Тема 19

https://github.com/v--/se2018

Проверка на хипотези.

Янис Василев

Оригинал: 28 юни 2019
Ревизия: 91d5bc9 от 09 юни 2021
За всеки случай проверете дали няма по-нова ревизия

1. Теория

Лемата на Нейман-Пирсън е базирана на изложението ѝ в Димитров и Янев, *Вероятности и статистика*.

1.1. Анотация

Изложената анотацията е взета от *Конспект за ДИ за спец. статистика*.

1. Определение за статистичека хипотеза.
2. Прости и сложни хипотези.
3. Определения за грешки от първи и твори род, критична област, мощност, значимост на тест и значимост на статистиката на теста.
4. Лема на Нейман–Пирсън.

1.2. Основни понятия

Считаме, че е зададено вероятностно пространство $(\Omega, \mathcal{F}, \mathrm{P})$.

Определение 1 (Извадки)**.** Нека ξ е случайна величина над $(\Omega, \mathcal{F}, \mathrm{P})$. Множеството от елементарни събития Ω в статистиката често се нарича **генерална съвкупност**.

- Ако случайните величини $\xi_1, \dots, \xi_n$ са независими две по две и имат същото разпределение като ξ, казваме, че $\xi_1, \dots, \xi_n$ са **наблюдения** над ξ и че те са **проста извадка** с обем n над генералната съвкупност Ω. Понякога ги разглеждаме и като случаен вектор $\overrightarrow{\xi_n} = (\xi_1, \dots, \xi_n)$.

- **Функция на правдоподобие** $l_\xi(x)$ на случайната величина ξ наричаме, в абсолютно непрекъснатия случай, плътността на ξ или, в дискретния случай, функцията на вероятностите на ξ.

- **Функция на правдоподобие** $l(x_1, \dots, x_n)$ на извадката $\xi_1, \dots, \xi_n$ наричаме функцията на правдоподобие на случайния вектор $\overrightarrow{\xi_n}$. При извадки от независими случайни величини, функцията на правдоподобие на извадката е произведение на индивидуалните функции на правдоподобие.

- **Извадково** пространство, съответстващо на извадката $\xi_1, \dots, \xi_n$, наричаме множеството $\mathcal{X} \subseteq \mathbb{R}^n$ от стойности на случайния вектор $\overrightarrow{\xi_n}$.

- **Реализации** на извадката наричаме вектори от $\mathcal{X}$. Те моделират истинските данни в математическата статистика, съпоставяйки ги на „теоретичната“ извадка $\xi_1, \dots, \xi_n$.

Определение 2 (Хипотези)**.** Нека $\xi_1, \dots \xi_n$ е проста извадка над случайната величина ξ, чието разпределение не ни е известно.

- Всяко предположение за разпределението на ξ наричаме **статистическа хипотеза**. Формално, хипотезата H често се представя като множество от възможни функции на разпределение на ξ. При повече от една хипотеза, искаме те да не се пресичат. Условна вероятност при условие, че $F_\xi \in H$, обикновена записваме чрез
$$\mathrm{P}(\cdot \mid H).$$
Обикновено се разглеждат само две хипотези: **нулевата хипотеза** H_0 и **алтернативната хипотеза** H_1.

- При **параметричната** статистика хипотезите се отнасят за параметри на семейства от разпределения, например за очакването μ на нормално разпределение или степента λ на поасоново разпределение. В противен случай говорим за **непараметрична** статистика. В непараметричния случай считаме, че двете хипотези съдържат едновременно само дискретни или само непрекъснати разпределения.

- За да направим заключение за верността на една хипотеза ни е необходим **статистически критерий**. Формално, статистическите критерии са изображения $\delta : \mathcal{X} \to \{H_0, H_1\}$, съпоставящи на реализация на извадка някоя хипотеза. Да отбележим, че критерият δ сам по себе си е случайна величина.

