

Оценивание вещественного параметра

Малов Сергей Васильевич

Санкт-Петербургский государственный электротехнический
университет

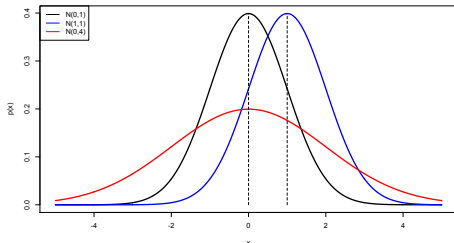
26 сентября / 3 октября 2020 г.

- 1 Параметрические семейства распределений
- 2 Выборка из нормального распределения
- 3 Постановка задачи оценивания вещественного параметра
- 4 Построение статистических оценок

Некоторые параметрические семейства

Семейство нормальных распределений $\xi \in \mathcal{N}(a, \sigma^2)$

- Тип: абсолютно непрерывное распределение
- Параметр распределения: $\theta = (a, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+$
- Плотность распределения: $p_{\theta}(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)$

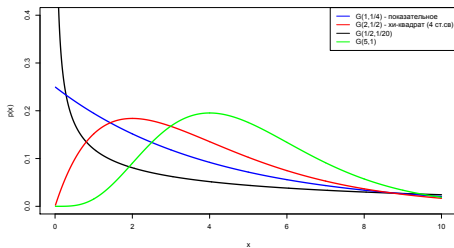


- Носитель распределения: \mathbb{R}
- a – параметр сдвига; σ – параметр масштаба
- Математическое ожидание: $\mathbb{E}_{\theta}\xi = a$
- Дисперсия: $\mathbb{D}_{\theta}\xi = \sigma^2$

Некоторые параметрические семейства

Семейство гамма распределений (Эрланга) $\xi \in \Gamma(a, b)$

- Тип: абсолютно непрерывное распределение
- Параметр распределения: $\theta = (a, b) \in \mathbb{R}_+ \times \mathbb{R}_+$
- Плотность распределения: $p_\theta(x) = \frac{1}{\Gamma(a)b^a} x^{a-1} \exp(-x/b) 1_{\{x \geq 0\}}$
- $\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx$ – гамма функция Эйлера, $a > 0$.

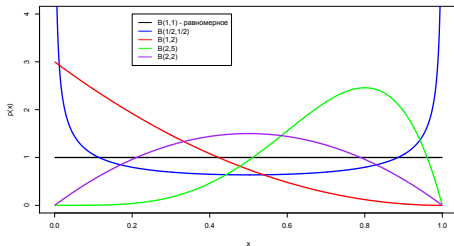


- Носитель распределения: $[0, \infty)$
- a – параметр формы; b – параметр масштаба
- Математическое ожидание: $\mathbb{E}_\theta \xi = ab$
- Дисперсия: $\mathbb{D}_\theta \xi = ab^2$

Некоторые параметрические семейства

Семейство бета распределений $\xi \in B(a, b)$

- Тип: абсолютно непрерывное распределение
- Параметр распределения: $\theta = (a, b) \in \mathbb{R}_+ \times \mathbb{R}_+$
- Плотность распределения: $p_\theta(x) = \frac{x^{a-1}(1-x)^{b-1}1_{\{x \in (0,1)\}}}{\beta(a,b)}$
- $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 x^{a-1}(1-x)^{b-1} dx$ – бета-функция, $a, b > 0$.

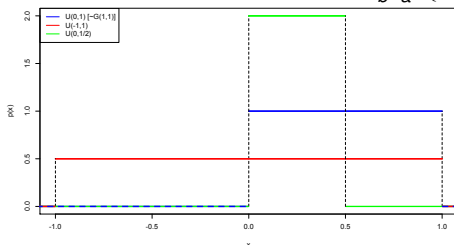


- Носитель распределения: $[0, 1]$
- a, b – параметры формы
- Математическое ожидание: $\mathbb{E}_\theta \xi = a/(a+b)$
- Дисперсия: $\mathbb{D}_\theta \xi = \frac{ab}{(a+b)^2(a+b+1)}$

