theta^2_{\eta\xi} = \frac{1}{\sigma^2_\eta} \mathbf{E}[m(\xi) - \mathbf{E}\eta]^2 = \frac{1}{\sigma^2_\eta} \sum_i [m_i - \mathbf{E}\eta]^2 p_{i\cdot},$$

където

$$p_{i\cdot} = \sum_j p_{ij}; \quad m_i = \mathbf{E}(\eta \mid \xi = x_i) = \frac{\sum_j y_j p_{ij}}{\sum_j p_{ij}}.$$

Коефициентът на корелация и корелационното отношение служат за характеристики на степента на зависимост между две величини.

В практиката е прието да се говори [7] за следните степени на корелационна зависимост:

$$
\begin{array}{ll}
0 \leqq \theta^2_{\eta\xi} < 0,3 & \text{– слаба,} \\
0,3 \leqq \theta^2_{\eta\xi} < 0,5 & \text{– умерена,} \\
0,5 \leqq \theta^2_{\eta\xi} < 0,7 & \text{– значима,} \\
0,7 \leqq \theta^2_{\eta\xi} < 0,9 & \text{– силна,} \\
0,9 \leqq \theta^2_{\eta\xi} & \text{– много силна.}
\end{array}
$$

Такава класификация на зависимостта обаче може да се използва само когато свободният коефициент на корелация е значим (вж. § 13.5).

Степента на зависимост може да се характеризира и с въведения от К. Пирсън коефициент на средно-квадратична свързаност, който в случай на дискретно двумерно разпределение се определя с израза

$$
\varphi^2 = \sum_{i,j} \frac{(p_{ij} - p_{i\bullet}p_{\bullet j})^2}{p_{i\bullet}p_{\bullet j}} \quad \text{при} \quad p_{i\bullet} = \sum_j p_{ij}; \; p_{\bullet j} = \sum_i p_{ij}.
$$

Ако случайните величини ξ и η са независими, то $\varphi^2 = 0$. Вярно е и обратното.

Пример. Нека дискретните случайни величини ξ и η приемат само две възможни стойности

$\eta = y$ / $\xi = x$	y_1	y_2
x_1	p_{11}	p_{12}
x_2	p_{21}	p_{22}

Тогава

$$
\varphi^2 = \frac{(p_{11}p_{22} - p_{12}p_{21})^2}{p_{1\bullet}p_{2\bullet}p_{\bullet 1}p_{\bullet 2}} = \rho^2.
$$

Максималната стойност на φ^2 е 1 и се получава в два случая: $p_{12} = p_{21} = 0$ или $p_{11} = p_{22} = 0$ (вж. и § 13.8).

Б. В статистиката са разработени методи за оценка на споменатите по-горе коефициенти и методи за проверка на техните стойности, използващи като материал статистически наблюдения върху отделните случайни величини. Тези методи са предмет на изследване в т. нар. *корелационен анализ.*

Приложението на корелационния анализ с експериментални данни се състои в следните дейности: 1) построяване на корелационно поле и съставяне на корелационни таблици; 2) пресмятане на емпиричните коефициенти на корелация или на корелационното отношение; 3) проверка на статистически хипотези за значимостта на връзката. По-нататъшното изследване се изразява в установяване на конкретния вид на зависимостта между величините и е предмет на регресионния анализ. Зависимости между три и повече случайни признаци, величини или фактори се изучават с методите на многомерния корелационен анализ.

Корелационното поле и корелационната таблица (вж. § 13.5) са спомагателни средства за анализ на емпирични данни. При нанасяне в координатна равнина на двойките наблюдения (ξ_i, η_i) като точки се получава корелационното поле. По неговата форма и разположението на точките може да се състави мнение за вида на евентуалната зависимост между величините (линейна, параболична и пр.). За числена обработка резултатите се групират и представят в корелационна таблица (вж. § 13.5). Всяка нейна клетка посочва броя n_{ij} на онези двойки (ξ_i, η_i), чиито компоненти попадат в съответните интервали на групиране по всяка от променливите (в i-тия интервал за ξ и j-тия интервал за η).

Обикновено дължините на интервалите на групиране по всяка променлива са равни помежду си. Нека центровете им са x_i, y_j. Числата n_{ij} се считат за брой на случаите, при които двойката (ξ, η) приема стойност (x_i, y_j).

Коефициентът на корелация и корелационното отношение дават информация за характера и силата на връзката. Емпиричният коефициент на корелация се смята по формулата

$$\widehat{\rho}^2_{\xi\eta} = \frac{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n}(x_i - \overline{x})(y_i - \overline{y})n_{ij}}{\sqrt{\sum\limits_{i=1}^{n} n_{i\bullet}(x_i - \overline{x})^2 \sum\limits_{j=1}^{n} n_{\bullet j}(y_j - \overline{y})^2}}.$$

Тук са използвани означенията

$$n_{i\bullet} = \sum_{j=1}^{n} n_{ij}; \qquad n_{\bullet j} = \sum_{i=1}^{n} n_{ij};$$

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i n_{i\bullet}; \quad \overline{y} = \frac{1}{n}\sum_{j=1}^{n} y_j n_{\bullet j}.$$

Оценката $\widehat{\rho}^2_{\xi\eta}$ е състоятелна оценка за истинския коефициент на корелация.

Във всички случаи препоръчително е за характеристика на силата на връзката между величините ξ и η да се използва корелационното отношение. Неговата емпирична оценка се дава с израза

$$\widehat{\theta}^2_{\eta/\xi} = \frac{1}{n} \cdot \frac{\sum\limits_{i=1}^{n} n_{i\bullet}(\overline{y}_i - \overline{y})^2}{\frac{1}{n}\sum\limits_{j=1}^{n} n_{\bullet j}(y_j - \overline{y})^2}.$$

Тук числителят характеризира разсейването на условното средно $\overline{y}_i = \frac{1}{n_{i\bullet}} \sum\limits_{j=1}^{n} n_{ij} y_j$ около безусловното средно $\overline{y}$. Величината $\widehat{\theta}^2_{\eta/\xi} - \widehat{\rho}^2_{\xi\eta}$ се използва като мярка за отклонението на зависимостта между ξ и η от линейната.

Проверката на хипотези за значимостта на връзката се базира върху знанията за разпределението на емпиричните оценки. В случай на нормално разпределени ξ и η емпиричният коефициент на корелация се счита за значимо различен от нула, ако е изпълнено неравенството

$$\widehat{\rho}^2_{\xi\eta} > \left[1 + \frac{n-2}{t^2_\alpha}\right]^{-1},$$

където t_α е критичната стойност на t-разпределението на Стюдънт с $n-2$ степени на свобода, съответна на избрано ниво на съгласие α. Ако е известно, че $\rho \neq 0$, необходимо е да се използва F-трансформацията на Фишер:

$$F = \frac{1}{2} \ln \frac{1 + \widehat{\rho}_{\xi\eta}}{1 - \widehat{\rho}_{\xi\eta}}.$$

При големи n величината F е приблизително нормално разпределена. С помощта на квантилите $F_{\frac{\alpha}{2}}$ и $F_{1-\frac{\alpha}{2}}$ може да се построят доверителни интервали за истинския коефициент на корелация ρ.

