Заметки по курсу «Математическая Статистика»

Лектор: Райгородский А. М.
Автор заметок: Хоружий Кирилл

От: *29 августа 2022 г.*

Содержание

1 Распределения и их свойства

1.1 Распределение χ^2

Пусть z_1, ..., z_k – совместно независимые: $z_i \sim \mathcal{N}(0,1)$, тогда

$$x = z_1^2 + \ldots + z_k^2,$$

имеет распределение χ^2 с k степенями свободы, то есть $x \sim f_{\chi^2(k)}(x)$ с плотностью распределения

$$f_{\chi^2(k)}(x) = \frac{(1/2)^{k/2}}{\Gamma\left(\frac{k}{2}\right)} x^{\frac{1}{2}k-1} e^{-\frac{1}{2}x}.$$

2 Точечные оценки параметров распределения

Рассмотрим выборку x_1, ..., x_n. Предположим, что есть набор

$$\xi_1, \ldots, \xi_n \colon \Omega \mapsto \mathbb{R}, \qquad \xi_1(\omega) = x_1, \ldots, \xi_n(\omega) = x_n.$$

Верим, что ξ_i *независимы в совокупности* и одинаково распределены.

Предполагаем, что $F_{\xi_i} \in \{F_{\boldsymbol{\theta}}\}_{\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^m}$, где Θ – параметрическое множество.

Def 2.1. Оценка (*точечная*) – любая функция $\hat{\theta}_n$ элементов выборки : $\operatorname{Im} \hat{\theta}_n \in \Theta$.

Вообще оценка это $\hat{\theta}_n(\xi_1, \ldots, \xi_n)$ – случайная величина. Естественно определить *несмещенную* оценку

$$\mathrm{E}\, \hat{\theta}_n = \theta$$

и *состоятельную* оценку

$$\hat{\theta}_n \xrightarrow[n\to\infty]{\mathrm{P}} \theta,$$

где имеется ввиду *сходимость по вероятности*:

$$\forall \varepsilon > 0 \ P(|\hat{\theta}_n - \theta| > \varepsilon) \xrightarrow[n\to\infty]{} 0.$$

Рассмотрим, например, нормальное распределение $\xi_i \sim \mathcal{N}(\theta_1, \theta_2^2)$. Положим $\hat{\theta}_n = \bar{x} = \frac{1}{n}\sum_{i=1}^n \xi_i$, тогда

$$\mathrm{E}\, \hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n \mathrm{E}\, \xi_i = \theta_1,$$

то есть оценка несмещенная. Более того, оценка состоятельна

$$\hat{\theta}_n = \frac{1}{n}(\xi_1 + \ldots + \xi_n) \xrightarrow[n\to\infty]{\mathrm{P}} \mathrm{E}\, \xi_i = \theta_1,$$

из закона больших чисел[1].

Теперь для дисперсии в $\xi_i \sim \mathcal{N}(a, \theta_2^2)$ разумно взять

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n (\xi_i - a)^2,$$

которая является состоятельной и несмещенной.

Но, для неизвестного среднего $\xi_i \sim \mathcal{N}(\theta_1, \theta_2^2)$ уже по-другому

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^n \left(\xi_i - \frac{1}{n}\sum_{j=1}^n \xi_j\right)^2, \qquad \mathrm{E}\, \hat{\theta}_n = \frac{n-1}{n}\theta_2^2,$$

получили смещенную оценку (асимптотически несмещенную), а для несмещенной оценки необходимо брать множитель $\frac{1}{n-1}$.

2.1 Функции риска

Хочется научиться сравнивать разные оценки, рассмотрим разность

$$\hat{\theta}_n - \theta,$$

и введем *функцию потерь* U такую, что $U(0) = 0$, $U(x) = U(-x)$ и монотонную. Например $U(x) = x^2$, $U(x) = |x|$ – квадратичная и абсолютная функция потерь. Тогда *функция риска*

$$R_{\hat{\theta}_n}(\theta) = \mathrm{E}\, U(\hat{\theta}_n - \theta).$$

[1] Может быть усилено до сходимость почти наверное. Тогда оценка будет называться сильно состоятельной.

