Тема 20

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

Точкови и интервални оценки за параметрите на нормалното разпределение.

Янис Василев

Оригинал: 4 юли 2019
Ревизия: 9c26f56 от 04 ноември 2024
За всеки случай проверете дали няма по-нова ревизия

1. Теория

Общата теория е базирана на Димитров и Янев, *Вероятности и статистика*, а оценките за нормалното разпределение - на Въндев, *Приложна статистика 1*.

1.1. Анотация

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

1. Определения за точкови и интервални оценки.
2. Свойства на точковите оценки.
3. Неравенство на Рао-Крамер.
4. Доверителни интервали за параметрите на нормалното разпределение.

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 (Оценки)**.** Нека $\xi_1, \dots \xi_n$ е проста извадка над случайната величина ξ, чието разпределение не ни е известно. Целта ни е на базата на тази извадка да оценим някакви числени характеристики $\theta = (\theta_1, \dots, \theta_m)$ на ξ, които напълно да определят разпределението на ξ. Обикновено θ е вектор от моменти на ξ или, в параметричната статистика, θ е някой от параметрите на класът от разпределения, на който се предполага, че принадлежи ξ. Условна вероятност при условие, че $\theta = \alpha$ е конкретна стойност на вектора, бележим с

$$\mathrm{P}(\cdot \mid \theta = \alpha) \qquad \text{или} \qquad \mathrm{P}(\cdot \mid \alpha).$$

Множеството от всички възможни стойности на θ ще бележим с Θ.

- **Статистика** наричаме всяка измерима функция на извадката. По определение статистиките са случайни величини.

- Ако $\theta = (\theta_1, \dots, \theta_m)$ са някакви числени характеристика на ξ и статистиката $t_n = t_n(x_1, \dots, x_n)$ не зависи от θ, казваме, че t_n е **точкова оценка** за θ.

- Стойността $t_n - \mathrm{E}(t_n \mid \theta)$ наричаме **случайна грешка**, а стойността $\mathrm{E}(t_n \mid \theta) - \theta$ наричаме **систематична грешка** или **изместване** на оценката. Разликата

  $$t_n - \theta = t_n - \mathrm{E}(t_n \mid \theta) + \mathrm{E}(t_n \mid \theta) - \theta$$

  се разпада на систематична и случайна грешка.

- Точковата оценка t_n за θ наричаме **неизместена**, ако $\mathrm{E}(t_n|\theta) = \theta$, и асимптотично неизместена, ако $\mathrm{E}(t_n \mid \theta) \xrightarrow[n\to\infty]{} \theta$.

- Точковата оценка t_n за θ наричаме **състоятелна**, ако $t_n \xrightarrow[n\to\infty]{} \theta$ по вероятност, т.е.

  $$\mathrm{P}(|t_n - \theta| > \varepsilon \mid \theta) \xrightarrow[n\to\infty]{} 0 \quad \forall \varepsilon > 0.$$

Оценката наричаме **силно състоятелна**, ако сходимостта е почти сигурна, т.е.

$$\mathrm{P}(\sup_{k\geq n}|t_k - \theta| > \varepsilon \mid \theta) \xrightarrow[n\to\infty]{} 0 \quad \forall \varepsilon > 0.$$

- Точковата оценка t_n за θ с изместване $\mathrm{E}(t_n) - \theta$ има **минимална дисперсия**, ако $\mathrm{var}(t_n) \leq \mathrm{var}(u_n)$ за произволна друга точкова оценка u_n със същото изместване.
- Ако векторът θ е едномерен (т.е. θ е число), **интервална оценка** с ниво на доверие γ за θ наричаме двойка точкови оценки a_n и b_n за θ, за които

  $$\mathrm{P}(a_n \leq \theta \leq b_n \mid \theta) = \gamma$$

  независимо от стойността на θ.

