Московский физико-технический институт

факультет инноваций и высоких технологий

Математическая статистика

Лектор: М.Е. Жуковский

конспект лекций

автор: Александр Марков

23 мая 2017 г.

Содержание

1 Сходимость случайных векторов

Определение 1.1. Пусть $\xi,\ \xi_1,\ \ldots,\ \xi_n$ — k-мерные случайные вектора. Как и в случае случайных величин, существуют следующие виды сходимости:

1. $\xi_n \xrightarrow{\text{п.н.}} \xi$ если $\mathsf{P}(\xi_n \to \xi) = 1$ (сходимость почти наверное)
2. $\xi_n \xrightarrow{\mathsf{P}} \xi$ если $\forall \varepsilon > 0:\ \mathsf{P}(||\xi_n - \xi||_2 > \varepsilon) \to 0$, где $||x||_t = \sqrt[t]{\sum\limits_{i=1}^{k} |x_i|^t}$ для $x = (x_1,\ \ldots,\ x_k) \in \mathbb{R}^k$ (сходимость по вероятности)
3. $\xi_n \xrightarrow{d} \xi$ если для любой непрерывной ограниченной функции $f : \mathbb{R}^k \to \mathbb{R}$ верно $\mathsf{E} f(\xi_n) = \mathsf{E} f(\xi)$ (сходимость по распределению, слабая сходимость)
4. $\xi_n \xrightarrow{L_p} \xi$ если $\mathsf{E}\left(||\xi_n - \xi||_p\right)^p \to 0$ (сходимость в L_p)

Утверждение 1.0.1. Пусть $\xi,\ \xi_1,\ \ldots$ — случайные k-мерные вектора. Тогда верны следующие взаимосвязи между сходимостью векторов и их компонент:

$$\left.\begin{aligned} &\xi_n \xrightarrow{\text{п.н.}} \xi \\ &\xi_n \xrightarrow{\mathsf{P}} \xi \\ &\xi_n \xrightarrow{L_p} \xi \end{aligned}\right\} \iff \forall i \in \{1,\ \ldots,\ k\} \begin{cases} \xi_n^{(i)} \xrightarrow{\text{п.н.}} \xi^{(i)} \\ \xi_n^{(i)} \xrightarrow{\mathsf{P}} \xi^{(i)} \\ \xi_n^{(i)} \xrightarrow{L_p} \xi^{(i)} \end{cases}$$

Доказательство.

1. *сходимость почти наверное.* $\Rightarrow$: $\{\xi_n^{(i)} \to \xi^{(i)}\} \supset \{\xi_n \to \xi\}$ и вероятность события справа равна 1.
   $\Leftarrow$: $\{\xi_n \to \xi\} = \bigcap\limits_{j=1}^{k} \{\xi_n^{(j)} \to \xi^{(j)}\}$ (известно из матана) и вероятность справа просто равна 1.
2. *сходимость по вероятности.* $\Rightarrow$: $\{|\xi_n^{(i)} - \xi^{(i)}| > \varepsilon\} \subset \{||\xi_n - \xi||_2 > \varepsilon\}$
   $\Leftarrow$: $\bigcup\limits_{i=1}^{k} \{|\xi_n^{(i)} - \xi^{(i)}| > \frac{\varepsilon}{k}\} \supset \{||\xi_n - \xi|| > \varepsilon\}$
3. *сходимость в L_p.* Очевидна цепочка неравенств

$$0 \leqslant \left|\xi_n^{(i)} - \xi^{(i)}\right|^p \leqslant \left|\xi_n^{(1)} - \xi^{(1)}\right|^p + \ldots + \left|\xi_n^{(k)} - \xi^{(k)}\right|^p$$

   Тогда $\Leftarrow$ следует из линейности мат.ожидания, а $\Rightarrow$ из свойства мат.ожидания $f \leqslant g \Rightarrow \mathsf{E} f \leqslant \mathsf{E} g$.

$\square$

Напоминание: критерием сходимости по распределению может служить теорема Александрова: если F_ξ непрерывна, то $\xi_n \xrightarrow{d} \xi \iff F_{\xi_n}(x) \to F_\xi(x)$ для любого $x \in \mathbb{R}^k$.

Теорема 1.1. *(теорема о наследовании сходимостей)*

1. *Пусть $\xi_n \to \xi$ почти наверное или по вероятности, а $h : \mathbb{R}^k \to \mathbb{R}^m$, такая что $\mathsf{P}(h$ непрерывна$) =$ 1. Тогда $h(\xi_n) \to h(\xi)$ почти наверное или по вероятности.*

2. *Пусть* $\xi_n \xrightarrow{d} \xi$, $h : \mathbb{R}^k \to \mathbb{R}^m$ *и непрерывна (Замечание: это не тоже самое, что и первом пункте). Тогда* $h(\xi_n) \xrightarrow{d} h(\xi)$.

Доказательство. Для доказательства теоремы воспользуемся следующей леммой:

Лемма 1.1. *Если последовательность случайных векторов сходится по вероятности, то из нее можно извлечь подпоследовательность, сходящуюся почти наверное.*

которая является прямым следствием одномерного случая (Выделим подпоследовательность для 1 координаты, из нее подпоследовательность для 2 координаты и так далее. В итоге получим сходимость почти наверное всех координат). Приступим к доказательству теоремы:

1. $\xi_n \xrightarrow{\text{п.н.}} \xi$:
$$\mathsf{P}(h(\xi_n) \to h(\xi)) \geqslant \mathsf{P}(h(\xi_n) \to h(\xi), \xi \in B) \geqslant \mathsf{P}(\xi_n \to \xi, \xi \in B) = 1$$
где $B = \{h \text{ непрерывна}\}, \mathsf{P}(\xi \in B) = 1$.

2. $\xi_n \xrightarrow{\mathsf{P}} \xi$: Предположим, что $h(\xi_n) \stackrel{\mathsf{P}}{\not\to} h(\xi)$. Это означает, что
$$\exists \varepsilon > 0 \exists \delta > 0 : \ \mathsf{P}(||h(\xi_n) - h(\xi)|| > \varepsilon) > \delta - (1)$$
для бесконечно многих n. Пусть $\{n_j\}$ это те номера, при которых верно неравенство выше. Из условия $\xi_{n_j} \xrightarrow{\mathsf{P}} \xi$. По лемме можно выделить подпоследовательность $\xi_{n_{j_k}} \xrightarrow{\text{п.н.}} \xi$. По доказанному ранее, $h(\xi_{n_{j_k}} \xrightarrow{\text{п.н.}} h(\xi)$, что противоречит (1).

3. $\xi_n \xrightarrow{d} \xi$: Рассмотрим непрерывную ограниченную функцию $f : \mathbb{R}^m \to \mathbb{R}$. Тогда $f(h) = f \circ h : \mathbb{R}^k \to \mathbb{R}$ — непрерывная и ограниченная функция, а значит
$$\mathsf{E} f(h(\xi_n)) = \mathsf{E}(f \circ h)(\xi_n) \to \mathsf{E}(f \circ h)(\xi) = \mathsf{E} f(h(\xi))$$
и $h(\xi_n) \xrightarrow{d} h(\xi)$.