Некоторые параметрические семейства

Семейство равномерных распределений $\xi \in U(a, b)$

- Тип: абсолютно непрерывное распределение
- Параметр распределения: $\theta = (a, b) \in \{(x, y) \in \mathbb{R} : x < y\}$
- Плотность распределения: $p_{\theta}(x) = \frac{1}{b-a} 1_{\{x \in [a, b]\}}$

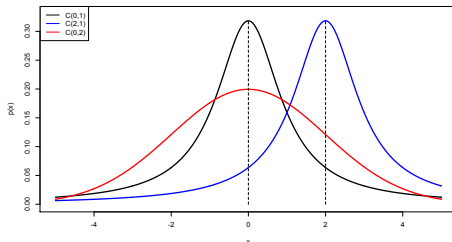


- Носитель распределения: $[a, b]$
- a – левая граница носителя; b – правая граница носителя
- Математическое ожидание: $\mathbb{E}_{\theta}\xi = (a + b)/2$
- Дисперсия: $\mathbb{D}_{\theta}\xi = (b - a)^2/12$

Некоторые параметрические семейства

Семейство распределений Коши $\xi \in C(a, b)$

- Тип: абсолютно непрерывное распределение
- Параметр распределения: $\theta = (a, b) \in \mathbb{R} \times \mathbb{R}_+$
- Плотность распределения: $p_{\theta}(x) = \frac{b}{\pi((x-a)^2 + b^2)}$



- Носитель распределения: \mathbb{R}
- a – параметр сдвига; b – параметр масштаба
- Математическое ожидание: $\mathbb{E}_{\theta}\xi$ не существует

- 1 Параметрические семейства распределений
- 2 Выборка из нормального распределения
- 3 Постановка задачи оценивания вещественного параметра
- 4 Построение статистических оценок

Определение

Пусть $\xi_0, \xi_1, \dots, \xi_n$ – выборка из стандартного нормального распределения $\mathcal{N}(0, 1)$.

- Распределение случайной величины $\chi_n^2 = \sum_{i=1}^n \xi_i^2$ называется хи-квадрат с n степенями свободы.
- Распределение случайной величины $S_n = \frac{\xi_0}{\sqrt{\frac{1}{n} \chi_n^2}}$ называется распределением Стьюдента (W. Gosset) с n степенями свободы

Определение

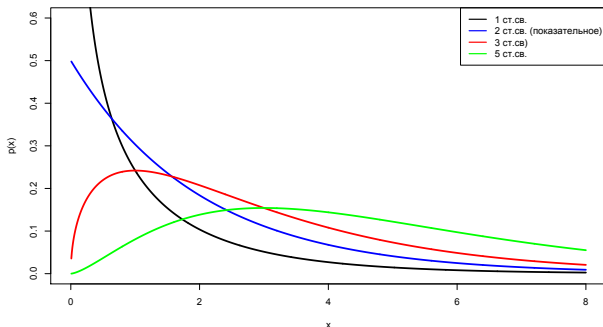
Пусть χ_n^2 и χ_m^2 – независимые случайные величины, имеющие распределения χ^2 с n и m степенями свободы соответственно.

- Распределение случайной величины $\psi_{n,m} = \frac{\chi_n^2/n}{\chi_m^2/m}$ называется распределением Фишера–Снедекора с n и m степенями свободы.

Распределения χ^2

Плотность хи-квадрат распределения ξ_n^2

$$p_n(x) = \frac{1}{\Gamma(n/2)2^{n/2}} x^{n/2-1} \exp(-x/2) 1_{\{x \geq 0\}}$$



- Хи-квадрат - частный случай гамма распределения: $\chi_n^2 \sim \Gamma(n/2, 2)$
- Носитель распределения: \mathbb{R}_+
- Математическое ожидание: $\mathbb{E}_\theta \chi_n^2 = n$
- Дисперсия: $\mathbb{D}_\theta \chi_n^2 = 2n$