Твърдение 3. *За произволна случайна величина ξ с крайно очакване средноаритметичното*

$$m_n(x_1, \dots, x_n) \coloneqq \frac{1}{n}\sum_{k=1}^{n} x_k$$

е неизместена оценка за $\mathrm{E}(\xi)$.

Доказателство. Следва директно от линейността на очакването. □

Твърдение 4. *За произволна случайна величина ξ с крайна дисперсия оценката*

$$s_n^2(x_1, \dots, x_n) \coloneqq \frac{1}{n-1}\sum_{k=1}^{n} \left(x_k - \hat{\xi}(x_1, \dots, x_n)\right)^2$$

е неизместена оценка за $\mathrm{var}(\xi)$.

Доказателство. Разписваме s_n^2

$$\begin{aligned}
s_n^2(\xi_1, \dots, \xi_n) &= \frac{1}{n-1}\sum_{i=1}^{n} \left(\xi_i - \hat{\xi}(\xi_1, \dots, \xi_n)\right)^2 = \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\left[\xi_i^2 - \frac{2}{n}\sum_{j=1}^{n}\xi_i\xi_j + \frac{1}{n^2}\left(\sum_{j=1}^{n}\xi_m\right)^2\right] = \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\left[\xi_i^2 - \frac{2}{n}\sum_{j=1}^{n}\xi_i\xi_j + \frac{1}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}\xi_j\xi_k\right] = \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\xi_i^2 - \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}\xi_i\xi_j + \frac{n}{n^2(n-1)}\sum_{j=1}^{n}\sum_{k=1}^{n}\xi_j\xi_k = \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\xi_i^2 - \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}\xi_i\xi_j = \\
&= \left(\frac{1}{n-1} - \frac{1}{n(n-1)}\right)\sum_{i=1}^{n}\xi_i^2 - \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\xi_i\xi_j =
\end{aligned}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\xi_i^2 - \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\xi_i\xi_j.$$

Като вземем очакване, получаваме

$$\begin{aligned}
\mathrm{E}(s_n^2 \mid \theta) &= \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(\xi_i^2 \mid \theta) - \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\mathrm{E}(\xi_i\xi_j \mid \theta) = \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(\xi_i^2 \mid \theta) - \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\mathrm{E}(\xi_i \mid \theta)\,\mathrm{E}(\xi_j \mid \theta) = \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(\xi^2 \mid \theta) - \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\mathrm{E}(\xi \mid \theta)^2 = \\
&= \frac{n}{n}\,\mathrm{E}(\xi^2 \mid \theta) - \frac{n(n-1)}{n(n-1)}\mathrm{E}(\xi \mid \theta)^2 = \\
&= \mathrm{E}(\xi^2 \mid \theta) - \mathrm{E}(\xi \mid \theta)^2 = \\
&= \mathrm{var}(\xi^2 \mid \theta).
\end{aligned}$$

□

1.3. Информация на Фишър

Определение 5. Нека $\theta \in \Theta$ е числена характеристика на ξ, т.е. θ е едномерен вектор с възможни стойности в Θ.

Разглеждаме **логаритмичната функция на правдоподобие** $\ln l_\xi(x \mid \theta)$. За нея ще изискаме допълнителни условия за регулярност:

1. Множеството $\{x \in \mathbb{R} \mid l_\xi(x \mid \theta) = 0\}$, в което не е дефинирана $\ln l_\xi(x \mid \theta)$, има вероятност 0 независимо от θ. Така очакването на $\ln l_\xi(x \mid \theta)$ е дефинирано почти навсякъде.

2. Функцията $\ln l_\xi(x \mid \theta)$ е диференцируема по θ за (почти) всяко x от дефиниционната си област.