§ 4. Установяване на зависимости между съвкупности от случайни величини

Разглеждаме n случайни величини $\xi_1, \ldots, \xi_n$, имащи разпределение от непрекъснат тип. Условната средна стойност на величината ξ_1 при условие, че $\xi_i = x_i$, $i = 2, \ldots, n$, е

$$\mathbf{E}(\xi_1 \,|\xi_2 = x_2, \ldots, \xi_n = x_n) = u(x_2, \ldots, x_n) = \frac{\int\limits_{-\infty}^{\infty} x_1 f(x_1, \ldots, x_n) dx_1}{\int\limits_{-\infty}^{\infty} f(x_1, \ldots, x_n) dx_1}.$$

Геометричното място на точките $(u, x_2, \ldots, x_n)$ се нарича *регресионна повърхнина* за средната стойност на ξ_1 при фиксирани стойности $x_2, \ldots, x_n$ на останалите случайни величини и има уравнение $x_1 = u(x_2, \ldots, x_n)$.

Равнината на средно–квадратичната регресия на величината ξ_1 относно $\xi_2, \ldots, \xi_n$ се определя като хиперравнина

$$\xi_1 = \beta_{12\bullet 34\ldots n}\xi_2 + \beta_{13\bullet 24\ldots n}\xi_3 + \cdots + \beta_{1n\bullet 23\ldots(n-1)}\xi_n,$$

която дава най-добра апроксимация на n-мерната регресионна повърхнина $u(x_2, \ldots, x_n)$ в смисъл, че величината

$$\text{(1)} \qquad \mathbf{E}[\xi_1 - \beta_{12\bullet 34\ldots n}\xi_2 - \cdots - \beta_{1n\bullet 23\ldots(n-1)}\xi_n]^2$$

има най-малката възможна стойност.

За да определим коефициента на регресия $\vec{\beta}$, диференцираме (1) по всеки от неизвестните коефициенти β_{1i} ($n - 1$ на брой) и получаваме $n - 1$ уравнения

$$\text{(2)} \qquad \left|\begin{array}{l} \lambda_{22}\beta_{12} + \lambda_{23}\beta_{13} + \cdots + \lambda_{2n}\beta_{1n} = \lambda_{21}, \\ \lambda_{32}\beta_{12} + \lambda_{33}\beta_{13} + \cdots + \lambda_{3n}\beta_{1n} = \lambda_{31}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \lambda_{n2}\beta_{12} + \lambda_{n3}\beta_{13} + \cdots + \lambda_{nn}\beta_{1n} = \lambda_{n1}. \end{array}\right.$$

С β_{1j} сме означили за краткост $\beta_{1j\bullet 3\ldots(j-1)(j+1)\ldots n}$.

Нека $\Lambda = (\Lambda_{ij})$ е детерминантата на (2), а Λ_{ij} — адюнгираното количество на β_{ij}. Тогава, ако Λ е положително дефинитна, единственото решение на (2) се записва така:

$$\text{(3)} \qquad \widehat{\beta}_{1j} = -\frac{\Lambda_{1j}}{\Lambda_{11}} = -\frac{\sigma_{\xi_1}}{\sigma_{\xi_j}} \cdot \frac{p_{1j}}{p_{11}}; \quad \widehat{\beta}_{ij} = -\frac{\sigma_{\xi_i}}{\sigma_{\xi_j}} \cdot \frac{p_{ij}}{p_{ii}},$$

където $\mathbf{P}$ е матрицата от коефициентите на корелация.

Равнината на регресия на всяка случайна величина относно останалите е определена еднозначно и коефициентите на регресия се определят по (3).

Разликата

$$\varepsilon_{12\ldots n} = \xi_1 - \widehat{\beta}_{12}\xi_2 - \cdots - \widehat{\beta}_{1n}\xi_n = \frac{1}{\Lambda_{11}} \sum_{j=1}^{n} \Lambda_{1j}\xi_j,$$

където коефициентите на регресия $\widehat{\beta}_{1j}$ се определят от (3), се нарича остатък на величината ξ_1 от регресионната равнина спрямо $\xi_2, \ldots, \xi_n$. Този остатък е некорелиран с $\xi_2, \ldots, \xi_n$.

Остатъчната дисперсия $\sigma^2_{1\bullet 23\ldots n}$ се определя от израза

$$\sigma^2_{1\bullet 23\ldots n} = \frac{|\Lambda|}{\Lambda_{11}} = \sigma_1^2 \frac{|\mathbf{P}|}{p_{11}} \tag{4}$$

и се разглежда като мярка за приближение на представянето на ξ_1 чрез най-добра линейна комбинация на величините $\xi_2, \ldots, \xi_n$. При $n = 2$ равенството (4) се свежда към $\sigma^2_{1\bullet 2} = \sigma^2_{\xi}(1 - \rho^2)$.

Коефициентът на корелация между остатъците на ξ_1 и ξ_2 след изваждане от тях на най-добрите линейни комбинации от $\xi_3, \xi_4 \ldots, \xi_n$, (използва се терминът „след отстраняване на влиянието на случайните величини $\xi_3, \ldots, \xi_n$“) се нарича *коефициент на частна корелация* и се означава с $\rho_{12\bullet 34\ldots n}$:

$$\rho_{12\bullet 34\ldots n} = \frac{-\Lambda_{12}}{\sqrt{\Lambda_{11}\Lambda_{22}}} = \frac{-p_{12}}{\sqrt{p_{11}p_{22}}}. \tag{5}$$

Всеки коефициент на корелация може да бъде изразен чрез пълните коефициенти на частна корелация. Така например при $n = 3$ имаме

$$\rho_{12\bullet 3} = \frac{\rho_{11}\rho_{33} - \rho_{13}\rho_{23}}{\sqrt{(1 - \rho_{13}^2)(1 - \rho_{23}^2)}}.$$

В частния случай при некорелирани величини от (5) следва, че всички частни коефициенти са равни на нула.