□

Теорема 1.2. *(лемма Слуцкого)*

1. *Если* $\xi_n \xrightarrow{d} \xi$, *а* $\eta_n \xrightarrow{d} \eta = c = const$ *— случайные величины. Тогда* $\xi_n + \eta_n \xrightarrow{d} \xi + c$, $\xi_n \eta_n \xrightarrow{d} c\xi$

2. *Если* $\xi_n \xrightarrow{d} \xi = const$ *— случайные вектора, то* $\xi_n \xrightarrow{\mathsf{P}} \xi$.

Доказательство. Докажем только второе утверждение.

Поскольку функция проектор непрерывна, то, по теореме о наследовании сходимости $\xi_n \xrightarrow{d} \xi \Rightarrow \xi_n^{(i)} \xrightarrow{d} \xi^{(i)}$, откуда
$$\xi_n^{(i)} \xrightarrow{d} C^{(i)} \Rightarrow \xi_n^{(i)} \xrightarrow{\mathsf{P}} C^{(i)} \Rightarrow \xi_n \xrightarrow{\mathsf{P}} \xi$$
поскольку в одномерном случае сходимость к константе по распределению эквивалентна сходимости по вероятности (*тем, кто забыл: теорема Александрова*). □

Утверждение 1.0.2. Пусть $\xi_n \xrightarrow{d} \xi$ — случайные вектора размерности $m \geqslant 1$, $h : \mathbb{R}^m \to \mathbb{R}$ — функция, дифференцируемая в точке $a \in \mathbb{R}^m$. Пусть $b_n \to 0$, $b_n \neq 0$. Тогда

$$\frac{h(a + \xi_n b_n) - h(a)}{b_n} \xrightarrow{d} \langle \xi, \nabla h|_a \rangle$$

Доказательство. $b_n \to 0 \Rightarrow b_n \xi_n \xrightarrow{\mathsf{P}} 0$ по лемме Слуцкого. По формуле Тейлора справедливо представление

$$h(a + x) = h(a) + \langle \nabla h|_a,\ x \rangle + \varphi(x)$$

где $\varphi(x) = o(||x||)$ и непрерывна в 0. Поскольку $\frac{\varphi(x)}{||x||} \to 0$, то по теореме о наследовании сходимости $\frac{\varphi(\xi_n b_n)}{||b_n \xi_n||} \xrightarrow{\mathsf{P}} 0$.

Подставим в формулу Тейлора $x = \xi_n b_n$:

$$\frac{h(a + \xi_n b_n) - h(a)}{b_n} = \frac{\langle \nabla h|_a, \xi_n b_n \rangle}{b_n} + \frac{\varphi(\xi_n b_n)}{b_n}$$

По теореме о наследовании сходимостей $||\xi_n|| \xrightarrow{d} ||\xi||$. Тогда по лемме Слуцкого $\frac{\varphi(\xi_n b_n)}{b_n} = \frac{\varphi(\xi_n b_n)}{b_n ||\xi_n||} \cdot ||\xi_n|| \xrightarrow{\mathsf{P}} 0$, а $\frac{\langle \nabla h|_a, \xi_n b_n \rangle}{b_n} = \langle \nabla h|_a, \xi_n \rangle \xrightarrow{d} \langle \nabla h|_a, \xi \rangle$ по теореме о наследовании сходимостей.

Объединяя все вышесказанное, имеем

$$\frac{h(a + \xi_n b_n) - h(a)}{b_n} \xrightarrow{d} \langle \xi, \nabla h|_a \rangle$$

$\square$

2 Вероятностно-статистическая модель

Предположим, что мы наблюдаем некоторый эксперимент. Пусть $\mathscr{X}$ — множество всех возможных значений эксперимента.

Определение 2.1. Множество $\mathscr{X}$ называется выборочным пространством.

Обозначим за $\mathscr{B}(\mathscr{X})$ некоторую σ-алгебру на $\mathscr{X}$ (в случае, когда $\mathscr{X} = \mathbb{R}^k$ — борелевскую). $\mathcal{P}$ — семество некоторыех вероятностных мер (распределений) на измеримом пространстве $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$ (например все абсолютно непрерывные распределения) и пусть $\mathsf{P} \in \mathcal{P}$ — некоторое заданное распределение вероятностей на $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$.

Определение 2.2. *Наблюдением* называется функция $X : \mathscr{X} \to \mathscr{X}$, такая что $\forall x \in \mathscr{X} : X(x) = x$ — случайная величина.

Мотивировка: заметим, что $\mathsf{P}(X \in B) = \mathsf{P}_X(B) \Rightarrow \mathsf{P}_X(x) = \mathsf{P}(x)$, где P — заданное распределение на $(\mathscr{X}, \mathscr{B}(\mathscr{X}))$.

Рассмотрим теперь $\mathscr{X}^n$. Зададим на нем $\mathscr{B}(\mathscr{X}^n) = \sigma(B_1 \times \ldots \times B_n,\, B_i \in \mathscr{B}(\mathscr{X}))$. Зададим распределение вероятностней P^n на $(\mathscr{X}^n, \mathscr{B}(\mathscr{X}^n))$ по правилу $\mathsf{P}^n(B_1 \times \ldots \times B_n) = \mathsf{P}(B_1) \ldots \mathsf{P}(B_n)\, \forall B_i \in \mathscr{B}(\mathscr{X})$.

Утверждение 2.0.1. (б/д, следствие теоремы о продолжении меры). Существует единственная вероятностная мера P^*, заданная на всем $(\mathscr{X}^n, \mathscr{B}(\mathscr{X}^n))$, такая что $\forall B_i \in \mathscr{B}(\mathscr{X}) : P^*(B_1 \times \ldots \times B_n) = \mathsf{P}^n(B_1 \times \ldots \times B_n)$. Будем обозначать P^* тем же символом P^n.

Определение 2.3. Функция $X : \mathscr{X}^n \to X^n$; $X(x) = x$ называется *наблюдением.* Аналогично одномерному случаю, $\mathsf{P}_X = \mathsf{P}^n$.

Утверждение 2.0.2. X — вектор из независимых одинаково распределенных случайных величин, такой что любая его координата имеет распределение P.

Доказательство. Сначала покажем, что каждая координата X имеет распределение P:

$$\mathsf{P}(X_i \in B) = \mathsf{P}^n(\mathscr{X}_{j \neq i} \in \mathscr{X},\, X_i \in B) = \mathsf{P}(B) \cdot \prod_{j \neq i} \mathsf{P}(\mathscr{X}) = \mathsf{P}(B)$$

Теперь установим независимость:

$$\mathsf{P}^n(X_1 \in B_1, X_2 \in B_2, X_{i>2} \in \mathscr{X}) = \mathsf{P}^n(B_1 \times B_2 \times \mathscr{X} \times \ldots \times \mathscr{X}) = \mathsf{P}(B_1)\mathsf{P}(B_2) = \mathsf{P}^n(X_1 \in B_1)\mathsf{P}(X_2 \in B_2)$$

□

Определение 2.4. $X = (X_1, \ldots, X_n)$ — выборка из $\mathscr{X}$ размера n.