3. За произволно $\theta \in \Theta$ са изпълнени условията за диференциране под очакването

$$\frac{\partial\,\mathrm{E}(\ln l_\xi(x \mid \theta) \mid \theta)}{\partial\theta} = \mathrm{E}\left(\frac{\partial \ln l_\xi(x \mid \theta) \mid \theta)}{\partial\theta}\right).$$

Информация на Фишър за θ на ξ наричаме очакването

$$\mathcal{I}_\xi(\theta) \coloneqq \mathrm{var}\left(\frac{\partial \ln l_\xi(\xi \mid \theta)}{\partial\theta}\,\middle|\,\theta\right).$$

Естествено, $\mathcal{I}_\xi(\theta)$ може и да не съществува.

Информация на Фишър за θ на извадката $\xi_1, \dots, \xi_n$ над ξ наричаме информация на Фишър за θ на случайния вектор $(\xi_1, \dots, \xi_n)$. За прости извадки от независими случайни величини имаме

$$\mathcal{I}_{(\xi_1,\dots,\xi_n)}(\theta) = n\mathcal{I}_\xi(\theta).$$

Забележка 6. Третото условие за регулярност не е необходимо за определението на информация на Фишър, но е необходимо за доказателство на свойствата ѝ.

Теорема 7.

1. *Ако информацията $\mathcal{I}_\xi(\theta)$ съществува, тогава*

$$\mathcal{I}_\xi(\theta) = \mathrm{E}\left(\left(\frac{\partial \ln l_\xi(\xi \mid \theta)}{\partial \theta} \,\middle|\, \theta\right)^2\right).$$

2. *Ако освен това логаритмичното правдоподобие $\ln l_\xi(x \mid \theta)$ е двукратно диференцируемо по θ и $\ln l_\xi$ позволява двукратно диференциране под очакването, имаме*

$$\mathcal{I}_\xi(\theta) = -\mathrm{E}\left(\frac{\partial^2 \ln l_\xi(\xi \mid \theta)}{\partial \theta^2} \,\middle|\, \theta\right).$$

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 различните параметри $\theta \in \Theta$.

1. Разглеждаме очакването

$$\begin{aligned}\mathrm{E}\left(\frac{\partial \ln l_\xi(\xi \mid \theta)}{\partial \theta} \,\middle|\, \theta\right) &= \int_{\mathbb{R}} \frac{\partial \ln l_\xi(x \mid \theta)}{\partial \theta} l_\xi(x \mid \theta) dx = \\ &= \int_{\mathbb{R}} \frac{\partial l_\xi(x \mid \theta)}{\partial \theta} \frac{l_\xi(x \mid \theta)}{l_\xi(x \mid \theta)} dx = \\ &= \frac{\partial}{\partial \theta} \int_{\mathbb{R}} l_\xi(x \mid \theta) dx = \\ &= \frac{\partial 1}{\partial \theta} = 0.\end{aligned}$$

Тогава

$$\begin{aligned}\mathcal{I}_\xi(\theta) = \operatorname{var}\left(\frac{\partial \ln l_\xi(\xi \mid \theta)}{\partial \theta} \,\middle|\, \theta\right) &= \\ = \mathrm{E}\left(\left(\frac{\partial \ln l_\xi(\xi \mid \theta)}{\partial \theta} \,\middle|\, \theta\right)^2\right) &+ \left(\mathrm{E}\left(\frac{\partial \ln l_\xi(\xi \mid \theta)}{\partial \theta} \,\middle|\, \theta\right)\right)^2 = \\ = \mathrm{E}\left(\left(\frac{\partial \ln l_\xi(\xi \mid \theta)}{\partial \theta} \,\middle|\, \theta\right)^2\right)&.\end{aligned}$$