Например, для несмещенной $\hat{\theta}_n$ и $U(x) = x^2$: $R_{\hat{\theta}_n}(\theta) = \mathrm{D}\,\hat{\theta}_n$. Глобально хотим минимизировать функцию риска.

Есть *интегральный* (*байесовский*) подход:

$$\int_\Theta R_{\hat{\theta}_n}(\theta)\,d\theta \to \min_{\hat{\theta}_n},$$

а для другой априорной информации корректно рассмотреть другую меру $d\mu$.

И есть минимаксный подход:

$$\max_{\theta\in\Theta} R_{\hat{\theta}_n}(\theta) \to \min_{\theta_n}.$$

2.2 Эмпирическая функция распределения

Хотим по элементам выборки определить функцию распределения $\hat{F}_n(x_1, \ldots, x_n)$ – *эмпирическую функцию распределения*, которую можем определить в виде

$$\hat{F}_n(x;\xi_1, \ldots, \xi_n) \overset{\text{def}}{=} \frac{1}{n}\sum_{i=1}^n \mathrm{I}_{\{\xi_i \leqslant x\}},$$

где *индикаторная функция*

$$\mathrm{I}_{\{\text{cond}\}} \overset{\text{def}}{=} \begin{cases} 1, & \text{if cond;} \\ 0, & \text{else.} \end{cases}$$

Можем упорядочить элементы выборки $x_{(1)} \leqslant \ldots \leqslant x_{(n)}$ – вариационный ряд, его элементы $x_{(k)}$ – k-е порядковые статистики.

Lem 2.2. *Для $\forall x \in \mathbb{R}$ верно, что $\hat{F}_n(x;\xi_1, \ldots, \xi_n)$ – несмещенная и состоятельная оценка $F_{\xi_i}(x)$.*

Верно более сильное утверждение – *теорема Гливенко–Кантелли*:

$$\mathrm{P}\left(\sup_{x\in\mathbb{R}} |\hat{F}_n(x;\xi_1, \ldots, \xi_n) - F_{\xi_i}(x)| \xrightarrow[n\to\infty]{} 0\right) = 1.$$

Умножив на $\sqrt{n}$, перейдём к *теореме Колмогорова*: пусть F_{ξ_i} непрерывна и

$$\sqrt{n}D_n \overset{\text{def}}{=} \sqrt{n}\sup_{x\in\mathbb{R}} |\hat{F}_n(x;\xi_1, \ldots, \xi_n) - F_{\xi_i}(x)|,$$

тогда

$$\sqrt{n}D_n \xrightarrow[n\to\infty]{\mathrm{d}} \eta: \quad F_\eta(y) = \begin{cases} 0, & y \leqslant 0; \\ K(y), & y > 0, \end{cases} \qquad K(y) = \sum_{i=-\infty}^{\infty} (-1)^i e^{-2i^2y^2}, y > 0,$$

где[2] $K(y)$ – функция распределения Колмогорова.

2.3 Асимптотическая нормальность

Оценка $\hat{\theta}_n$ называется *асимптотически нормальной*, если

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow[n\to\infty]{d} \eta \sim \mathcal{N}(0, \sigma^2[\theta]),$$

где σ^2 – асимптотическая дисперсия.

Def 2.3. *Медианной точкой* распределения называют $x_{1/2}$: $F(x_{1/2}) = 1/2$, аналогично вводят α-квантиль с x_α.

Рассмотрим $F_{\xi_i}(x) = F(x-\theta)$, тогда для $p(0) \neq 0$ и $p(-x) = p(x)$ верно, что $\theta = \mathrm{E}\,\xi_i = x_{1/2}$. В качестве одной из оценок можем рассматривать $\hat{\theta}_n^{[1]} = \bar{x}$, которая будет асимптотически нормальной:

$$\sigma^2_{\bar{x}} = \mathrm{D}\,\xi_1.$$

Для $x_{1/2}$ хорошей оценкой выступит $\hat{\theta}_n^{[2]} = x_{(n/2)}$ – выборочная медиана, для которой верна теорема:

$$\sqrt{n}\left(\hat{\theta}_n^{[2]} - x_{1/2}\right) \xrightarrow[n\to\infty]{d} \eta \sim \mathcal{N}\left(0, \tfrac{1}{4p^2[0]}\right).$$

Как выбирать из двух асимптотически нормальных оценок? Просто брать ту, что с меньшим значением σ^2, она будет более эффективной.