Поскольку многие из рассматирваемых в будущем свойств статистик и распределений ассимптотические, необходимо уметь получать выборку любого конечного размера n. Для этого введем следующие определения:

Определение 2.5. $\mathscr{X}^\infty = \mathscr{X} \times \mathscr{X} \times \ldots = (x_1,\, x_2,\, \ldots)$, $\forall i\, x_i \in \mathscr{X}$ — множество бесконечных последовательностей элементов из $\mathscr{X}$.

$\mathscr{B}(\mathscr{X}^\infty) = \sigma(\{(x_1,\, \ldots,\, x_n,\, \ldots)|\, (x_1,\, \ldots,\, x_n) \in B,\, B \in \mathscr{B}(\mathscr{X}^n)\},\, \forall n \in \mathbb{N})$ — цилиндрическая σ-алгебра. Под знаком σ рассматриваются все множества из $\mathscr{X}^\infty$, такие что для некоторого n, первые n их координат являются координатами множества из $\mathscr{B}(\mathscr{X}^n)$.

Определение 2.6. Обозначим P^∞ распределение на $(\mathscr{X}^\infty,\, \mathscr{B}(\mathscr{X}^\infty))$, заданное по следующему правилу: пусть $B \in \mathscr{B}(\mathscr{X}^n)$. Тогда $\mathsf{P}^\infty(B) = \mathsf{P}^\infty(B \times \mathscr{X} \times \ldots) = \mathsf{P}^n(B)$.

Утверждение 2.0.3. Существует единственная вероятностная мера P^*, заданная на всем $(\mathscr{X}^\infty, \mathscr{B}(\mathscr{X}^\infty))$, совпадающая на элементах $\mathscr{B}(\mathscr{X}^n)$ с P^n. — аналогично n-мерному случаю, будем обозначать P^* так же P^∞.

Определение 2.7. Функция $X : \mathscr{X}^\infty \to X^\infty$ такая что $X(x) = x$, как и прежде, называется наблюдением.

Утверждение 2.0.4. *(б/д, аналогично конечномерному случаю)*Пусть $X = (X_1,\, X_2,\, \ldots)$. Тогда $\{X_i\}_{i=1}^\infty$ это независимые одинаково распределенные случайные величины с распределением P каждая.

Будем в дальнешем для простоты обозначений писать $(\mathscr{X},\, \mathscr{B}(\mathscr{X}),\, \mathsf{P})$ вместо $(\mathscr{X}^\infty, \mathscr{B}(\mathscr{X}^\infty),\, \mathsf{P}^\infty)$ и называть выборку наблюдениеми наоборот.

Определение 2.8. Тройка

$$(\mathscr{X},\, \mathscr{B}(\mathscr{X}),\, \mathcal{P})$$

где

a) $\mathscr{X}$ — выборочное пространство,

b) $\mathscr{B}(\mathscr{X})$ — σ-алгебра на $\mathscr{X}$,

c) $\mathcal{P}$ — множество вероятностых мер на измеримом простанстве $(\mathscr{X},\, \mathscr{B}(\mathscr{X}))$

называется *вероятностно-статистической моделью.*

3 Статистики. Непараметрические статистики

3.1 Определение статистики. Примеры

Определение 3.1. Пусть дано измеримое пространство $(E, \mathscr{E})$ и $(\mathscr{B}(\mathscr{X})|\mathscr{E})$-измеримая функция $S : \mathscr{X} \to E$. Тогда композиция функций $S \circ X = S(X)$ называется *статистикой.*

Пример 3.1. $\overline{X} = \frac{X_1+\ldots+X_n}{n}$ — выборочное среднее.

Пример 3.2. Пусть g — некоторая $(\mathscr{B}(\mathscr{X})|\mathscr{E})$-измеримая функция. Тогда статистикой является $\overline{g(X)} = \frac{g(X_1)+\ldots+g(X_n)}{n}$. Такая статистика называется выборочной характеристикой.

Пример 3.3. Различные функции от выборочных характеристик тоже являются статистиками. Для примера рассмотрим $h(x,y) = x - y^2$, $h : \mathbb{R}^2 \to \mathbb{R}$, $\mathscr{X} = \mathbb{R}$. Тогда $h(\overline{X^2}, \overline{X}) = \overline{X^2} - \overline{X}^2$ является статистикой, называется *выборочной дисперсией* и обозначается s^2.

Утверждение 3.1.1. $s^2 = \frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$

Доказательство. Рассмотрим числа $x_1, x_2, \ldots, x_n \in \mathbb{R}$ и случайную величину $\xi \sim U(\{x_1, \ldots, x_n\})$. Посчитаем дисперсию ξ двумя способами:

$$\begin{aligned}\mathsf{D}\xi &= \mathsf{E}(\xi - \mathsf{E}\xi)^2 = \mathsf{E}(\xi - \overline{x})^2 = \frac{1}{n}\sum(x_i - \overline{x})^2 \\ &= \mathsf{E}\xi^2 - (\mathsf{E}\xi)^2 = \overline{x^2} - (\overline{x})^2\end{aligned}$$

□

Пример 3.4. *Порядоквые статистики.* Рассмотрим случай $\mathscr{X} = \mathbb{R}$. Тогда

$$\begin{aligned}&X_{(1)} = \min\{X_1, \ldots, X_n\} \\ &X_{(2)} = \min\{\{X_1, \ldots, X_n\} \setminus \{X_{(1)}\}\} \\ &\ldots \\ &X_{(n)} = \max\{X_1, \ldots, X_n\}\end{aligned}$$

Эти n статистик называются *порядковыми статистиками*, $X_{(k)}$ — k-ая порядковая статистика, а $(X_{(1)}, \ldots, X_{(n)})$ — *вариационный ряд.*

3.2 Непараметрические статистики

Пусть $X_1, \ldots, X_n$ выборка из неизвестного распределения P, а $B \in \mathscr{B}(\mathscr{X})$ и перед нами стоит задача восстановить $P(B)$.

Определение 3.2. Вероятностная мера P_n^*, заданная по правилу

$$\mathsf{P}_n^*(B) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \in B)$$

называется *эмпирическим распределением*, построенным по выборке $X_1, \ldots, X_n$.

Утверждение 3.2.1. Пусть $\{X_n\}_{n=1}^{\infty}$ — выборка неограниченного размера на $(\Omega, \mathscr{F}, \mathsf{P})$. Тогда

$$\forall B \in \mathscr{B}(\mathscr{X}) : \mathsf{P}_n^*(B) \to \mathsf{P}_X(B) \text{ при } n \to \infty$$

Доказательство. Зафиксируем множество B. Тогда $\mathsf{P}_n^*(B) = \frac{1}{n}\sum\limits_{i=1}^{n} I(X_i \in B)$. По УЗБЧ, $\mathsf{P}_n^* \xrightarrow{\text{п.н.}} \mathsf{E}I(X_i \in B)$, но поскольку X_i имеют распределение P_X, то $P_n^*(B) \xrightarrow{\text{п.н.}} \mathsf{E}I(X \in B) = \mathsf{P}(X \in B) = \mathsf{P}_X(B)$ $\square$

Рассмотрим случай $(\mathscr{X}, \mathscr{B}(\mathscr{X})) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$

Определение 3.3. Функия $F_n^*(x) = \mathsf{P}_n^*((-\infty, x]) = \frac{1}{n}\sum\limits_{i=1}^{n} I(X_i \leqslant x)$ называется эмпирической функцией распределения, построенной по выборке $X_1, \ldots, X_n$.