2. Директно пресмятаме

$$
\begin{aligned}
&\mathrm{E}\left(\frac{\partial^2 \ln l_\xi(\xi \mid \theta)}{\partial \theta^2} \,\middle|\, \theta\right) = \\
&= \int_{\mathbb{R}} \frac{\partial^2 \ln l_\xi(x \mid \theta)}{\partial \theta^2} l_\xi(x \mid \theta) dx = \\
&= \int_{\mathbb{R}} \frac{\partial}{\partial \theta}\left(\frac{1}{l_\xi(x \mid \theta)} \cdot \frac{\partial l_\xi(x \mid \theta)}{\partial \theta}\right) \cdot l_\xi(x \mid \theta) dx = \\
&= \int_{\mathbb{R}} \frac{1}{(l_\xi(x \mid \theta))^2} \cdot \left(\frac{\partial^2 l_\xi(x \mid \theta)}{\partial \theta^2} \cdot l_\xi(x \mid \theta) - \left(\frac{\partial l_\xi(x \mid \theta)}{\partial \theta}\right)^2\right) \cdot l_\xi(x \mid \theta) dx = \\
&= \int_{\mathbb{R}} \frac{\partial^2 l_\xi(x \mid \theta)}{\partial \theta^2} dx - \int_{\mathbb{R}} \left(\frac{\partial l_\xi(x \mid \theta)}{\partial \theta}\right)^2 \frac{1}{l_\xi(x \mid \theta)} dx = \\
&= \frac{\partial^2}{\partial \theta^2} \int_{\mathbb{R}} l_\xi(x \mid \theta) dx - \int_{\mathbb{R}} \left(\frac{\partial \ln l_\xi(x \mid \theta)}{\partial \theta} \cdot l_\xi(x \mid \theta)\right)^2 \frac{1}{l_\xi(x \mid \theta)} dx = \\
&= -\mathcal{I}_\xi(\theta).
\end{aligned}
$$

□

1.4. Неравенство на Рао-Крамер

Теорема 8. *Ако за вектор от числени характеристики $\theta = (\theta_1, \dots, \theta_m)$ на случайната величина ξ съществува неизместена оценка с минимална дисперсия*[1]*, тя е единствена.*

Доказателство. Нека t_n и u_n са неизместени оценки за θ с минимална дисперсия d^2. Ще докажем, че те са равни.

Нека $u_n = (t_n + u_n)/2$. Поради линейността на очакването, v_n също е неизместена оценка на θ. За дисперсията на v_n имаме

$$\mathrm{var}(v_n \mid \theta) = \frac{1}{4} \mathrm{E}\big((t_n + u_n)^2 \mid \theta\big) = \frac{1}{4}[\mathrm{var}(t_n \mid \theta) + \mathrm{var}(u_n \mid \theta) + 2\,\mathrm{cov}(t_n, u_n \mid \theta)].$$

Неравенството на Коши-Буняковски-Шварц ни дава

$$\mathrm{cov}(t_n, u_n \mid \theta) = \mathrm{E}(t_n u_n \mid \theta) \leq \sqrt{\mathrm{E}(t_n^2 \mid \theta)\, \mathrm{E}(u_n^2 \mid \theta)} = d^2.$$

Така получихме $\mathrm{var}(v_n \mid \theta) \leq d^2$. Тъй като дисперсията d^2 е минимална, строгото неравенство $\mathrm{var}(v_n|\theta) < d^2$ няма как да е изпълнено. Остава $\mathrm{var}(v_n|\theta) = \mathrm{cov}(t_n, u_n|\theta) = d^2$ и значи в горното неравенство се достига равенство.

Но равенство в неравенството на Коши-Буняковски-Шварц може да се достигне единствено когато t_n и u_n са линейно зависими, т.е. $t_n = \alpha u_n$ за някое $\alpha \in \mathbb{R} \setminus \{0\}$. Тогава

$$d^2 = \mathrm{var}(t_n \mid \theta) = \alpha^2\, \mathrm{var}(u_n \mid \theta) = d^2,$$

следователно $\alpha = 1$ и $t_n = u_n$. □

[1] Такава оценка наричаме **ефективна**.