- Един критерий ни казва коя хипотеза да **приемем** и коя да **отхвърлим**, обикновено на базата на данни от експеримент, т.е. на някоя реализация на извадка над ξ. Поради случайния характер на експериментите, обаче, при практически задачи е възприета терминологията „имаме/нямаме основание да отхвърлим хипотезата H на база на данните“ или „хипотезата H е/не е съвместима с данните“.

- **Статистически тест** наричаме набор от хипотези и съгласуван с тях критерии.

- Вероятността α за **грешка от първи род** или **ниво на съгласие** или **ниво на значимост** е вероятността да отхвърлим вярна нулева хипотеза, т.е.

  $$\alpha := \mathrm{P}(\delta = H_1 \mid H_0).$$

  Стойността $\gamma := 1 - \alpha$ наричаме **значимост** или **ниво на доверие** на теста.

- „Вероятността β за грешка от втори род“ е вероятността да приемем грешна нулева хипотеза, т.е.

  $$\beta := \mathrm{P}(\delta = H_0 \mid H_1).$$

  Стойността $\pi := 1 - \beta$ наричаме **мощност** на теста.

- **Критична област** W_α с ниво на значимост α наричаме произволно множество $W_\alpha \subsetneq \mathcal{X}$ от реализации на извадката с $\mathrm{P}(\overrightarrow{\xi_n} \in W \mid H_0) = \alpha$, попадайки в което **отхвърляме** нулевата хипотеза.

  Всяка критична област задава критерия

  $$\delta(x) = \begin{cases} H_1, x \in W, \\ H_0, x \notin W. \end{cases}$$

- Нека са зададени хипотезите H_0 и H_1 и е фиксирано ниво на значимост α. Критичната област

  $$W_\alpha^* := \operatorname*{argmax}_{W_\alpha \subsetneq \mathcal{X}} \mathrm{P}(\overrightarrow{\xi_n} \in W_\alpha \mid H_1),$$

  осигуряваща най-голяма мощност на теста, наричаме **оптимална критична област**.

- Хипотезата наричаме **проста**, ако на нея отговаря точно едно разпределение. В противен случая я наричаме **сложна**. Ако имаме една проста хипотеза и една сложна, избираме нулевата да бъде проста. Така имаме три случая:

  1. Проста хипотеза срещу проста алтернатива,
  2. Проста хипотеза срещу сложна алтернатива,
  3. Сложна хипотеза срещу сложна алтернатива.

- Разглеждаме тест с две хипотези H_0 и H_1 като множества от възможни функции на разпределение на ξ. Имаме три случая

  – Ако $F_1(x) \leq F_0(x) \quad \forall F_0 \in H_0, F_1 \in H_1$, наричаме теста **ляв едностранен**
  – Ако $F_1(x) \geq F_0(x) \quad \forall F_0 \in H_0, F_1 \in H_1$, наричаме теста **десен едностранен**

– В противен случай, наричаме теста **двустранен**

- Нека H_0 е проста хипотеза и $x = (x_1, \dots, x_n) \in \mathcal{X}$ е реализация на извадка. **Значимост** или p-**стойност** на реализацията x наричаме условната вероятност спрямо нулева хипотеза H_0 на опашките на разпределението на ξ, определени от типа на теста.

  В зависимост от типа на теста разполагаме с няколко формални определения за значимост на реализация:

$$p := \begin{cases} \mathrm{P}(x \geq \xi \mid H_0), & \text{за леви едностранни тестове} \\ \mathrm{P}(x \leq \xi \mid H_0), & \text{за десни едностранни тестове} \\ 2\min(\mathrm{P}(x \leq \xi \mid H_0), \mathrm{P}(x \geq \xi \mid H_0)) & \text{за двустранни тестове} \end{cases}$$

  Често се казва, че значимостта на x е вероятността да наблюдаваме „по-екстремна“ стойност от x.