Теорема 3.1. *(Гливенко-Кантелли)*

Пусть $\{X_n\}_{n=1}^{\infty}$ — выборка из неизвестного распределения P с функцией распределения F. Тогда

$$D_n = \sup_{x\in\mathbb{R}} |F_n^*(x) - F(x)| \xrightarrow{\text{п.н.}} 0$$

Доказательство. Поскольку F_n^* равна константе на каждом из отрезков $[X_{(k)}, X_{(k+1)}]$, то

$$D_n = \sup_{0\leqslant k\leqslant n} \left\{ \left| F(X_{(k)} - \frac{k}{n} \right|, \left| F_(X_{(k+1)}) - \frac{k}{n} \right| \right\},$$

где $X_{(0)} = -\infty$, $X_{(n+1)} = +\infty$, а значит D_n — действительно случайная величина.

Зафиксируем $N \in \mathbb{N}$. Определим число $x_{k,N} := \min\{x \in \mathbb{R} | \ F(x) \geqslant \frac{k}{N}\}$ (определение корректно, поскольку F непрервына справа) для $k \in \{1, \ldots, N-1\}$, $x_{0,N} := -\infty$, $x_{N,N} := +\infty$.

Пусть $x \in [x_{k,N}, x_{k+1,N})$. Тогда

$$\begin{aligned} F_n^*(x) - F(x) &\leqslant F_n^*(x_{k+1,N} - 0) - F(x_{k,N}) = \\ &= F_n^*(x_{k+1,N} - 0) + \underbrace{F(x_{k+1,N} - 0)}_{\leqslant (k+1)/N} - \underbrace{F(x_{k,N})}_{\geqslant k/N} - F(x_{k+1,N}) \\ &\leqslant F_n^*(x_{k+1,N} - 0) - F(x_{k+1,N} - 0) + \frac{1}{N} \end{aligned}$$

Аналогично $F_n^*(x) - F(x) \geqslant F_n^*(x_{k,N}) - F(x_{k,N}) - \frac{1}{N}$, откуда $\forall x \in \mathbb{R}$:

$$|F_n^*(x) - F(x)| \leqslant \max_{0\leqslant k,l\leqslant N} \{|F_n^*(x_{k+1,N} - 0) - F(x_{k+1,N} - 0)|, |F_n^*(x_{l,N}) - F(x_{l,N})|\} + \frac{1}{N}$$

однако, по УЗБЧ, $F_n^*(x_{k,N}) \xrightarrow{\text{п.н.}} F(x_{k,N})$, $F_n^*(x_{k+1,N} - 0) \xrightarrow{\text{п.н.}} F(x_{k+1,N} - 0)$ откуда

$$\overline{\lim}\, D_n = \overline{\lim} \sup_{x\in\mathbb{R}} |F_n^*(x) - F(x)| \leqslant \frac{1}{N} \text{ почти наверное}$$

В силу произвольности N получаем, что $D_n \xrightarrow{\text{п.н.}} 0$. $\square$

Теорема 3.2. *(б/д, Колмогорова-Смирнова)*

Пусть $\{X_n\}_{n=1}^{\infty}$ *— выборка неограниченного размера из распределения с непрерывной функцией распределения* F*. Тогда*

$$\sqrt{n} \sup_{x \in \mathbb{R}} |F_n^*(x) - F(x)| \xrightarrow{d} \xi$$

где ξ *имеет распределение Колмогорова, т.е.*

$$F_\xi(x) = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2x^2},\ x > 0$$

3.3 Ядерные оценки плотности

В данном разделе будем считать, что $\mathcal{P}$ это все абсолютно-непрерывные распределения, $P \in \mathcal{P}$ — неизвестное распределение, имеющее плотность p.

Пусть $X_1,\ \ldots,\ X_n$ — выборка из распределения P

Определение 3.4. Пусть Q — некоторое распределение вероятностей с плотностью $q(x)$. Тогда если $q(x)$ симметрична относительно 0, то $q(x)$ называется *ядром.*

Пример 3.5. $q(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ — гауссово ядро.

Пример 3.6. $q(x) = \frac{1}{2} I(|x| \leqslant 1)$ — прямоугольное ядро.

Пример 3.7. $q(x) = (1 - |x|) I(|x| \leqslant 1)$ — треугольное ядро.

Пример 3.8. $q(x) = \frac{3}{4}(1 - x^2) I(|x| \leqslant 1)$ — ядро Епанечникова.

Определение 3.5. Рассмотрим выборку $X_1,\ \ldots,\ X_n$ из неизвестного распределения P. Вероятностная мера $\tilde{\mathsf{P}_n}$, заданная по правилу

$$\tilde{P_n}(B) = \frac{1}{n} \sum_{i=1}^{n} Q\left(\frac{B - X_i}{h_n}\right)$$

где $\frac{B - X_i}{h_n} = \{\frac{x - X_i}{h_n} \,\big|\, x \in B\}$ и $h_n \to 0,\ h_n > 0$ называется *сглаженным эмпирическим распределением.*

Сглаженное эмпирическое распределение обладает следующим набором свойств:

1. $\tilde{\mathsf{P}_n}$ имеет плотность $\tilde{p_n}(x) = \frac{1}{nh_n} \sum\limits_{i=1}^{n} q\left(\frac{x - x_i}{h_n}\right)$

2. $\tilde{\mathsf{P}_n}$ — свертка распределений P_n^* и $Q(\frac{B}{h_n})$

3. Пусть $\alpha = \int\limits_{\mathbb{R}} q^2(x)dx < +\infty,\ h_n \to 0,\ nh_n \to +\infty$ и $p(x)$ — непрерывна и ограничена. Тогда $\tilde{p_n}(x) = p_n(x) + \frac{\xi_n}{\sqrt{nh_n}}$, где $p_n(x) = \mathsf{E}\tilde{p_n}(x) = \frac{1}{h_n} \int\limits_{\mathbb{R}} q\left(\frac{x - y}{h_n}\right) p(y)dy$ и $\xi_n(x) \xrightarrow{d} \mathcal{N}(0,\ \alpha p(x))$

4 Параметические распределения. Оценки параметров

4.1 Определение и свойства оценок

Рассмотрим $(\mathscr{X}, \mathscr{B}(\mathscr{X}), \mathcal{P})$, где $\mathcal{P} = \{\mathsf{P}_\theta,\ \theta \in \Theta\}$ — все параметризованные распределения (например, все нормальные распределения или экспоненциальные распределения).