Теорема 9 (Рао-Крамер)**.** *Нека θ е някаква числена характеристика на ξ е $r'(\theta)$ е диференцируема в Θ. Нека са изпълнени условията за регулярност от определение 5.*

За всяка неизместена оценка t_n на $r(\theta)$ е изпълнено

$$\operatorname{var}(t_n \mid \theta) \geq \frac{[r'(\theta)]^2}{n\mathcal{I}_\xi(\theta)},$$

като равенство се достига тогава и само тогава, когато производната на логаритмичното правдоподобие на извадката $\xi_1, \dots, \xi_n$ допуска представяне

$$\frac{\partial \ln l(x_1, \dots, x_n \mid \theta)}{\partial \theta} = k(\theta)[t_n(x_1, \dots, x_n) - r(\theta)],$$

където $k : \Theta \to \mathbb{R}$ не зависи от $x_1, \dots, x_n$.

Доказателство. Отново ще докажем теоремата само за абсолютно непрекъснати разпределения. В общия случайримановите интеграли могат да се заменят с интеграли по вероятностни мерки, съответстващи на различните параметри $\theta \in \Theta$.

Диференцираме интегралите

$$1 = \int_{\mathbb{R}^n} l(x_1, \dots, x_n \mid \theta) dx_1 \dots dx_n,$$

$$\mathrm{E}(t_n \mid \theta) = \int_{\mathbb{R}^n} l(x_1, \dots, x_n \mid \theta) t(x_1, \dots, x_n) dx_1 \dots dx_n.$$

по θ

$$0 = \int_{\mathbb{R}^n} \frac{\partial l(x_1, \dots, x_n \mid \theta)}{\partial \theta} dx_1 \dots dx_n,$$

$$r'(\theta) = \int_{\mathbb{R}^n} \frac{\partial l(x_1, \dots, x_n \mid \theta)}{\partial \theta} t(x_1, \dots, x_n) dx_1 \dots dx_n.$$

Тогава

$$\begin{aligned}
r'(\theta) &= r'(\theta) - 0 \cdot r(\theta) = \\
&= \int_{\mathbb{R}^n} \frac{\partial l(x_1, \dots, x_n \mid \theta)}{\partial \theta} \cdot [t(x_1, \dots, x_n) - r(\theta)] dx_1 \dots dx_n = \\
&= \int_{\mathbb{R}^n} \frac{\partial \ln l(x_1, \dots, x_n \mid \theta)}{\partial \theta} \cdot l(x_1, \dots, x_n) \cdot [t(x_1, \dots, x_n) - r(\theta)] dx_1 \dots dx_n = \\
&= \mathrm{E}\left(\frac{\partial \ln l(\xi_1, \dots, \xi_n \mid \theta)}{\partial \theta} \cdot [t(\xi_1, \dots, \xi_n) - r(\theta)] \,\middle|\, \theta \right).
\end{aligned}$$

От неравенството на Коши-Буняковски-Шварц получаваме

$$\begin{aligned}
(r'(\theta))^2 &= \left(\mathrm{E}\left(\frac{\partial \ln l(\xi_1, \dots, \xi_n \mid \theta)}{\partial \theta} \cdot [t(\xi_1, \dots, \xi_n) - r(\theta)] \,\middle|\, \theta \right) \right)^2 \leq \\
&\leq \mathrm{E}\left(\left(\frac{\partial \ln l(\xi_1, \dots, \xi_n \mid \theta)}{\partial \theta} \right)^2 \,\middle|\, \theta \right) \cdot \mathrm{E}\left([t(\xi_1, \dots, \xi_n) - r(\theta)]^2 \,\middle|\, \theta \right) = \\
&= n\mathcal{I}_\xi(\theta) \operatorname{var}(t_n \mid \theta),
\end{aligned}$$

като равенство се достига, когато за някоя зависеща от θ константа $k(\theta)$ е изпълнено

$$\frac{\partial \ln l(x_1, \dots, x_n \mid \theta)}{\partial \theta} = k(\theta)(t(x_1, \dots, x_n) - r(\theta)).$$

$\square$

1.5. Нормално разпределение

Нека $\xi \in \mathbf{No}(\mu, \sigma^2)$ и $\xi_1, \dots, \xi_n$ е проста извадка над ξ.