Ако пълните коефициенти на корелация са известни, частните коефициенти могат да бъдат изчислени от (5). За пресмятането се използват рекурентните съотношения

$$\rho_{12\bullet 34\ldots n} = \frac{\rho_{12\bullet 34\ldots n-1} - \rho_{1n\bullet 34\ldots(n-1)}\rho_{2n\bullet 34\ldots(n-1)}}{\sqrt{\left(1 - \rho^2_{1n\bullet 34\ldots(n-1)}\right)\left(1 - \rho^2_{2n\bullet 34\ldots(n-1)}\right)}}.$$

Остатъчните дисперсии и частните коефициенти на регресия са свързани с изразите

$$\sigma^2_{1\bullet 23\ldots n} = \sigma^2_1(1-\rho^2_{12})(1-\rho^2_{13\bullet 2})\cdots(1-\rho^2_{1n\bullet 23\ldots(n-1)}),$$

$$\beta_{12\bullet 34\ldots n} = \rho_{12\bullet 34\ldots n}\frac{\rho_{1\bullet 34\ldots n}}{\rho_{2\bullet 34\ldots n}},$$

$$\rho^2_{12\bullet 34\ldots n} = \beta_{12\bullet 34\ldots n}\beta_{21\bullet 34\ldots n}.$$

Да разгледаме остатъка

$$\eta_{1\bullet 23\ldots n} = \xi_1 - \widehat{\beta}_{12}\xi_2 - \cdots - \widehat{\beta}_{1n}\xi_n = \xi_1 - \xi_1^*,$$

където $\xi_1^* = \widehat{\beta}_{12}\xi_2 + \cdots + \widehat{\beta}_{1n}\xi_n$ е най-добрата линейна оценка на ξ_1. Коефициентът на корелация между ξ_1 и величината ξ_1^* означен с $\rho_{1(23\ldots n)}$, се нарича ***свободен коефициент на корелация*** и се пресмята по формулата

$$\rho_{1(23\ldots n)} = \frac{\mathbf{E}(\xi_1\xi_1^*)}{\sqrt{\mathbf{E}\xi_1^2}\sqrt{\mathbf{E}(\xi_1^*)^2}} = \sqrt{1-\frac{|\Lambda|}{\lambda_{11}\Lambda_{11}}} = \sqrt{1-\frac{|\mathbf{P}|}{p_{11}}}. \tag{6}$$

При $\rho_{1(23\ldots n)} = 1$ случайната величина ξ_1 е равна на линейна комбинация от $\xi_2, \ldots, \xi_n$, а при $\rho_{1(23\ldots n)} = 0$, т. е. при $\rho_{12} = \cdots = \rho_{1n} = 0$, случайната величина ξ_1 е некорелирана с всички величини $\xi_2, \ldots, \xi_n$. При числено пресмятане на (6) е удобно да се използва съотношението

$$\rho^2_{1(23\ldots n)} = 1 - \frac{\sigma^2_{1\bullet 23\ldots n}}{\sigma^2_1}.$$

§ 5. Рангова корелация

В практиката често се срещат признаци, които не се поддават на количествена оценка. Така например, ако се опитаме да оценим съотношението между математическите и музикалните способности на група учащи се, то „нивото на способностите“ е променлива величина, в смисъл че варира от един индивид към друг. То може да се измери, ако поставим на всеки индивид оценки, обаче този начин е лишен от обективност, тъй като различни изпитващи могат да поставят на един и същ индивид различни оценки. Субективният елемент може да се изключи, ако подредим учащите по степените на техните способности и дадем на всеки от тях пореден номер, който ще наречем „ранг“. Корелацията между ранговете много по-точно отразява съотношението между способностите на учащите се, отколкото корелацията между оценките.

Нека n индивида имат по качеството А рангове: $x_1, \ldots, x_n$, а по качеството Б — рангове $y_1, \ldots, y_n$. Да означим разликата $x_k - y_k = d_k$. Стойността на d_k ни дава мярката за близост на съответствието между А и Б. Ако всички d_k са равни на нула, то съответствието е пълно.

Коефициентът на рангова корелация R, предложен от Спирмън, се пресмята по формулата

$$R = 1 - \frac{6\sum\limits_{k=1}^{n} d_k^2}{n(n^2-1)}.$$

Той се изменя от $+1$ до -1. Ако всички стойности на d_k са равни на нула, то $R = +1$. Ако ранговете са такива, че на $1, 2, \ldots, n$ по единия признак съответстват $n, n-1, \ldots, 1$ по другия признак, то $R = -1$.

Пример 1. Проверете има ли корелационна връзка между средните добиви от единица площ и фонда „техническа въоръженост“, взети за 10 държави.

Добив x_k	5	8	1	6	3	4	4	10	2	7	$\sum x_k = 55$
Фонд y_k	3	7	2	5	1	9	6	8	4	10	$\sum y_k = 55$
d_k	2	1	−1	1	2	0	−2	2	−2	−3	$\sum d_k = 0$
d_k^2	4	1	1	1	4	0	4	4	4	9	$\sum d_k^2 = 32$

За коефициентите на корелация имаме

$$R = 1 - \frac{6.32}{10.(100-1)} = 0,81.$$

По такъв начин можем да направим извода, че връзката между средните добиви и фонда „техническа въоръженост“ съществува и тя е достатъчно тясна.

Нека имаме n индивида, подредени по два качествени признака. Нека номерата на първата редица са от 1 до n и под всеки от тях напишем номерата от втората редица, т. е.

$$(1) \qquad \left[\begin{array}{llll} \xi_1 = 1, & \xi_2 = 2, & \dots, & \xi_n = n \\ \eta_1, & \eta_2, & \dots, & \eta_n. \end{array}\right.$$

Тогава за всяко η_k, $k = 1, 2, \dots, n$, да определим величината a_k, равна на броя на тези ξ_i, стоящи надясно от η_k и по-големи от η_k (инверсиите на η_k с членовете на редицата (ξ), вж. § 13.3).

Ако означим с $a = \sum\limits_{k=1}^{n} a_k$, то $\tau = 1 - \dfrac{4a}{n(n+1)}$ се нарича *рангов коефициент на корелация*, предложен от Кендал. Стойностите на τ принадлежат на интервала $[-1, 1]$. При това $\tau = 0$ само когато поредните номера на редиците от (1) съвпадат (разликата от поредните номера на всеки индивид е нула) и $\tau = 1$ само когато всеки индивид има обратно наредени поредни номера по двата признака.

Пример 2. Определете τ за пример 1 при:

ξ_k	η_k	a_k
$\xi_1 = 5$	3	$a_1 = 7$
$\xi_2 = 8$	7	$a_2 = 3$
$\xi_3 = 1$	2	$a_3 = 6$
$\xi_4 = 6$	5	$a_4 = 4$
$\xi_5 = 3$	1	$a_5 = 5$
$\xi_6 = 9$	9	$a_6 = 1$
$\xi_7 = 4$	6	$a_7 = 2$
$\xi_8 = 10$	8	$a_8 = 1$
$\xi_9 = 2$	4	$a_9 = 0$
$\xi_{10} = 7$	10	$a_{10} = 0$

Решение.

$$p = \sum_{k=1}^{10} a_k = 30; \qquad \tau = \frac{4.30}{10.9} - 1 = 0,33.$$

При отсъствие на връзка за коефициентите на Спирмън имаме

$$\mathbf{E}R = 0; \qquad \mathbf{D}R = \frac{1}{n-1},$$

а за коефициента на Кендал:

$$\mathbf{E}\tau = 0; \quad \mathbf{D}\tau = \frac{2n+5}{9}.$$

§ 6. Коефициент на конкордация

Нека разгледаме случая, когато имаме m редици, всяка от които да има брой на ранговете n. Ще разгледаме общото съотношение между горните редици върху един пример. Предполагаме, че четирима експерти подреждат 6 обекта по следния начин:

Обекти / Експерти	А	Б	В	Г	Д	Е
А1	5	4	1	6	3	2
А2	2	3	1	5	6	4
А3	4	1	6	3	2	5
А4	4	3	2	5	1	6
Обща сума на ранговете	15	11	10	19	12	17

Тук ще се интересуваме от общата мярка за съгласуваност (конкордация) между групата експерти. Нагледно тази мярка вероятно може да се получи, като усредним всички възможни стойности на R и τ, изчислени за всяка двойка експерти; но когато броят m е голям, е ясно, че такава процедура изисква много труд и време. В случаите, подобни на описаните по-горе, по-просто е да се съпоставят сумите на ранговете, „предписани" на всеки обект от различните експерти: такива данни са приведени в последния ред на таблицата. Сумата на тези числа е 84; в общия случай тя е равна на $\frac{1}{2}mn(n+1)$. Тази величина се получава чрез сумирането на m събираеми, всяко от които представлява сума от натурални числа от 1 до n. Тогава средната стойност на сумата от ранговете на един обект е $\frac{1}{2}m(n+1)$; в нашия случай тя е 14. Разглеждаме индивидуалните отклонения от тази средна стойност:

$$+1, \ -3, \ -4, \ +5, \ -2, \ +3.$$

Ако всички нареждания, дадени от експертите, съвпадат, сумите от ранговете в таблицата ще изглеждат така:

$$m,\ 2m,\ \ldots,\ nm$$

(тези суми, разбира се, не е задължително да са разположени именно в такъв ред), а техните отклонения ще са

$$-\frac{1}{2}\,m(n-1),\ -\frac{1}{2}\,m(n-3),\ \ldots,\ \frac{1}{2}\,m(n-1).$$

Сумата от квадратите на тези отклонения ще бележим с S_r:

$$S_r = \frac{1}{12}\,m^2(n^3-n).$$

Това е максималната стойност, която може да приеме сумата S_r. Другата екстремална стойност на S_r в случая на хаотично разположени, несъгласувани мнения на експертите е равна на нула, когато m е четно число или ако n е нечетно (или едното и другото едновременно). В останалите случаи S_r, макар и не равно на нула, е сравнително малко.

Означаваме с S сумата от квадратите на фактически срещащите се отклонения. Величината

$$W = \frac{S}{S_r} = \frac{12\,S}{m^2\,(n^3-n)} \tag{1}$$

се нарича *коефициент на конкордация* (*съгласуваност*) и също е въведена от М. Кендал. За нашия пример $S = 64$, така че

$$W = \frac{12.64}{16.210} = 0,229.$$

В някакъв смисъл W служи като мярка за съгласуваност на m-те експерти. Ако $W = 1$, то имаме пълна съгласуваност. Ако различията между експертите са големи, сумите от ранговете ще се окажат (много или малко) близки помежду си и затова S ще бъде по-малко от възможната максимална стойност, а следователно и коефициентът W ще бъде малка величина. Увеличаването на W от нула до едно означава, че различията между отклоненията отслабват, и в това се проявява все по-голямата съгласуваност на ранговете.

Ако означим с R_{av} средната стойност на коефициента на Спирмън, изчислен за $\binom{m}{2}$ възможни двойки експерти, то

$$R_{av} = \frac{mW-1}{m-1}. \tag{2}$$

§ 7. Корелация при номинално скаларни признаци

В някои случаи разглежданите признаци са алтернативни — имат само две състояния: „да“ или „не“. Разглеждани по двойки, те също могат да бъдат зависими. Търсим носител на информация за евентуална зависимост между наблюденията над голям брой обекти, при всяка от които се наблюдава наличието или липсата на някои от признаците. Тези наблюдения се представят с т. нар. двумерни таблици на контингентност, върху които се построява и количественият измерител на зависимостта, изложен по-долу. Таблицата на контингентност схематично се представя така:

	Признак B	Признак $\overline{B}$
Признак A	a	b
Признак $\overline{A}$	c	d

Числата в таблицата са честотите на наблюденията, притежаващи едновременно съвместната комбинация от стойности на признаците A, B, $\overline{A}$ и $\overline{B}$. Числото

$$\Phi = \frac{bc - ad}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} \tag{1}$$

се нарича ***коефициент на контингенция*** и е въведен от Пирсън и Браве.

Стойностите на коефициента Φ винаги принадлежат на интервала $[-1, 1]$, както при повечето от изброените измерители за стохастична зависимост. Почти всички те се използват еднотипно по начина, изложен в тринадесета глава.

Пример. За някои епидемии при животни и хора е необходимо да се определи колко съществена е зависимостта между ваксинираните (обработени) и неваксинираните (необработени) животни и хора. Нека при епидемия от шап имаме данните:

	Заболели	Незаболели	Суми
Обработени	5	95	100
Необработени	150	250	400

Тогава от първата таблица намираме по (1):

$$\Phi = \frac{95.150 - 5.250}{\sqrt{100.400.155.345}} = 0,31,$$

което показва, че имаме умерена, но съществена зависимост на обработката и заболеваемостта. Обработените по-рядко заболяват от тези, които не са обработени.

Задачи

1. По данните от таблицата да се намерят:

а) коефициентът на корелация $\rho_{\xi\eta}$;

б) уравнението на най-добрата линейна регресия $u(x)$, $v(y)$;

в) корелационните отношения $\theta^2_{\xi\eta}$, $\theta^2_{\eta\xi}$.

η \ ξ	5	10	15	20	n_η
10	2	—	—	—	2
20	5	4	1	—	10
30	3	8	6	3	20
40	—	3	6	6	15
50	—	—	2	1	3
n_ξ	10	15	15	10	$n = 50$

Отг. а) $\rho_{\xi\eta} = 0,636$; б) $u(x) = 1,17x + 16,73$, $v(y) = 0,345y + 1,67$; в) $\theta^2_{\eta\xi} = 0,656$, $\theta^2_{\xi\eta} = 0,651$.

2. Намерете уравнението на най-добрата параболична регресия $U(x) = \beta_0 + \beta_1 x + \beta_2 x^2$ по данни от дадената таблица. Оценете силата на корелационната връзка между ξ и η по $\theta^2_{\eta\xi}$ съгласно указанията от § 4.А:

ξ \ η	10	15	20	35	60	n_ξ
0	8	1	—	—	—	9
2	1	5	2	—	—	8
3	—	1	5	4	—	10
4	—	—	1	8	1	10
50	—	—	—	—	13	13
n_η	9	7	8	12	14	$n = 50$

Отг. $U(x) = 10,82 - 3,14x + 2,55x^2$; почти функционална — $\theta^2_{\eta\xi} = 0,95$.