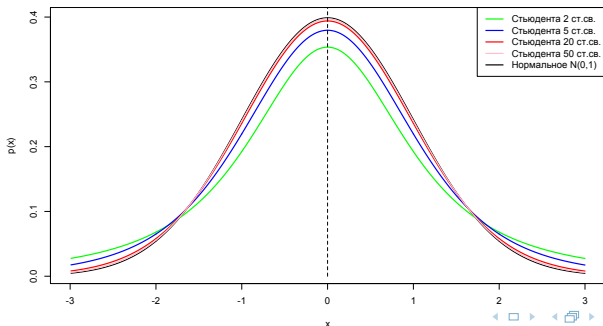
Распределения Стьюдента

Плотность распределения Стьюдента

$$t_n(x) = \frac{\Gamma((n+1)/2)}{\sqrt{\pi n} \Gamma(\frac{n}{2}) (1 + x^2/n)^{(n+1)/2}}, \quad x \in \mathbb{R}.$$

- Распределения Стьюдента также называются t -распределениями
- Распределения Стьюдента симметричны относительно нуля
- С ростом числа степеней распределение Стьюдента приближается (сходится) к стандартному нормальному распределению

Плотности распределений Стьюдента с различными ч.с.с.

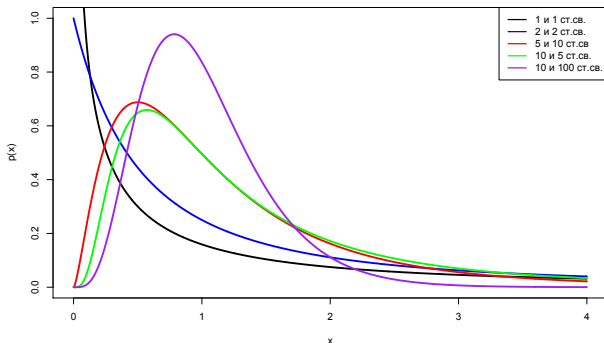


Распределения Фишера–Снедекора

Плотность распределения Фишера–Снедекора $F_{n,m}$

$$p_{n,m}(x) = \frac{(m/n)^{m/2}}{B(n/2, m/2)} \frac{x^{n/2-1}}{(x + m/n)^{(n+m)/2}} 1_{\{x \geq 0\}}$$

- $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 x^{a-1}(1-x)^{b-1} dx$ – бета-функция, $a, b > 0$.



- Распределение $F_{n,m}$ сходится к χ_n^2 при $m \rightarrow \infty$
- Если $\nu \sim S_m$ (распределение Стьюдента), то $\nu^2 \sim F_{1,m}$
- Математическое ожидание: $\mathbb{E}_\theta f_{n,m} = m/(m-2)$, $m > 2$
- Дисперсия: $\mathbb{D}_\theta f_{n,m} = 2m^2(m+n-2)/(n(m-2)^2(m-4))$, $m > 4$

Теорема (Лемма Фишера)

Пусть X_1, X_2, \dots, X_n – выборка из нормального распределения $\mathcal{N}(a, \sigma^2)$. Тогда

- (i) $\sqrt{n} \frac{\bar{X} - a}{\sigma} \in \mathcal{N}(0, 1)$;
- (ii) \bar{X} и s^2 – независимые статистики;
- (iii) $\frac{ns^2}{\sigma^2}$ имеет χ^2 -распределение с $(n - 1)$ степенью свободы;
- (iv) $\sqrt{n - 1} \frac{\bar{X} - a}{s}$ имеет распределение Стьюдента с $n - 1$ степенями свободы.

Доказательство леммы Фишера

Доказательство. Введем случайные величины $Y_k = X_k - \mathbb{E}X_k$, $1 \leq k \leq n$. Они независимы и имеют нормальное распределение $N(0, \sigma^2)$.