Определение 4.1. Пусть $S : \mathscr{X} \to \Theta$ — измеримая функция, такая что $S(X)$ — статистика. Тогда $S(X)$ называется оценкой параметра θ.

Если $S : \mathscr{X} \to \tau(\Theta)$ — измеримая функция, такая что $S(X)$ — статистика, то $S(X)$ — оценка параметра $\tau(\theta)$.

Определение 4.2. Пусть $X_1, \ldots, X_n$ — выборка из неизвестного распределения P_θ. Оценка $\theta^* = S(X)$ называется *несмещенной*, если $\forall \theta \in \Theta$

$$\mathsf{E}_\theta \theta^* = \theta$$

где запись E_θ подразумевает, что математическое ожидание зависит от параметра θ.

Пример 4.1. Рассмотрим оценку $\overline{X}$. $\mathsf{E}_\theta \overline{X} = \frac{1}{n} \sum \mathsf{E}_\theta X_i = \mathsf{E}_\theta X_1$, а значит $\overline{X}$ это несмещенная оценка параметра $\tau(\theta) = \mathsf{E}_\theta X_1$.

Определение 4.3. Очевидно, что при различных n (размерах выборки) оценка $\theta_n^* = \theta^*(X_1, \ldots, X_n)$ принимает различные значения. Рассмотрим последовательность оценок $\{\theta_n^*\}_{n=1}^{\infty}$. Оценка θ^* называется *состоятельной (сильно состоятельной)* если

$$\forall \theta \in \Theta :\ \theta_n^* \xrightarrow{\mathsf{P}_\theta} \theta \ (\theta_n^* \xrightarrow{\mathsf{P}_\theta \text{ п.н.}} \theta\ )$$

где символ P_θ означает, что вероятность событий зависит от конкретного значения θ.

Пример 4.2. Оценка $\overline{X}$ является состоятельной оценкой по ЗБЧ для $\mathsf{E}_\theta X_1$, и даже сильно состоятельной оценкой для $\mathsf{E}_\theta X_1$ по УЗБЧ

Факт. *Сильно состоятельные оценки являются состоятельными.*

Определение 4.4. Оценка θ^* является *асимптотически нормальной* оценкой θ, если

$$\sqrt{n}\,(\theta_n^*(X_1, \ldots, X_n) - \theta) \xrightarrow{d_\theta} \mathcal{N}(0,\ \sigma^2(\theta))$$

Функция $\sigma^2(\theta)$ называется *асимптотической дисперсией.*

Верно и аналогичное определение в многомерном случае, с той лишь разницей, что случайный вектор слева сходиться к случайному вектору $\xi \sim \mathcal{N}(0,\ \Sigma(\theta))$, но в данном курсе мы будем рассматривать лишь одномерный случай.

Пример 4.3. $\sqrt{n}\left(\overline{X} - \mathsf{E}_\theta X_1\right) \xrightarrow{d_\theta} \mathcal{N}(0,\ \mathsf{D}_\theta X_1)$ по ЦПТ

Утверждение 4.1.1. Пусть оценка θ^* является асипмтотически нормальной оценкой параметра θ. Тогда оценка θ^* — состоятельная.

Доказательство.

$$\left.\begin{aligned} &\sqrt{n}\,(\theta^* - \theta) \xrightarrow{d_\theta} \xi \sim \mathcal{N}(0,\ \sigma^2(\theta)) \\ &\tfrac{1}{\sqrt{n}} \to 0 \end{aligned}\right\} \Rightarrow \text{ по лемме Слуцкого } \theta^* - \theta \xrightarrow{d_\theta} 0 \Rightarrow \theta^* - \theta \xrightarrow{\mathsf{P}_\theta} 0$$

□

Утверждение 4.1.2. Пусть θ^* — (сильно) состоятельная оценка параметра θ, $\tau : \Theta \to E$ — непрерывная функция. Тогда $\tau(\theta^*)$ — (сильно) состоятельная оценка параматера $\tau(\theta)$.

Доказательство. Прямое следствие теоремы о наследовании сходимости. □

Утверждение 4.1.3. Пусть θ^* — асимптотически нормальная оценка θ с асимптотической дисперсией $\sigma^2(\theta)$, а $\tau : \Theta \to E$ — дифференцируемая функция (мы считаем, что $\Theta \subset \mathbb{R}$). Тогда $\tau(\theta^*)$ — асимптотически нормальная оценка $\tau(\theta)$ с асимптотической дисперсией $\sigma^2(\theta)\big[\tau'(\theta)\big]^2$

Доказательство. Применим утверждение 1.0.2 для $h = \tau, b_n = \frac{1}{\sqrt{n}}, \xi_n = \sqrt{n}\,(\theta^* - \theta) \xrightarrow{d_\theta} \eta \sim \mathcal{N}(0, \sigma^2(\theta))$ и $a = \theta$. Имеем:

$$\frac{h(a + \xi_n b_n) - h(a)}{b_n} = \sqrt{n}\,(\tau(\theta^*) - \tau(\theta)) \xrightarrow{d_\theta} \mathcal{N}\left(0,\ \sigma^2(\theta)\big[\tau'(\theta)\big]^2\right)$$

□

Пример 4.4. $X_1, \ldots, X_n$ — выборка из экспоненциального распределения с неизвестным параметром $\theta > 0$. По ЦПТ выполнена сходимость

$$\sqrt{n}\left(\overline{X} - \frac{1}{\theta}\right) \xrightarrow{d_\theta} \mathcal{N}\left(0, \frac{1}{\theta^2}\right)$$

Рассмотрим функцию $\tau(x) = \frac{1}{x}$, дифференцируемую на $(0,\ +\infty) = \Theta$. Применяя утверждение 4.1.3, получаем

$$\sqrt{n}\left(\frac{1}{\overline{X}} - \theta\right) \xrightarrow{d_\theta} \mathcal{N}\left(0,\ \frac{1}{\theta^2}\Big[-\frac{1}{x^2}\Big]^2\Bigg|_{\frac{1}{\theta}}\right) = \mathcal{N}(0, \theta^2)$$

что означает, что оценка $\frac{1}{\overline{X}}$ является асимптотически нормальной оценкой параметра θ с дисперсией θ^2.

4.2 Методы нахождения оценок

1) Метод подстановки

Рассмотрим функцию G, такую что $G(\mathsf{P}_\theta) = \theta$. Предположим, что мы знаем такую функцию G в явном виде. Тогда сделаем оценку $\theta^* = G_n(\mathsf{P}_\theta^n)$. Такой метод называется *методом подстановки.*

Пример 4.5. Пусть $X_1,\, \dots,\, X_n$ — выборка из $\mathcal{N}(\theta, 1)$. Плотность такого распределения

$$p_\theta = \frac{1}{2\pi} \exp\big[-\frac{(x-\theta)^2}{2}\big]$$

Тогда $(x-\theta)^2 = -2\ln(\sqrt{2\pi}p_\theta)$ и значение θ явно выражается.

Однако, зачастую такой метод непременим в виду сложности функции G, поэтмоу рассмотрим другие методы.