3. Намерете уравнението на регресия $V(y) = a + by + cy^2$ и корелационното отношение $\theta^2_{\xi\eta}$ на данните от таблицата:

$\eta \backslash \xi$	6	30	50	n_η
1	15	—	—	15
3	1	14	—	15
4	—	2	18	20
n_ξ	16	16	18	$n = 50$

Отг. $V(y) = 3,18 + 0,02y + 2,8y^2$; $\theta^2_{\xi\eta} = 0,96$.

4. Резултатите от измерванията на стойностите на признаците x, y, z при 50 индивида са дадени в следната таблица:

x	10	10	15	15	20	20	20	20	25	25	30	30	35	40	45	50
y	40	45	40	45	45	45	50	50	25	55	55	55	60	65	65	60
z	5	5	10	10	10	15	15	20	20	30	35	40	45	45	45	45
n	8	1	4	7	2	1	2	2	1	3	2	3	4	1	1	1

Определете:

а) свободния коефициент на линейна корелация между променливите z и x, y;

б) частния коефициент на корелация $\rho_{xz}(y)$ и $\rho_{yz}(x)$;

в) уравнението на линейната регресия $z = ax + by + c$.

Отг. а) $\rho_z(x, y) = 0,87$; б) $\rho_{xz}(y) = 0,23$, $\rho_{yz}(x) = 0,54$;
в) $z = 0,52x + 1,07y - 43,5$.

5. Намерете коефициентите на корелация и напишете уравнението на най-добра линейна регресия $U(x)$ и $V(y)$ за данните от таблицата:

$\xi \backslash \eta$	1	2	4	6	7	n_ξ
1	1	2	3	—	—	6
2	—	1	2	3	—	6
3	—	—	2	4	2	8
4	—	—	—	1	2	3
n_η	1	3	7	8	4	23

Отг. $\rho_{\xi\eta} = 0,73$; $U(x) = 1,3x + 1,77$; $V(y) = 0,41y + 0,37$.

6. Учениците А, Б, В, Г, Д, Е, Ж, З, И, К от един клас са получили следните редици от рангове по математика и музика:

Ученици	А	Б	В	Г	Д	Е	Ж	З	И	К
Математика	7	4	3	10	6	2	9	8	1	5
Музика	5	7	3	10	1	9	6	2	8	4

Определете:

а) коефициента R на ранговата корелация на Спирмън;

б) коефициента τ на корелацията на Кендал.

Отг. а) $R = -0,103$; б) $\tau = -0,24$.

7. Определете коефициента на конкордация W за трите редици от рангове:

P	4,5	2	4,5	3	7,5	6	9	7,5	10
Q	1	2,5	4,5	4,5	8	9	6,5	10	6,5
R	1	4,0	4,5	4,5	4,5	8	8	0	10

Отг. $W = 0,828$.

8. Нека са дадени три редици от по осем ранга:

Експерти	А	Б	В	Г	Д	Е	Ж	З
P	4	2	1	7	6	3	5	8
Q	7	2	1	6	4	5	3	8
R	7	4	2	6	5	3	1	8
	18	8	4	19	15	11	9	24

Определете:

а) коефициента на конкордация W;

б) кое от подрежданията В Б Ж Е Д А Г З и В Ж Б Д Е Г А З е „по-добро" съобразно мненията на експертите?

Отг. а) $W = 0,88$; б) второто подреждане е „по-добро".

9. Като използвате коефициента на Пирсън – Браве, проверете дали има зависимост между успеха на студентите (над средния A, под средния $\overline{A}$) и получаването (B) или неполучаването ($\overline{B}$) на стипендия във вашата учебна група (курс).

10. Покажете, че за всеки две случайни величини ξ и η винаги могат да се намерят две некорелирани случайни величини X и Y така, че да е валидно представянето $\xi = aX + bY + c$, $\eta = a_1X + b_1Y + c$.

11. Покажете, че ако случайните величини ξ и η са свързани със строга функционална връзка $\eta = g(\xi)$, корелационното отношение е $\theta^2_{\eta\xi} = 1$.

12. Покажете, че в коефициента R на ранговата корелация на Спирмън от § 5 при условие за независимост на двата признака A и B, величината $\sum_{k=1}^{n} d_k^2$ е равномерно разпределена върху числата

$$1^2,\ 1^2+2^2,\ 1^2+2^2+3^2, \dots, 1^2+2^2+\dots+n^2,$$

представляващи възможните стойности на сумата от квадратите на първите k естествени числа, $k = 0, 1, \dots, n$.

13. Покажете, че величината a, участваща в пресмятането на ранговия коефициент на корелация τ на Кендал от § 5:

а) при условие, че и двата признака A и B са независими, приема стойности, равномерно разпределени върху числата $0, 1, 3, \ldots, \binom{n-1}{2}$, представящи възможните стойности на сумата от първите k естествени числа $\sum_{i=0}^{k} i$, $k = 0, 1, \ldots, n$;

б) при условие, че $A = kB$, $k > 0$, с вероятност 1 са изпълнени

$$\mathbf{P}\{a = 0\} = 1 \quad \text{и} \quad \mathbf{P}\{\tau = 1\} = 1;$$

в) при условие, че $A = kB$ и $k < 0$, с вероятност 1 са изпълнени

$$\mathbf{P}\left\{a = \binom{n+1}{2}\right\} = 1 \quad \text{и} \quad \mathbf{P}\{\tau = 1\} = 1.$$

14. Докажете формула (2) от § 6 за връзката между средната стойност на коефициента R на Спирмън, получен за m експерти, и коефициента на конкордация W.

15. Покажете, че ако признаците A и B са независими, то разпределението на числата в таблицата на контингентност е полиномно с параметри n, $p_1 = \mathbf{P}(A)\mathbf{P}(B)$, $p_2 = \mathbf{P}(A)\mathbf{P}(\overline{B})$, $p_3 = \mathbf{P}(\overline{A})\mathbf{P}(B)$ и $p_4 = \mathbf{P}(\overline{A})\mathbf{P}(\overline{B})$.

16. Покажете, че ако признаците A и B са независими, коефициентът на средно квадратична свързаност φ^2 от § 3 приема стойност 0, а ако A и B са функционално свързани (съществува монотонна функция $g(\bullet)$, такава че $g(A) = B$), то $\varphi^2 = m - 1$.

17. Покажете, че ако A и B приемат само две възможни стойности, то коефициентът на средно–квадратична свързаност φ^2 от § 3 и коефициентът на контингенция Φ от § 7 са свързани с израза $\varphi^2 = \Phi^2$.

18. Покажете, че изразът (коефициентът на Чупров)

$$T = \frac{\varphi^2}{\sqrt{(n-1)(m-1)}}$$

определя мярка на зависимост на признаците A и B, която е равна на 0, ако A и B са независими, и равна на 1, когато между A и B има функционална зависимост.