Твърдение 10. *Оценките m_n и s_n^2 за $\mathbf{No}(\mu, \sigma^2)$-разпределена извадка са независими. За разпределенията им имаме*

$$m_n \in \mathbf{No}\left(\mu, \frac{\sigma^2}{n}\right), (n-1)\frac{s_n^2}{\sigma^2} \in {}^2(n-1).$$

Доказателство. Този факт следва от теоремата на Кокрън. Наистина, нека $\eta_k = \frac{\xi_k - \mu}{\sigma}, k = 1, \dots, n$.

Разглеждаме квадратичните форми

$$\begin{aligned}
(n-1)s_n^2(\eta_1, \dots, \eta_n) &= \sum_{k=1}^{n} \xi_k^2 - [\sqrt{n} \cdot m_n(\eta_1, \dots, \eta_n)]^2, \\
\sum_{i=1}^{n}\left(\eta_i - \frac{1}{n}\sum_{j=1}^{n}\eta_j\right)^2 &= \sum_{k=1}^{n}\eta_k^2 - \frac{1}{n}\left(\sum_{k=1}^{n}\eta_k\right)^2, \\
\overrightarrow{\eta_n}^T\left(I_n - \frac{1}{n}1_n\right)\overrightarrow{\eta_n} &= \overrightarrow{\eta_n}^T I_n \overrightarrow{\eta_n} - \overrightarrow{\eta_n}^T\left(\frac{1}{n}1_n\right)\overrightarrow{\eta_n},
\end{aligned}$$

където с I_n сме означили единичната $n \times n$ матрица и

$$1_n \coloneqq \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}.$$

Имаме равенство и за ранговете на матриците на квадратичните форми, т.е.

$$\left(I_n - \frac{1}{n}1_n\right) = \operatorname{rank} I_n - \operatorname{rank}\left(\frac{1}{n}1_n\right).$$

От теоремата на Кокрън следва, че

$$\begin{aligned}
&[\sqrt{n} m_n(\eta_1, \dots, \eta_n)]^2 \in {}^2(1), \\
&(n-1)s_n^2(\eta_1, \dots, \eta_n) \in {}^2(n-1)
\end{aligned}$$

и освен това двете са независими.

Изразено чрез оригиналната извадка $\xi_1, \dots, \xi_n$, получаваме:

1. Оценката на очакването m_n е $\mathbf{No}(\mu, n\sigma^2)$-разпределена. Наистина,

$$\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\frac{\xi_k-\mu}{\sigma}\right)^2 = \left(\sqrt{n}\frac{m_n(\xi_1,\dots,\xi_n)-\mu}{\sigma}\right)^2 =$$
$$= \left[\sqrt{n}\cdot m_n(\eta_1,\dots,\eta_n)\right]^2 \in {}^{2}(1),$$

   следователно

$$\sqrt{n}\frac{m_n(\xi_1,\dots,\xi_n)-\mu}{\sigma} \in \mathbf{No}(0,1),$$
$$m_n(\xi_1,\dots,\xi_n) \in \mathbf{No}\left(\mu, \frac{\sigma^2}{n}\right).$$

2. Коригираната оценка на дисперсията $(n-1)\frac{s_n^2}{\sigma^2}$ е ${}^{2}(n-1)$-разпределена. Наистина,

$$(n-1)\frac{s_n^2(\xi_1,\dots,\xi_n)}{\sigma^2} = (n-1)s_n^2(\eta_1,\dots,\eta_n) \in {}^{2}(n-1).$$

3. Оценките m_n и s_n^2 за оригиналната извадка са независими, тъй като съответните оценки за стандартизираната извадка са техни афинни преобразования и са независими.