2) Метод моментов

Будем считать, что $\Theta \subset \mathbb{R}^k$. Рассмотрим борелевские функции $g_1,\, \dots,\, g_k$, действующие из $\mathbb{R}$ в $\mathbb{R}$, такие что функция $m(\theta)$, заданная по правилу

$$m(\theta) = (\mathsf{E}_\theta g_1(X_1),\, \dots,\, \mathsf{E}_\theta g_k(X_1))$$

является биекцией с обратной функцией m^{-1}.

Найдем $m^{-1}\begin{pmatrix} g_1(\overline{X}) \\ \dots \\ g_k(\overline{X}) \end{pmatrix} = \theta^*$ — это и будет оценкой для θ, полученной *методом моментов*

Замечание. Часто $g_k(x) = x^k$ — *стандартные пробные функции.* Иногда стоит рассматривать в качестве функций g_i индикаторы.

Пример 4.6. Пусть $X_1,\, \dots,\, X_n$ — выборка из гамма распределения с параметрами $(\alpha,\, \lambda)$, $g_1(x) = x$, $g_2(x) = x^2$. В таком случае

$$\mathsf{E}_\theta X_1 = \int\limits_0^{+\infty} x\frac{\alpha^\lambda e^{-\alpha x}}{\Gamma(\lambda)}dx = \frac{\Gamma(\lambda+1)}{\Gamma(\lambda)\alpha}\int\limits_0^{+\infty}\frac{x^\lambda \alpha^{\lambda+1}e^{-\alpha x}}{\Gamma(\lambda)}dx = \frac{\lambda}{\alpha}$$

$$\mathsf{E}_\theta X_1^2 = \int\limits_0^{+\infty} x^2\frac{\alpha^\lambda e^{-\alpha x}}{\Gamma(\lambda)}dx = \frac{\Gamma(\lambda+2)}{\Gamma(\lambda)}\frac{1}{\alpha^2} = \frac{\lambda(\lambda+1)}{\alpha^2}$$

Тогда $m(\theta) = \begin{pmatrix} \lambda/\alpha \\ \lambda(\lambda+1)/\alpha^2 \end{pmatrix}$ и $\theta = (\alpha, \lambda)$. Решим систему

$$\begin{cases} \dfrac{\lambda^*}{\alpha^*} = \overline{X} \\ \dfrac{\lambda(\lambda+1)}{\alpha^{*2}} \end{cases} \Rightarrow \begin{cases} \alpha^* = \dfrac{\overline{X}}{s^2} \\ \lambda^* = \dfrac{(\overline{X})^2}{s^2} \end{cases}$$

Установим несколько важных свойств оценки, полученной методом моментов

Утверждение 4.2.1. Пусть m^{-1} непрерывна на $m(\Theta)$. Тогда оценка, полученная методом моментов, является сильно состоятельной.

Доказательство. По УЗБЧ $\overline{g_i(X)} \xrightarrow{\mathsf{P}_\theta \text{ п.н.}} \mathsf{E}_\theta g_i(X)$, откуда $\begin{pmatrix} \overline{g_1(X)} \\ \dots \\ \overline{g_k(X)} \end{pmatrix} \xrightarrow{\mathsf{P}_\theta \text{ п.н.}} m(\theta)$, а значит, по теореме о наследовании сходимости $m^{-1}\begin{pmatrix} \overline{g_1(X)} \\ \dots \\ \overline{g_k(X)} \end{pmatrix} \xrightarrow{\mathsf{P}_\theta \text{ п.н.}} \theta$. □

Утверждение 4.2.2. *(б/д)* Пусть $\Theta \subset \mathbb{R}$, m^{-1} дифференцируема на $m(\Theta)$ и существует $\mathsf{E}_\theta\left[\,(g_1(X_1))^2\,\right]$. Тогда оценка θ^*, полученная по методу моментов, является а.н.о. параметра θ.

Замечание. Оценка по методу моментов не обязательно является несмещенной.

3) Метод выборочных квантилей

Определение 4.5. Рассмотрим распределение вероятностей P на $\mathbb{R}$ с функцией распределения F и число $p \in (0,\ 1)$. Тогда *квантилем уровня* p называется число

$$z_p := \min\{x,\ F(x) \geqslant p\}$$

В случае, если F непрерывна, $z_p = F^{-1}(p)$. Если F разрывна, то либо $z_p = F^{-1}(p)$, либо, если $F^{-1}(p)$ не существует, то существует точка z, в которой у F разрыв, такая что $F(z-0) < p$, $F(z+0) > p$. В таком случае $z_p = z$.

Определение 4.6. Рассмотрим выборку $X_1,\ \dots,\ X_n$ из распределения P. *Выборочным квантилем уровня p называется* число

$$z_p^* := \begin{cases} X_{(np)} & np \in \mathbb{Z} \\ X_{(\lfloor np \rfloor + 1)} & np \notin \mathbb{Z} \end{cases}$$

Теорема 4.1. *(б/д)*

Пусть f — плотность распределения P, причем f — непрерывно дифференцируема на $\mathbb{R}$, $p \in (0,\ 1)$, $f(z_p) > 0$. Тогда z_p^* — асимптотически нормальная оценка z_p с асимптотической дисперсией $\frac{p(1-p)}{f^2(z_p)}$

Определение 4.7. *Медианой* называется число $\mu = z_{\frac{1}{2}}$. Для выборки $X_1,\ \dots,\ X_n$ выборочной медианой называется число μ^*, равное $X_{(k+1)}$, если $n = 2k+1$ и равное $\frac{X_{(k)} + X_{(k+1)}}{2}$ для $n = 2k$.

Теорема 4.2. *Пусть f — плотность распределения P, причем f — непрерывно дифференцируема на $\mathbb{R}$ $f(\mu) > 0$. Тогда μ^* — асимптотически нормальная оценка μ с асимптотической дисперсией* $\frac{1}{4f^2(\mu)}$

Пример 4.7. Пусть $X_1,\ \dots,\ X_n$ — выборка из распределения Коши со сдвигом θ, $f(x) = \frac{1}{\pi(1+(x-\theta)^2)}$. Нетрудно заметить, что плотность симметрична относительно θ, а значит $F(\theta) = \frac{1}{2}$ и θ является медианой $\mu = \theta$.

Тогда μ^* это а.н.о. θ с а.д. $\frac{\pi^2}{4}$

Пример 4.8. Пусть $X_1, \ldots, X_n$ — выборка из $\mathcal{N}(\theta, 3)$. Найдем оценки для θ по методу моментов и методу квантилей: по методу моментов это $\overline{X}$, а по методу квантилей: μ^*. Для $\theta^* = \overline{X}$ а.д. равна 3. $p(\theta) = \frac{1}{\sqrt{6\pi}}$, а значит а.д. $\theta^* = \mu^*$ равна $\frac{3\pi}{2}$.