Упътване. Използвайте зад. 16, като за второто твърдение е необходимо $m = n$.

19. Ако двумерното разпределение на (ξ, η) е

$$\mathbf{P}\{\xi = k_1, \eta = k_2\} = \frac{n!}{k_1!\, k_2!\, (n - k_1 - k_2)!}\, p_1^{k_1}\, p_2^{k_2}\, (1 - p_1 - p_2)^{n-k_1-k_2},$$

то

$$\rho_{\xi\eta} = -\frac{\sqrt{p_1 p_2}}{(1-p_1)(1-p_2)}, \qquad \mathbf{E}(\xi\eta) = n(n-1)p_1 p_2.$$

ЛИТЕРАТУРА

1. Беляев, Ю. К., Е. В. Чепурин. Основы математической статистики. МГУ, М., 1983.

2. Бернули, Я., П. С. Лаплас, А. Н. Колмогоров. Вероятности. Наука и изкуство, С., 1982.

3. Большев, Л. Н., Н. В. Смирнов. Таблицы математической статистики. Наука, М ., 1983.

4. Боровков, А. А. Теория вероятностей. Наука, М., 1976.

5. Гешев, А. А., Б. Н. Димитров. Математическа статистика. ПУ „П. Хилендарски", Пловдив, 1981.

6. Гнеденко, Б. В. Курс теории вероятностей. Наука, М., 1969.

7. Кендалл, М., А. Стюарт. Статистические выводы и связи. Наука, М., 1973.

8. Кендалл, М., А. Стюарт. Многомерный статистический анализ и временные ряды. Наука, М., 1976.

9. Климов, Г. П. Приложна математическа статистика. Наука и изкуство, С., 1976.

10. Климов, Г. П. Теория вероятностей и математическая статистика. МГУ, М., 1983.

11. Кокс, Д., Д. Хинкли. Теоретическая статистика. Мир, М., 1978.

12. Колмогоров, А. Н. Основные понятия теории вероятностей . Наука, М.,1974.

13. Колмогоров, А. Н. Теория вероятностей. В: сб. Математика, ее содержание, методы и значение. Т. 2. АН СССР, М., 1956.

14. Королюк, В. С., Н. И. Портенко, А. В. Скороход, А. Ф. Турбин. Справочник по теории вероятностей и математической статистике. Наукова Думка, Киев, 1978.

15. Крамер, Г. Математические методы статистики. Мир, М., 1975.

16. Леман, Э. Проверка статистических гипотез. Наука, М., 1979.

17. Неве, Ж. Математические основы теории вероятностей. Мир, М., 1969.

18. Обретенов, А. Теория на вероятностите. Наука и изкуство, С., 1974.

19. Обретенов, А. Вероятностни и статистически методи. Наука и изкуство, С., 1978.

20. Обретенов, А., Д. Хаджиев. Теория на вероятностите. II част, СУ „Кл. Охридски", С., 1983.

21. Обрешков, Н. Теория на вероятностите. Наука и изкуство, С.,1963.

22. Севастьянов, Б. А. Курс теории вероятностей и математической статистики. Наука, М., 1982.

23. Севастьянов, Б. А., В. П. Чистяков, А. М. Зубков. Сборник задач по теории вероятностей. Наука, М., 1980.

24. Смирнов, Н. В., И. В. Дунин-Барковски. Курс теории вероятностей и математической статистики. Наука, М., 1965.

25. С о л о д я н н и к о в, Ю. В. Математическая статистика. Куйб. ГУ, Куйбишев, 1982.

26. С т о я н о в, Й., И. М и р а з ч и й с к и, Ц. И г н а т о в, М. Т а н у ш е в. Ръководство за упражнения по теория на вероятностите. Наука и изкуство, С., 1985.

27. Т ъ к е р, Х. Математическа статистика. Наука и изкуство, С., 1968.

28. Ф е л л е р, В. Введение в теорию вероятностей и ее применения. Т. 2, Мир, М., 1984.

29. Ф е л ъ р, У. Увод в теория на вероятностите и нейните приложения. Част I, Наука и изкуство, С., 1986.

30. Я н е в, Н. М., М. С. Т а н у ш е в. Ръководство за упражнения по математическа статистика. СУ „Кл. Охридски“, С., 1989.

ПРИЛОЖЕНИЯ

Таблица 1

Нормално разпределение $\mathcal{N}(0,1) : F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-t^2/2)\, dt$

x	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0.7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8551	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8914	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0.9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0.9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0.9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0.9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0.9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998

χ^2-разпределение с n степени на свобода $F(u) = \int\limits_0^u \frac{x^{(n-2)/2}e^{-x/2}}{2^{n/2}\Gamma(n/2)}dx$

$n \setminus F$	0,005	0,010	0,025	0,050	0,100	0,250
1	$0,0^4393$	$0,0^3157$	$0,0^3982$	$0,0^2393$	0,0158	0,102
2	0,0100	0,0201	0,0506	0,103	0,211	0,575
3	0,0717	0,115	0,216	0,352	0,584	1,21
4	0,207	0,297	0,484	0,711	1,06	1,92
5	0,412	0,554	0,831	1,15	1,61	2,67
6	0,676	0,872	1,24	1,64	2,20	3,45
7	0,989	1,24	1,69	2,17	2,83	4,25
8	1,34	1,65	2,18	2,73	3,49	5,07
9	1,73	2,09	2,70	3,33	4,17	5,90
10	2,16	2,56	3,25	3,94	4,87	6,74
11	2,60	3,05	3,82	4,57	5,58	7,58
12	3,07	3,57	4,40	5,23	6,30	8,44
13	3,57	4,11	5,01	5,89	7,04	9,30
14	4,07	4,66	5,63	6,57	7,79	10,2
15	4,60	5,23	6,26	7,26	8,55	11,0
16	5,14	5,81	6,91	7,96	9,31	11,9
17	5,70	6,41	7,56	8,67	10,1	12,8
18	6,26	7,01	8,23	9,39	10,9	13,7
19	6,84	7,63	8,91	10,1	11,7	14,6
20	7,43	8,26	9,59	10,9	12,4	15,5
21	8,03	8,90	10,3	11,6	13,2	16,3
22	8,64	9,54	11,0	12,3	14,0	17,2
23	9,26	10,2	11,7	13,1	14,8	18,1
24	9,89	10,9	12,4	13,8	15,7	19,0
25	10,5	11,5	13,1	14,6	16,5	19,9
26	11,2	12,2	13,8	15,4	17,3	20,8
27	11,8	12,9	14,6	16,2	18,1	21,7
28	12,5	13,6	15,3	16,9	18,9	22,7
29	13,1	14,3	16,0	17,7	19,8	23,6
30	13,8	15,0	16,8	18,5	20,6	24,5