Рассмотрим произвольную ортонормированную $n \times n$ -матрицу A , первая строка которой состоит из чисел, равных $\frac{1}{\sqrt{n}}$ (построение ортонормированного базиса). Пусть ξ_k , $1 \leq k \leq n$ – новые случайные величины, такие что $\xi = (\xi_1, \dots, \xi_n)' = \frac{1}{\sigma} AY$, $Y = (Y_1, \dots, Y_n)'$. В этом случае

$$\xi_1 = \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^n Y_k = \frac{\sqrt{n}}{\sigma} \bar{Y}, \quad \mathbb{E}\xi_k = 0, \quad \mathbb{D}\xi_k = 1, \quad 1 \leq k \leq n, \\ \text{cov}(\xi_k, \xi_l) = 0 \quad (k \neq l).$$

Кроме того, можно выразить s^2 через случайные величины ξ_k , $1 \leq k \leq n$, которые независимы, как некоррелированные нормальные $N(0, 1)$. Поскольку A – ортогональна, суммы квадратов Y_k/σ^2 и ξ_k равны, так что

$$\frac{ns^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{k=1}^n Y_k^2 - \left(\frac{\bar{Y}}{\sigma} \right)^2 = \xi_2^2 + \dots + \xi_n^2.$$

Отсюда уже следует независимость ns^2/σ^2 и $\bar{X} = \frac{\sigma}{\sqrt{n}} \xi_1 + a$.

Утверждение о распределениях ns^2/σ^2 и \bar{X} теперь очевидно. ■

Следствие

Пусть X_1, X_2, \dots, X_n и Y_1, Y_2, \dots, Y_m – независимые выборки из нормальных распределений $\mathcal{N}(a, \sigma_1^2)$ и $\mathcal{N}(b, \sigma_2^2)$ соответственно.

Тогда

$F = \frac{ns_1^2}{ms_2^2} \frac{\sigma_2^2}{\sigma_1^2}$ имеет распределение Фишера–Снедекора $F_{n-1, m-1}$

- $s_1^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ – выборочная дисперсии, построенная по X_1, X_2, \dots, X_n
- $s_2^2 = \frac{1}{m} \sum_{j=1}^m (Y_j - \bar{Y})^2$ – выборочная дисперсии, построенная по Y_1, Y_2, \dots, Y_m .

- 1 Параметрические семейства распределений
- 2 Выборка из нормального распределения
- 3 Постановка задачи оценивания вещественного параметра
- 4 Построение статистических оценок

Пусть $(\mathfrak{X}, \mathfrak{F}, \mathcal{P})$, $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$, $\Theta \subseteq \mathbb{R}$ – статистический эксперимент

- Задача точечного оценивания: по результатам наблюдений найти приближенное значение параметра $\tilde{\theta}(X) \in \Theta$
- Точечная оценка – статистика $\tilde{\theta} : \mathfrak{X} \rightarrow \Theta$ (измеримая)
- Для определения близости оценки $\tilde{\theta}$ к истинному значению параметра θ вводится функция потерь $W(\tilde{\theta}, \theta)$, удовлетворяющая следующим условиям:
 - неотрицательность: $W(\tilde{\theta}, \theta) \geq 0$
 - если $\tilde{\theta} = \theta$, то потери нулевые: $W(\theta, \theta) = 0$
- Наиболее употребительные функции потерь:
 - $W(\tilde{\theta}, \theta) = (\tilde{\theta} - \theta)^2$ – функция потерь Гаусса
 - $W(\tilde{\theta}, \theta) = |\tilde{\theta} - \theta|$ – функция потерь Лапласа
- Риск оценки $R_{\tilde{\theta}} : \Theta \rightarrow [0, \infty)$
 - $R_{\tilde{\theta}}(\theta) = \mathbb{E}_\theta W(\tilde{\theta}(X), \theta)$ – среднее значение потерь при каждом фиксированном θ , $\theta \in \Theta$

Оптимальное оценивание

- Наиболее часто для измерения точности оценки $\tilde{\theta}$ используют гауссовский риск

$$R_{\tilde{\theta}}(\theta) = E_{\theta}(\tilde{\theta}(X) - \theta)^2$$

- Оценку, имеющую наименьший риск по сравнению с другими оценками при каждом значении θ , разумно считать оптимальной
 - в классе всевозможных оценок не существует оценки, минимизирующей риск при каждом значении θ
- Различают два подхода в задаче оптимального оценивания:
 - поиск оценки, минимизирующей риск при всех значениях $\theta \in \Theta$ в ограниченном классе разумных оценок
 - например, среди несмещенных оценок
 - использование функционалов от риска (байесовский, минимаксный подходы)