5 Способы сравнения статистик

5.1 Сравнения произвольных оценок

Определение 5.1. Пусть $\theta \in \Theta$ — оцениваемый параметр, а θ^* — его оценка. Тогда функция $g : \Theta^2 \to \mathbb{R}_+$ называется *функцией потерь*, а $\mathsf{E}_\theta g(\theta^*,\ \theta)$ — *функцией риска* для функции потерь g.

Замечание. Как правило, $g(x,y) = |x-y|$ или $g(x,y) = (x-y)^2$. В многомерном случае часто $g(x,y) = \langle A(x-y),(x-y)\rangle$, где A — некоторая неотрицательно определенная матрица.

Определение 5.2. Пусть $\mathcal{K}$ — некоторый класс оценок. Оценка $\theta^* \in \mathcal{K}$ называется *наилучшей* в классе $\mathcal{K}$, если она лучше всех других оценок из $\mathcal{K}$.

Существует несколько подходов определения какая из двух оценок является лучшей. Приведем здесь их.

1) Равномерный подход

Определение 5.3. Оценка θ^* лучше оценки $\widehat{\theta}$, если $\forall \theta \in \Theta :\ \mathsf{E}_\theta g(\theta^*,\ \theta) \leqslant \mathsf{E}_\theta g(\widehat{\theta},\ \theta)$ и хотя бы для одного $\theta \in \Theta$ неравенство строгое.

Утверждение 5.1.1. В классе всевозможных оценок $\mathcal{K}$ нет наилучшей в равномерном подходе. (считаем $g(x,y) = (x-y)^2$ или $|x-y|$)

Доказательство. Поскольку класс $\mathcal{K}$ содержит константы, то достаточно рассмотреть их. Действительно, зафиксируем $\theta_0 \in \Theta$ и рассмотрим оценку $\theta^* = \theta_0$. Любая другая оценка либо совпадает с θ^* на θ, либо хуже нее на θ_0, а любая другая оценка-константа θ_1^* лучше оценки θ^* на $\theta_1 \neq \theta_0$. □

2) Байесовский подход

Определение 5.4. Пусть Q — некоторое распределение вероятностей на Θ. Тогда оценка θ^* лучше оценки $\widehat{\theta}$ в байесовском подходе, если для любого $\theta \in \Theta$ выполнено неравенство $\mathsf{E}_Q g(\theta^*,\ \theta) \leqslant \mathsf{E}_Q g(\widehat{\theta},\ \theta)$.

Очевидно, что если оценка является наилучшей в равномерном подходе, то она является лучшей и в байесовском. Обратное же неверно.

3) Минимаксный подход

Определение 5.5. Оценка θ^* лучше оценки $\widehat{\theta}$, если $\sup\limits_{\theta\in\Theta} g(\theta^*,\theta) < \sup\limits_{\theta\in\Theta} g(\widehat{\theta},\theta)$

5.2 Поиск наилучшей оценки в классе несмещенных оценок

В этом разделе используется равномерный подход с функцией потерь $g(x,y) = (x-y)^2$.

Рассмотрим сначала некоторое дискретное распределение P (будем считать б.о.о, что P определено на $\mathbb{Z}_+$).

Определение 5.6. Положим $\mathsf{P}(B) = \sum\limits_{k\in B\cap\mathbb{Z}_+} \mathsf{P}(\{k\}) = \sum\limits_{k\in B\cap\mathbb{Z}_+} p(k) =: \int\limits_B p(x)\mu(dx)$, где $\mu(dx)$ — *считающая мера*, т.е. $\mu : \mathscr{B}(\mathbb{R}) \to \mathbb{Z}_+ \cup \{+\infty\}$ и $\mu(B) = |B \cap \mathbb{Z}|$.

Определение 5.7. Семейство распределений $\mathcal{P}$ *доминируемо* относительно меры μ, если

1. либо все распределения абсолютно непрерывные и μ — мера Лебега,
2. либо все распределения дискретные и μ — считающая мера.

Будем считать для таких семейств, что $\mathsf{P}(B) = \int\limits_B p(x)\mu(dx)$.

Далее считаем, что имеющееся семество распределений $\mathcal{P}$ — доминируемо относительно некоторой меры μ и $X_1, \ldots, X_n = X$ — выборка из исследуемого распределения $\mathsf{P}_\theta \in \mathcal{P}$ с плотностью $p_\theta(x)$.

Определение 5.8. Функция $u_\theta(x) = \frac{\partial}{\partial\theta} \ln p_\theta(x)$ называется *вкладом* наблюдения x, а функция $I_X(\theta) = \mathsf{E}_\theta \left[u_\theta(X)\right]^2$ — *информацией Фишера*

Введем *условия регулярности*

R1: Θ — открытый интервал (возможно, бесконечный) в $\mathbb{R}$.

R2: Множество $A = \{x \in \mathbb{R} \mid p_\theta(x) > 0\}$ не зависит от $\theta \in \Theta$.

R3: $\theta \in \Theta$ и для любой статистики $S(X)$ с конечным вторым моментом справедливо дифференцирование под знаком интеграла, т.е. верно равенство:

$$\frac{\partial}{\partial\theta}\mathsf{E}_\theta S(x) = \mathsf{E}_\theta \left[S(x)\frac{\partial}{\partial\theta}\ln p_\theta(x)\right]$$

обосновать которое можно так:

$$\frac{\partial}{\partial\theta}\int\limits_{\mathbb{R}} S(x)p_\theta(x)\mu(dx) = \int\limits_A S(x)\frac{\partial}{\partial\theta}p_\theta(x)\cdot\frac{1}{p_\theta(x)}p_\theta(x)dx =$$
$$= \int\limits_{\mathbb{R}} S(x)\left[\frac{\partial}{\partial\theta}\ln p_\theta(x)\right]p_\theta(x)dx = \mathsf{E}_\theta\left[S(x)\frac{\partial}{\partial\theta}\ln p_\theta(x)\right]$$

R4: $0 < I_X(\theta) < +\infty$ — информация Фишера существует, конечна и отлична от 0.

Теорема 5.1. *(неравенство Рао-Крамера)*

Пусть выполнены условия регулярности **R1-R4**, τ *— дифференцируемая на* Θ *функция и* $\widehat{\theta}$ *— несмещенная оценка параметра* $\tau(\theta)$*. Тогда выполнено неравенство*

$$\mathsf{D}_\theta\widehat{\theta} \geqslant \frac{\left[\tau'(\theta)\right]^2}{I_X(\theta)} \ \forall\theta\in\Theta$$

Доказательство. Рассмотрим статистику $S(X) = 1$. Используя **R3**, имеем

$$0 = \frac{\partial}{\partial\theta}\mathsf{E}_\theta 1 = \mathsf{E}_\theta\frac{\partial}{\partial\theta}\ln p_\theta(x) \Rightarrow \mathsf{E}_\theta u_\theta(X) = 0 \tag{1}$$

Применим теперь **R3** для статистики $S(X) = \widehat{\theta}$. Помня, что эта оценка несмещенная, имеем:

$$\tau'(\theta) = \mathsf{E}_\theta \widehat{\theta} u_\theta(X) \tag{2}$$

Вычтем из второго равенства первое, домноженное на $\tau(\theta)$:

$$\tau'(\theta) = \mathsf{E}_\theta \left[\widehat{\theta} - \tau(\theta)\right] u_\theta(X)$$

возведем в квадрат и воспользуемся неравенством Коши-Буняковского:

$$\left(\tau'(\theta)\right)^2 = \left(\mathsf{E}_\theta \left[\widehat{\theta} - \tau(\theta)\right] u_\theta(X)\right)^2 \leqslant \mathsf{E}_\theta \left[\widehat{\theta} - \tau(\theta)\right]^2 \mathsf{E}_\theta u_\theta(X)^2 = \mathsf{D}_\theta \widehat{\theta} I_X(\theta)$$

откуда следует требуемое неравенство. □

Следствие 5.1.1. *Наилучшей оценкой является та, для которой достигается равенство.*

Определение 5.9. Если $\forall \theta \in \Theta$ для несмещенной оценки $\widehat{\theta}$ параметра $\tau(\theta)$ в неравенстве Рао-Крамера достигается равенство, то оценка $\widehat{\theta}$ называется *эффективной*.