□

Това твърдение мотивира следната дефиниция

Определение 11. Стандартизация на извадката $\xi_1, \dots, \xi_n$ наричаме случайната величина

$$\sqrt{n}\frac{m_n-\mu}{\sigma} \in \mathbf{No}(0,1).$$

Аналогично на стандартизацията, **стюдентизация** наричаме случайната величина

$$\sqrt{n}\frac{m_n-\mu}{s_n} \in \mathbf{t}(n-1).$$

Забележка 12. Стюдентизацията има разпределение на Стюдънт, тъй като според твърдение 10 оценките m_n и s_n са независими.

1.5.1. Доверителен интервал за очакването при известна дисперсия

Със z_p ще означаваме p-квантила на стандартното нормално разпределение, т.е.

$$\Phi(z_p) = p.$$

Да забележим, че стандартното нормално разпределение е симетрично около нулата, за нас е достатъчно да знаем само един от двата квантила

$$-z_{\frac{1-\gamma}{2}} = z_{1-\frac{1+\gamma}{2}} = z_{\frac{1+\gamma}{2}}.$$

Разглеждаме следния γ-доверителен интервал с ниво на доверие

$$\mathrm{P}\left(z_{\frac{1-\gamma}{2}} \le \sqrt{n}\frac{m_n - \mu}{\sigma} \le -z_{\frac{1-\gamma}{2}} = z_{\frac{1+\gamma}{2}}\right) = \gamma,$$
$$\mathrm{P}\left(-z_{\frac{1-\gamma}{2}} \le \sqrt{n}\frac{\mu - m_n}{\sigma} \le z_{\frac{1-\gamma}{2}}\right) = \gamma,$$
$$\mathrm{P}\left(m_n - \frac{\sigma}{\sqrt{n}} z_{\frac{1-\gamma}{2}} \le \mu \le m_n + \frac{\sigma}{\sqrt{n}} z_{\frac{1-\gamma}{2}}\right) = \gamma.$$

1.5.2. Доверителен интервал за очакването при неизвестна дисперсия

С $\mathbf{t}(m)_p$ ще означаваме p-квантила на $\mathbf{t}(m)$ разпределението на Стюдънт.

Аналогично със случая с известно средно, но използвайки стюдентизация вместо стандартизация, получаваме интервала

$$\mathrm{P}\left(m_n - \frac{s_n}{\sqrt{n}} \cdot \mathbf{t}(n-1)_{\frac{1-\gamma}{2}} \le \mu \le m_n + \frac{s_n}{\sqrt{n}} \cdot \mathbf{t}(n-1)_{\frac{1-\gamma}{2}}\right) = \gamma.$$

1.5.3. Доверителен интервал за стандартното отклонение

С ${}^2(m)_p$ ще означаваме p-квантила на ${}^2(m)$ разпределението.

Тъй като според твърдение 10 имаме $(n-1)\frac{s_n^2}{\sigma^2} \in {}^2(n-1)$, директно намираме γ-доверителния интервал

$$\mathrm{P}\left({}^2(n-1)_{\frac{1-\gamma}{2}} \le (n-1)\frac{s_n^2}{\sigma^2} \le {}^2(n-1)_{\frac{1+\gamma}{2}}\right) = \gamma,$$
$$\mathrm{P}\left(\frac{{}^2(n-1)_{\frac{1-\gamma}{2}}}{(n-1)s_n^2} \le \frac{1}{\sigma^2} \le \frac{{}^2(n-1)_{\frac{1+\gamma}{2}}}{(n-1)s_n^2}\right) = \gamma,$$
$$\mathrm{P}\left(\frac{(n-1)s_n^2}{{}^2(n-1)_{\frac{1+\gamma}{2}}} \le \sigma^2 \le \frac{(n-1)s_n^2}{{}^2(n-1)_{\frac{1-\gamma}{2}}}\right) = \gamma.$$

2. Литература

Въндев, Димитър. *Приложна статистика 1*. 2002. URL: https://store.fmi.uni-sofia.bg/fmi/statist/personal/vandev/lectures/applstat.pdf.

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).