- 1 Параметрические семейства распределений
- 2 Выборка из нормального распределения
- 3 Постановка задачи оценивания вещественного параметра
- 4 Построение статистических оценок

Функция правдоподобия

- Пусть $\mathcal{P} \ll \mu$, т.е. любая мера семейства распределений доминирована мерой μ : $\mathbb{P}_\theta \ll \mu$
 - $\mathbb{P}_\theta \ll \mu$, если $\forall A \in \mathfrak{F}: \mu(A) = 0 \Rightarrow \mathbb{P}_\theta(A) = 0$
 - по теореме Радона–Никодима существует плотность $p_\theta \equiv \frac{d\mathbb{P}_\theta}{d\mu}$
 - $P_\theta(A) = \int_A p_\theta(x) d\mu(x)$, $A \in \mathfrak{F}$.
- В качестве доминирующей меры обычно используют
 - меру Лебега: p_θ – плотность распределения абсолютно непрерывной случайной величины
 - считающая мера на \mathbb{Z} : $p_\theta(x) = \mathbb{P}(X = x)$, $x \in \mathbb{Z}$
- Функция правдоподобия $L: \mathfrak{X} \times \Theta \rightarrow \mathbb{R}_+$:

$$L(x; \theta) = p_\theta(x), \quad x \in \mathfrak{X}, \theta \in \Theta.$$

Определение

Оценка $\hat{\theta}(X)$, максимизирующая функцию правдоподобия $L(X; \theta)$ по θ при каждом фиксированном X (т.е. $L(X; \hat{\theta}(X)) \geq L(X; \tilde{\theta}(X))$) при каждом $X \in \mathcal{X}$ для любой оценки $\tilde{\theta}(X)$, называется оценкой максимального правдоподобия (ОМП).

- Если $\hat{\theta}$ – оценка максимального правдоподобия параметра θ , g – параметрическая функция, то $g(\hat{\theta})$ является оценкой максимального правдоподобия для $g(\theta)$.
- Если X_1, \dots, X_n – выборка из распределения с плотностью распределения p_θ , $\theta \in \Theta$, то

- функция правдоподобия распадается в произведение

$$L(X; \theta) = \prod_{i=1}^n p_\theta(X_i)$$

- логарифм функции правдоподобия представляется в виде суммы

$$LL(X; \theta) = \log L(X; \theta) = \sum_{i=1}^n \log p_\theta(X_i)$$

Нахождение ОМП

- В силу монотонности логарифма задача максимизации функции правдоподобия сводится к задаче максимизации её логарифма по всем $\theta \in \Theta$

$$\log L(X; \theta) = \sum_{i=1}^n \log p_{\theta}(X_i).$$

- удобнее максимизировать сумму, а не произведение
- Если $\theta = (\theta_1, \dots, \theta_d)$ – d -мерный параметр и p_{θ} дифференцируема по θ , то для нахождения максимума надо найти решения системы уравнений

$$U(X; \theta) = \frac{\partial}{\partial \theta_i} \log L(X; \theta) = 0, \quad i = 1, \dots, d.$$

- Если функция правдоподобия не дифференцируема по параметру, то переход от функции правдоподобия к её логарифму обычно не имеет смысла

Пусть X_1, \dots, X_n – выборка из распределения P_θ , $\theta = (\theta_1, \dots, \theta_d)$, и существуют $\mu_k(\theta) = \mathbb{E}_\theta X_1^k$, $k = 1, \dots, n$.

Определение

Если существует единственное решение $\tilde{\theta}(X)$ системы уравнений

$$\hat{\mu}_k(\tilde{\theta}) = \mu_k(\tilde{\theta}), \quad k = 1, \dots, d,$$

где $\hat{\mu}_k(\tilde{\theta}) = \frac{1}{n} \sum_{i=1}^n X_i^k$ – выборочный момент k -го порядка, то $\tilde{\theta}(X)$ называется оценкой по методу моментов (ОММ).