Таблица 2 (продо

$n \setminus F$	0,500	0,750	0,900	0,950	0,975	0,990	0,995
1	0,455	1,32	2,71	3,84	5,02	6,63	7,88
2	1,39	2,77	4,61	5,99	7,38	9,21	10,6
3	2,37	4,11	6,25	7,81	9,35	11,3	12,8
4	3,36	5,39	7,78	9,49	11,1	13,3	14,9
5	4,35	6,63	9,24	11,1	12,8	15,1	16,7
6	5,35	7,84	10,6	12,6	14,4	16,8	18,5
7	6,35	9,04	12,0	14,1	16,0	18,5	20,3
8	7,34	10,2	13,4	15,5	17,5	20,1	22,0
9	8,34	11,4	14,7	16,9	19,0	21,7	23,6
10	9,34	12,5	16,0	18,3	20,5	23,2	25,2
11	10,3	13,7	17,3	19,7	21,9	24,7	26,8
12	11,3	14,8	18,5	21,0	23,3	26,2	28,3
13	12,3	16,0	19,8	22,4	24,7	27,7	29,8
14	13,3	17,1	21,1	23,7	26,1	29,1	31,3
15	14,3	18,2	22,3	25,0	27,5	30,6	32,8
16	15,3	19,4	23,5	26,3	28,8	32,0	34,3
17	16,3	20,5	24,8	27,6	30,2	33,4	35,7
18	17,3	21,6	26,0	28,9	31,5	34,8	37,2
19	18,3	22,7	27,2	30,1	32,9	36,2	38,6
20	19,3	23,8	28,4	31,4	34,2	37,6	40,0
21	20,3	24,9	29,6	32,7	35,5	38,9	41,4
22	21,3	26,0	30,8	33,9	36,8	40,3	42,8
23	22,3	27,1	32,0	35,2	38,1	41,6	44,2
24	23,3	28,2	33,2	36,4	39,4	43,0	45,6
25	24,3	29,3	34,4	37,7	40,6	44,3	46,9
26	25,3	30,4	35,6	38,9	41,9	45,6	48,3
27	26,3	31,5	36,7	40,1	43,2	47,0	49,6
28	27,3	32,6	37,9	41,3	44,5	48,3	51,0
29	28,3	33,7	39,1	42,6	45,7	49,6	52,3
30	29,3	34,8	40,3	43,8	47,0	50,9	53,7

Таблица

Разпределение на Стюдънт с n степени на свобода:

$$F(y)=\int_0^y \frac{\left(\frac{n-1}{2}\right)!}{\left(\frac{n-2}{2}\right)!\sqrt{\pi n\left(1+\frac{x^2}{n}\right)^{n+1}}}\,dx$$

$n \setminus F$	0,75	0,90	0,95	0,975	0,99	0,995	0,9995
1	1,000	3,078	6,314	12,706	31,821	63,657	636,619
2	0,816	1,886	2,920	4,303	6,965	9,925	31,598
3	0,765	1,638	2,353	3,182	4,541	5,841	12,941
4	0,741	1,533	2,132	2,776	3,747	4,604	8,610
5	0,727	1,476	2,015	2,571	3,365	4,032	6,859
6	0,718	1,440	1,943	2,447	3,143	3,707	5,959
7	0,711	1,415	1,895	2,365	2,998	3,499	5,405
8	0,706	1,397	1,860	2,306	2,896	3,355	5,041
9	0,703	1,383	1,833	2,262	2,821	3,250	4,781
10	0,700	1,372	1,812	2,228	2,764	3,169	4,587
11	0,697	1,363	1,796	2,201	2,718	3,106	4,437
12	0,695	1,356	1,782	2,179	2,681	3,055	4,318
13	0,694	1,350	1,771	2,160	2,650	3,012	4,221
14	0,692	1,345	1,761	2,145	2,624	2,977	4,140
15	0,691	1,341	1,753	2,131	2,602	2,947	4,073
16	0,690	1,337	1,746	2,120	2,583	2,921	4,015
17	0,689	1,333	1,740	2,110	2,567	2,898	3,965
18	0,688	1,330	1,734	2,101	2,552	2,878	3,922
19	0,688	1,328	1,729	2,093	2,539	2,861	3,883
20	0,687	1,325	1,725	2,086	2,528	2,845	3,850
21	0,686	1,323	1,721	2,080	2,518	2,831	3,819
22	0,686	1,321	1,717	2,074	2,508	2,819	3,792
23	0,685	1,319	1,714	2,069	2,500	2,807	3,767
24	0,685	1,318	1,711	2,064	2,492	2,797	3,745
25	0,684	1,316	1,708	2,060	2,485	2,787	3,725
26	0,684	1,315	1,706	2,056	2,479	2,779	3,707
27	0,684	1,314	1,703	2,052	2,473	2,771	3,690
28	0,683	1,313	1,701	2,048	2,467	2,763	3,674
29	0,683	1,311	1,699	2,045	2,462	2,756	3,659
30	0,683	1,310	1,697	2,042	2,457	2,750	3,646
40	0,681	1,303	1,684	2,021	2,423	2,704	3,551
60	0,679	1,296	1,671	2,000	2,390	2,660	3,460
120	0,677	1,289	1,658	1,980	2,358	2,617	3,373
∞	0,674	1,282	1,645	1,960	2,326	2,576	3,291

Таблица 4

F-разпределение с m степени на свобода в числителя и n — в знаменателя — $F(m,n)$:

$$G(y) = \frac{\Gamma\left(\frac{m+n}{2}\right) m^{m/2} n^{n/2}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \int_0^y x^{(m/2)-1}(n+mx)^{-(m+n)/2}\,dx$$

G	$n \backslash m$	1	2	3	4	5	6	7	8	9
0,90		39,9	49,5	53,6	55,8	57,2	58,2	58,9	59,4	59,9
0,95		161	200	216	225	230	234	237	239	241
0,975	1	648	800	864	900	922	937	948	957	963
0,99		4050	5000	5400	5620	5760	5860	5930	5980	6020
0,995		16200	20000	21600	22500	23100	23400	23700	23900	24100
0,90		8,53	9,00	9,16	9,24	9,29	9,33	9,35	9,37	9,38
0,95		18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4
0,975	2	38,5	39,0	39,2	39,2	39,3	39,3	39,4	39,4	39,4
0,99		98,5	99,0	99,2	99,2	99,3	99,3	99,4	99,4	99,4
0,995		199	199	199	199	199	199	199	199	199
0,90		5,54	5,46	5,39	5,34	5,31	5,28	5,27	5,25	5,24
0,95		10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81
0,975	3	17,4	16,0	15,4	15,1	14,9	14,7	14,6	14,5	14,5
0,99		34,1	30,8	29,5	28,7	28,2	27,9	27,7	27,5	27,3
0,995		55,6	49,8	47,6	46,2	45,1	44,8	44,4	44,1	43,9
0,90		4,54	4,32	4,19	4,11	4,05	4,01	3,98	3,95	3,93
0,95		7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00
0,975	4	12,2	10,6	9,98	9,60	9,36	9,20	9,07	8,98	8,90
0,99		21,2	18,0	16,7	16,0	15,5	15,2	15,0	14,8	14,7
0,995		31,3	26,3	24,3	23,2	22,5	22,0	21,6	21,4	21,1
0,90		4,06	3,78	3,62	3,52	3,45	3,40	3,37	3,34	3,32
0,95		6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77
0,975	5	10,0	8,43	7,76	7,39	7,15	6,98	6,85	6,76	6,68
0,99		16,3	13,3	12,1	11,4	11,0	10,7	10,5	10,3	10,2
0,995		22,8	18,3	16,5	15,6	14,9	14,5	14,2	14,0	13,8
0,90		3,78	3,46	3,29	3,18	3,11	3,05	3,01	2,98	2,96
0,95		5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10
0,975	6	8,81	7,26	6,60	6,23	5,99	5,82	5,70	5,60	5,52
0,99		13,7	10,9	9,78	9,15	8,75	8,47	8,26	8,10	7,98
0,995		18,6	14,5	12,9	12,0	11,5	11,1	10,8	10,6	10,4
0,90		3,59	3,26	3,07	2,96	2,88	2,83	2,78	2,75	2,72
0,95		5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68
0,975	7	8,07	6,54	5,89	5,52	5,29	5,12	4,99	4,90	4,82
0,99		12,2	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72
0,995		16,2	12,4	10,9	10,1	9,52	9,16	8,89	8,68	8,51