1.3. Лема на Нейман-Пирсън

Лема 3 (Нейман-Пирсън). *Нека са дадени две прости хипотези за функцията на правдоподобие на извадка $\xi_1, \dots, \xi_n$ над случайна величина ξ,*

$$\begin{cases} H_0: & l(x) = l_0(x) \\ H_1: & l(x) = l_1(x). \end{cases}$$

Считаме, че е зададено ниво на съгласие α. Ако за някоя реална константа $c > 0$ за критичната област W_α са изпълнени неравенствата

$$l_0(x) \leq c \cdot l_1(x), x \in W_\alpha,$$
$$l_0(x) \geq c \cdot l_1(x), x \notin W_\alpha,$$

тогава W_α е оптимална критична област.

Забележка 4. В общия случай стойностите $x \in \mathcal{X}$, за които $l_0(x) = cl_1(x)$, могат както да лежат в критичната област, така и извън нея.

Доказателство. Ще докажем теоремата само за абсолютно непрекъснати разпределения. В общия случай римановите интеграли могат да се заменят с интеграли по вероятностни мерки, съответстващи на двете хипотези.

Нека U_α е произволна друга критична област за същия тест с ниво на съгласие α. Означаваме $Q(A) := \mathrm{P}(\overrightarrow{\xi_n} \in A)$.

Ще докажем, че $Q(U_\alpha \mid H_1) \geq Q(W_\alpha \mid H_1)$. Поради адитивността на вероятностните мерки имаме

$$
\begin{aligned}
&Q(W_\alpha \mid H_1) - Q(U_\alpha \mid H_1) = \\
&= Q((U_\alpha \cap W_\alpha) \cup (W_\alpha \setminus U_\alpha) \mid H_1) - Q((U_\alpha \cap W_\alpha) \cup (U_\alpha \setminus W_\alpha) \mid H_1) = \\
&= Q(U_\alpha \cap W_\alpha \mid H_1) + Q(W_\alpha \setminus U_\alpha \mid H_1) - Q(U_\alpha \cap W_\alpha \mid H_1) - Q(U_\alpha \setminus W_\alpha \mid H_1) = \\
&= Q(W_\alpha \setminus U_\alpha \mid H_1) - Q(U_\alpha \setminus W_\alpha \mid H_1) = \\
&= \int_{W_\alpha \setminus U_\alpha} l_1(x)dx + \int_{U_\alpha \setminus W_\alpha} (-l_1(x))dx \geq \\
&\geq \frac{1}{c}\int_{W_\alpha \setminus U_\alpha} l_0(x)dx + \frac{1}{c}\int_{U_\alpha \setminus W_\alpha} (-l_0(x))dx = \\
&= \frac{1}{c}\left(\int_{W_\alpha \setminus U_\alpha} l_0(x)dx - \int_{U_\alpha \setminus W_\alpha} l_0(x)dx\right) = \\
&= \frac{1}{c}[Q(W_\alpha \setminus U_\alpha \mid H_0) - Q(U_\alpha \setminus W_\alpha \mid H_0)] = \\
&= \frac{1}{c}[Q(U_\alpha \cap W_\alpha \mid H_0) + Q(W_\alpha \setminus U_\alpha \mid H_0) - Q(U_\alpha \cap W_\alpha \mid H_0) - Q(U_\alpha \setminus W_\alpha \mid H_0)] = \\
&= \frac{1}{c}[Q(W_\alpha \mid H_0) - Q(U_\alpha \mid H_0)] = \\
&= \frac{1}{c}(\alpha - \alpha) = 0.
\end{aligned}
$$

□

2. Литература

Димитров, Боян и Николай Янев. *Вероятности и статистика.* 2-е изд. Софтех, 2007.

Конспект за ДИ за спец. статистика. 2018. URL: https://intranet.fmi.uni-sofia.bg/index.php/s/KOTdUnmqbrnd0sX (дата на посещ. 24.03.2019).