- Выборочное распределение иногда называют непараметрической оценкой максимального правдоподобия распределения выборки
- Для построения ОММ используются моменты выборочного распределения
- Довольно часто ОММ и ОМП совпадают

Упражнение

Пусть X_1, \dots, X_n – выборка из двухпараметрического нормального распределения $N(\mathbf{a}, \sigma^2)$. Найти ОМП и ОММ для параметра распределения $\theta = (\mathbf{a}, \sigma^2)$.

Решение. (ОМП) Логарифм функции правдоподобия имеет вид

$$\log L(X; \theta) = -\frac{n}{2} \log(2\pi\sigma^2) - \sum_{i=1}^n \frac{(X_i - \mathbf{a})^2}{2\sigma^2}.$$

Точка максимума функции правдоподобия находится из системы уравнений

$$\begin{cases} 2 \sum_{i=1}^n (X_i - \mathbf{a}) = 0, \\ n/\sigma - \sum_{i=1}^n \frac{(X_i - \mathbf{a})^2}{\sigma^3} = 0. \end{cases}$$

Получаем ОМП $(\hat{\mathbf{a}}, \hat{\sigma}^2) = (\bar{X}, s^2)$.

(ОММ) Известно, что $\mathbb{E}_\theta X_1 = \mathbf{a}$ и $\mathbb{D}_\theta X_1 = \sigma^2$. Следовательно, система метода моментов

$$\begin{cases} \frac{1}{n} \sum_{i=1}^n X_i = \mathbf{a}, \\ \frac{1}{n} \sum_{i=1}^n X_i^2 = \mathbf{a}^2 + \sigma^2. \end{cases} \Leftrightarrow \begin{cases} \mathbf{a} = \frac{1}{n} \sum_{i=1}^n X_i, \\ \sigma^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \left(\frac{1}{n} \sum_{i=1}^n X_i \right)^2 \end{cases}$$

Получаем ОМП $(\hat{\mathbf{a}}, \hat{\sigma}^2) = (\bar{X}, s^2)$. ■

Построение статистических оценок

Упражнение

Пусть X_1, \dots, X_n – выборка из двухпараметрического равномерного распределения на интервале (a, b) ($\theta = (a, b)$).

Решение. (ОМП) Функция правдоподобия имеет вид

$$L(x; \theta) = \prod_{i=1}^n \frac{1}{(b-a)} 1_{[a,b]}(x_i) = \frac{1}{(b-a)^n} 1_{[a,b]}(\min_i(x_i)) 1_{[a,b]}(\max_i(x_i)),$$

где

$$1_A(x) = \begin{cases} 1, & \text{если } x \in A; \\ 0, & \text{в остальных случаях.} \end{cases}$$

Функция правдоподобия не дифференцируема, но максимизация функции правдоподобия сводится к максимизации $(b-a) \rightarrow \max_{a,b}$ при условии $a \leq \min_{1 \leq i \leq n} (X_i) \leq \max_{1 \leq i \leq n} (X_i) \leq b$.

Решение данной задачи оптимизации – ОМП: $(\hat{a}, \hat{b}) = (\min_{1 \leq i \leq n} (X_i), \max_{1 \leq i \leq n} (X_i))$.

(ОММ) Известно, что $\mathbb{E}_\theta X_1 = \frac{a+b}{2}$ и $\mathbb{E}_\theta X_1^2 = \frac{a^2+ab+b^2}{3}$. Система ММ:

$$\begin{cases} \bar{X} = a, \\ \bar{X}^2 = \frac{a^2+ab+b^2}{3} \end{cases} \Leftrightarrow \begin{cases} b+a=2\bar{X}, \\ b-a=4\sqrt{3}\sqrt{\bar{X}-\bar{X}^2} \end{cases} \Leftrightarrow \begin{cases} a=\bar{X}-2\sqrt{3}\sqrt{\bar{X}-\bar{X}^2}, \\ b=\bar{X}+2\sqrt{3}\sqrt{\bar{X}-\bar{X}^2} \end{cases}$$

Таким образом, ОММ: $(\tilde{a}, \tilde{b}) = (\bar{X} - 2\sqrt{3}s, \bar{X} + 2\sqrt{3}s)$. ■