G	$n \backslash m$	1	2
0,90		3,46	3,11
0,95		5,32	4,46
0,975	8	7,57	6,06
0,99		11,3	8,65
0,995		14,7	11,0
0,90		3,36	3,01
0,95		5,12	4,26
0,975	9	7,21	5,71
0,99		10,6	8,02
0,995		13,6	10,1
0,90		3,29	2,92
0,95		4,96	4,10
0,975	10	6,94	5,46
0,99		10,0	7,56
0,995		12,8	9,43
0,90		3,18	2,81
0,95		4,75	3,89
0,975	12	6,55	5,10
0,99		9,33	6,93
0,995		11,8	8,51
0,90		3,07	2,70
0,95		4,54	3,68
0,975	15	6,20	4,77
0,99		8,68	6,36
0,995		10,8	7,70
0,90		2,97	2,59
0,95		4,35	3,49
0,975	20	5,87	4,46
0,99		8,10	5,85
0,995		9,94	6,99
0,90		2,88	2,49
0,95		4,17	3,32
0,975	30	5,57	4,18
0,99		7,56	5,39
0,995		9,18	6,35
0,90		2,79	2,39
0,95		4,00	3,15
0,975	60	5,29	3,93
0,99		7,08	4,98
0,995		8,49	5,80
0,90		2,75	2,35
0,95		3,92	3,07
0,975	120	5,15	3,80
0,99		6,85	4,79
0,995		8,18	5,54
0,90		2,71	2,30
0,95		3,84	3,00
0,975	∞	5,02	3,69
0,99		6,63	4,61
0,995		7,88	5,30

Таблица 4 (продължение)

G	$n \setminus m$	10	12	15	20	30	60	120	∞
0,90		60,2	60,7	61,2	61,7	62,3	62,8	63,1	63,3
0,95		242	244	246	248	250	252	253	254
0,975	1	969	977	985	993	1000	1010	1010	1020
0,99		6060	6110	6160	6210	6260	6310	6340	6370
0,995		24200	24400	24600	24800	25000	25200	25400	25500
0,90		9,39	9,41	2,42	9,44	9,46	9,47	9,48	9,49
0,95		19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5
0,975	2	39,4	39,4	39,4	39,4	39,5	39,5	39,5	39,5
0,99		99,4	99,4	99,4	99,4	99,5	99,5	99,5	99,5
0,995		199	199	199	199	199	199	199	199
0,90		5,23	5,22	5,20	5,18	5,17	5,15	5,14	5,1
0,95		8,79	8,74	8,70	8,66	8,62	8,57	8,55	8,53
0,975	3	14,4	14,3	14,3	14,2	14,1	14,0	13,9	13,9
0,99		27,2	27,1	26,9	26,7	26,5	26,3	26,2	26,1
0,995		43,7	43,4	43,1	42,8	42,5	42,1	42,0	41,8
0,90		3,92	3,90	3,87	3,84	3,82	3,79	3,78	3,76
0,95		5,96	5,91	5,86	5,80	5,75	5,69	5,66	5,63
0,975	4	8,84	8,75	8,66	8,56	8,46	8,36	8,31	8,26
0,99		14,5	14,4	14,2	14,0	13,8	13,7	13,6	13,5
0,995		21,0	20,7	20,4	20,2	19,9	19,6	19,5	19,3
0,90		3,30	3,27	3,24	3,21	3,17	3,14	3,12	3,11
0,95		4,74	4,68	4,62	4,56	4,50	4,43	4,40	4,37
0,975	5	6,62	6,52	6,43	6,33	6,23	6,12	6,07	6,02
0,99		10,1	9,89	9,72	9,55	9,38	9,20	9,11	9,02
0,995		13,6	13,4	13,1	12,9	12,7	12,4	12,3	12,1
0,90		2,94	2,90	2,87	2,84	2,80	2,76	2,74	2,72
0,95		4,06	4,00	3,94	3,87	3,81	3,74	3,70	3,67
0,975	6	5,46	5,37	5,27	5,17	5,07	4,96	4,90	4,85
0,99		7,87	7,72	7,56	7,40	7,23	7,06	6,97	6,88
0,995		10,2	10,0	9,81	9,59	9,36	9,12	9,00	8,88
0,90		2,70	2,67	2,63	2,59	2,56	2,51	2,49	2,47
0,95		3,64	3,57	3,51	3,44	3,38	3,30	3,27	3,23
0,975	7	4,76	4,67	4,57	4,47	4,36	4,25	4,20	4,14
0,99		6,62	6,47	6,31	6,16	5,99	5,82	5,74	5,65
0,995		8,38	8,18	7,97	7,75	7,53	7,31	7,19	7,08
0,90		2,54	2,5	2,46	2,42	2,38	2,34	2,31	2,29
0,95		3,35	3,28	3,22	3,15	3,08	3,01	2,97	2,93
0,975	8	4,30	4,20	4,10	4,00	3,89	3,78	3,73	3,67
0,99		5,81	5,67	5,52	5,36	5,20	5,03	4,95	4,86
0,995		7,21	7,01	6,81	6,61	6,40	6,18	6,06	5,95
0,90		2,42	2,38	2,34	2,30	2,25	2,21	2,18	2,16
0,95		3,14	3,07	3,01	2,94	2,86	2,79	2,75	2,71
0,975	9	3,96	3,87	3,77	3,67	3,56	3,45	3,39	3,33
0,99		5,26	5,11	4,96	4,81	4,65	4,48	4,40	4,31
0,995		6,42	6,23	6,03	5,83	5,62	5,41	5,30	5,19