[2] «$\xrightarrow[n\to\infty]{\mathrm{d}}$» – сходимость по распределению.

2.4 Построение статистических оценок: метод моментов

Понятно, что можем построить состоятельные несмещенные оценки, вида

$$\mathrm{E}\,\xi_i^k \;:\; \overline{x^k} = \frac{1}{n}\sum_{i=1}^n \xi_i^k.$$

Зная, что величины распределены по $F_{\xi_i} \in \{F_{\boldsymbol{\theta}}\}_{\boldsymbol{\theta}\in\Theta\subset\mathbb{R}^m}$, можем посчитать

$$\mathrm{E}\,\xi_i^k = f_k(\theta_1,\, ...,\, \theta_m).$$

Составим систему уравнений

$$f_1(\theta_1,\, ...,\, \theta_m) = \bar{x}, \qquad \dots, \qquad f_m(\theta_1,\, ...,\, \theta_m) = \overline{x^m},$$

которую будем считать разрешимой и с $\exists!$ решением – $\hat{\theta}_n^1,\, ...,\, \hat{\theta}_n^m$, искомой оценкой параметров распределения.

2.5 Построеник статистических оценок: метод максимального правдоподобия

Рассмотрим $f(x;\,\boldsymbol{\theta}) = P_\theta(\xi = x)$ для дискретного случая и $f(x;\,\boldsymbol{\theta}) = p_\xi(x;\,\boldsymbol{\theta})$ для абсолютно непрерывного случая. Определим *функцию правдоподобия*

$$L(x_1,\, ...,\, x_n; \boldsymbol{\theta}) = \prod_{i=1}^n f(x_i;\,\boldsymbol{\theta}),$$

то есть чем вероятнее для $\boldsymbol{\theta}$ наблюдение таких параметров, тем больше функция правдоподобия. Оценка метода правдоподобия (ОМП) – это точка $\hat{\theta}_n$ максимума функции L.

Удобно работать с $\ln L$, тогда будем решать систему

$$\frac{\partial \ln L}{\partial \theta_1} = 0, \qquad \dots, \qquad \frac{\partial \ln L}{\partial \theta_m} = 0,$$

и находить ОМП.

Например, для $\mathrm{Binom}(k, \theta)$

$$P(\xi_i = x_i) = C_k^{x_i}\theta^{x_i}(1-\theta)^{k-x_i}, \qquad L(x_1,\, ...,\, x_n; \theta) = \prod_{i=1}^n C_k^{x_i}\theta^{x_i}(1-\theta)^{k-x_i},$$

тогда можем найти

$$\frac{\partial \ln L}{\partial \theta} = \sum_{i=1}^n \left(\frac{x_i}{\theta} - \frac{k - x_i}{1-\theta}\right) = 0, \qquad \Rightarrow \qquad \hat{\theta}_n = \frac{1}{kn}\sum_{i=1}^n x_i.$$

2.6 Неравенство Рао-Крамера

Введем множество $A = \{x \mid f(x;\,\boldsymbol{\theta}) \neq 0\}$. *Первое условие регулярности* – $A \neq A(\boldsymbol{\theta})$. *Вторым условием регулярности* является дифференцируемость $f(x;\,\theta)$ по $\boldsymbol{\theta}$ на A. Наконец, введем $u = \partial_\theta \ln f(\xi;\,\boldsymbol{\theta})$, и условием регулярности будет выступать

$$\mathrm{E}\,u = 0 \;\; \forall \theta \in \Theta, \qquad 0 < \mathrm{D}\,u < \infty.$$

К слову про дипсерсию

$$\mathrm{D}\,u = \mathrm{E}\left(\frac{\partial \ln f}{\partial \theta}\right)^2 \overset{\text{def}}{=} I_1(\theta),$$

где $I_1(\theta)$ – информация Фишера, или $I_n(\theta) = nI_1(\theta)$:

$$I_n(\theta) \overset{\text{def}}{=} \mathrm{E}\left(\partial_\theta \ln L(\xi_1,\, ...,\, \xi_n; \theta)\right)^2.$$

Thr 2.4 (неравенство Рао-Крамера)**.** *Пусть выполнены условия регулярности, $\hat{\theta}_n$ – несмещенная оценка θ и $\mathrm{E}\,\hat{\theta}_n$ можно дифференцировать по θ под знаком интеграла. Тогда*[3] $\mathrm{D}\,\hat{\theta}_n \geqslant \frac{1}{I_n(\theta)}$.

Thr 2.5. *Пусть выполнены условия регулярности, $f(x;\boldsymbol{\theta})$ трижды дифференцируема по θ, $\mathrm{E}\,\partial_\theta^3 f(\xi;\theta) \leqslant h(\xi)$ с $\mathrm{E}\,h < \infty$. Тогда решение уравнения $\partial_\theta \ln L = 0$ является ОМП. Эта оценка асимптотически нормальна с наименьшей дисперсией $\frac{1}{I_1(\theta)}$ – асимптотически эффективна.*

[3] Где $I_n(\theta) \sim n \cdot \mathrm{const}$.

3 Проверка статистических гипотез

3.1 Гипотезы о согласии

Посмотрели на выборку, предполагаем H_0, что $F_{\xi_i} = F$. *Ошибка I рода* – отвергаем H_0, хотя она верна. *Ошибка II рода* – принять H_0, хотя верна $H \neq H_0$.

Ω-критейрий. Выберем $\Omega \subset \mathbb{R}^n$ – *критическое множество*, и считаем, что если $(x_1, ..., x_n) \in \Omega$, то H_0 неверна, если $(x_1, ..., x_n) \notin \Omega$, то H_0 верна.

Если Ω выбрана, то

$$P_0\left[(\xi_1, ..., \xi_n) \in \Omega\right] = \int_\Omega p_{\xi_1, ..., \xi_n}(x_1, ..., x_n)\, dx_1 \ldots dx_n.$$

Критерий Колмогорова. Предполагая F непрерывной, можем выбрать

$$\Omega = \{(x_1, ..., x_n) \mid \sqrt{n} D_n(x_1, ..., x_n) > c\}, \qquad D_n = \sup_{x \in \mathbb{R}} |\hat{F}_n(\xi_1, ..., \xi_n) - F(x_1, ..., x_n|,$$

где $\hat{F}_n$ – эмпирическая функция распределения. Осталось вспомнить, что по теореме Колмогорова

$$P_0\left[(\xi_1, ..., \xi_n) \in \Omega\right] \approx P(\eta > c) = 1 - K(c).$$

Осталось выбрать $c_{1-\alpha}$ так, чтобы $K(c_{1-\alpha}) = 1 - \alpha$, и получить $P_0 = \alpha$. Удобно определить

$$\Omega_\alpha = \{(x_1, ..., x_n) \mid \sqrt{n} D_n(x_1, ..., x_n) > c_{1-\alpha}\},$$

где α – уровень значимости.

Критерий χ^2. Считаем в рамках H_0, что $F_{\xi_i} = F$; $\xi_i \in \Delta$. Представим $\Delta = \Delta_1 \sqcup \ldots \sqcup \Delta_N$. Какие-то из элементов выборки $x_1, ..., x_n$ (ν_1 штук) попали в Δ_1, какие-то (ν_2 штук) в Δ_2 и т.д. Введем p_i – вероятности того, что $\xi_i \in \Delta$, если H_0 верна. Тогда, по *теореме Пирсона*

$$\sum_{i=1}^N \frac{(\nu_i - np_i)^2}{np_i} \xrightarrow[n \to \infty]{d} \eta \sim \chi^2_{N-1}.$$

Тогда аналогично строим

$$\Omega_\alpha = \left\{(x_1, ..., x_n) \mid \sum_{i=1}^N \frac{(\nu_i - np_i)^2}{np_i} > \chi^2_{N_1, 1-\alpha}\right\},$$

для уровня значимости α.