Теорема 5.2. *(критерий эффективности)*

В условиях неравенства Рао-Крамера оценка θ^* является эффективной оценкой параметра $\tau(\theta)$ $\iff$ $\theta^* - \tau(\theta) = c(\theta) \cdot u_\theta(X) \iff c(\theta) = \frac{\tau'(\theta)}{I_X(\theta)}$

Доказательство. Заметим, что равенство в Рао-Крамере $\iff$ равенство в КБШ $\iff$ случайные величины для которых применяется КБШ — линейно зависимы, т.е. $\theta^* - \theta = c(\theta)u(\theta) + a(\theta)$. Используя несмещенность θ^*, получаем $\forall \theta \in \Theta: \ 0 = \mathsf{E}_\theta a(\theta) = a(\theta) = 0$.

Имеем теперь

$$u_\theta(X)\left[\theta^* - \tau(\theta)\right] = c(\theta)\left(u_\theta(X)\right)^2$$

Посчитав мат.ожидание обеих частей равенства, справа имеем $\tau'(\theta)$ аналогично док-ву неравенства Рао-Крамера, а слева $c(\theta)I_X(\theta)$, а значит равенство возможно только при $c(\theta) = \frac{\tau'(\theta)}{I_X(\theta)}$ □

Следствие 5.2.1. *Если есть оценка $\widehat{\theta}$ не хуже θ^*, то к ней можно применить те же рассуждения и получить, что $\theta^* = \widehat{\theta}$.*

Следствие 5.2.2. *Эффективная оценка является наилучшей в классе несмещенных оценок в равномерном подходе с квадратичной функцией потерь.*

Исследуем D_θ на сходимость.

Пусть $X = (X_1, \ldots, X_n)$. Тогда

$$\begin{aligned} I_{(X_1, \ldots, X_n)}(\theta) &= \mathsf{E}_\theta u_\theta(X_1, \ldots, X_n)^2 = \mathsf{D}_\theta u_\theta(X_1, \ldots, X_n) = \\ &= \mathsf{D}_\theta \sum_{i=1}^n u_\theta(X_i) = \sum_{i=1}^n \mathsf{D}_\theta u_\theta(X_i) = n\mathsf{D}_\theta u_\theta(X_1) = nI_{X_1}(\theta) = ni(\theta) \end{aligned}$$

где $i(\theta)$ — информация Фишера одного элемента выборки. Взяв $\tau(\theta) = \theta$, имеем $\mathsf{D}_\theta \theta^* \geqslant \frac{1}{I_X(\theta)} = \frac{1}{ni(\theta)}$, а значит $\mathsf{D}_\theta \theta^* \to 0$ как $\frac{1}{n}$

6 Оценка максимального правдоподобия

Рассмотрим семейство параметрических распределений $\mathcal{P} = \{\mathsf{P}_\theta,\ \theta \in \Theta\}$, доминируемое относительно меры μ, и p_θ — плотность P_θ.

Определение 6.1. Пусть $X_1,\ \ldots,\ X_n$ — выборка из P_θ. Тогда *правдоподобием* называется функция $f_\theta(X) = \prod\limits_{i=1}^{n} p_\theta(X_i)$

Определение 6.2. Оценка $\theta^* = \arg\max f_\theta(X_1,\ \ldots,\ X_n)$ называется *оценкой максимального правдоподобия.*

Пример 6.1. Рассмотрим $\Theta = \mathbb{N}$ и $\mathsf{P}_\theta = U\{1,\ \ldots,\ \theta\}$. Тогда функция правдоподобия равна

$$f_\theta(X_1,\ \ldots,\ X_n) = \frac{I(X \in \{1,\ \ldots,\ \theta\}^n)}{\theta^n}$$

Откуда $\theta^* = X_{(n)}$.

Пример 6.2. $\Theta = \mathbb{R} \times (0,\ +\infty)$ и $X_1,\ \ldots,\ X_n$ — выборка из $\mathcal{N}(a,\ \sigma^2)$. Функция правдоподобия $f_\theta = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left[-\frac{\sum(X_i - a)^2}{2\sigma^2}\right]$. Как видно, функция правдоподобия устроена довольно трудно, поэтому часто имеет смысл рассматривать *логарифмическую функцию правдоподобия* $L_\theta = \ln f_\theta$. Тогда

$$L_\theta = -\frac{n}{2}\ln 2\pi\sigma^2 - \frac{\sum(X_i - a)^2}{2\sigma^2}$$

Найдем производные

$$\frac{\partial \ln f_\theta}{\partial a} = \frac{\sum(X_i - a)}{\sigma^2} = n\frac{\overline{X} - a}{\sigma^2} \tag{3}$$

$$\frac{\partial \ln f_\theta}{\partial \sigma^2} = -\frac{n}{2}\frac{2\pi}{2\pi\sigma^2} + \frac{\sum(X_i - a)^2}{2\sigma^4} = \frac{\sum(X_i - a)^2 - n\sigma^2}{2\sigma^4} \tag{4}$$

откуда о.м.п. $\theta^* = (a^*,\ \sigma^{2*}) = \left(\overline{X},\ \frac{\sum(X_i - a)^2}{n}\right)$

С этого момента считаем, что $\mathcal{P} = \{\mathsf{P}_\theta,\ \theta \in \Theta\}$ — произвольное семейство распределений, доминируемое относительно меры μ, плотность P_θ равна p_θ и если $\theta_1 \neq \theta_2$ то $\mathsf{P}_{\theta_1} \neq \mathsf{P}_{\theta_2}$. Введем *условия регулярности*

R1: Множество $A = \{x \in \mathbb{R} \mid p_\theta(x) > 0\}$ не зависит от θ.

R2: Будем считать, что $X_1,\ \ldots,\ X_n$ — выборка из $\mathsf{P} \in \mathsf{P}$.

R3: Θ — открытый интервал в $\mathbb{R}$ (возможно, бесконечный).

R4: Функция $p_\theta(x)$ дифференцируема по θ на множестве A.

R5: Функция $p_\theta(x)$ трижды непрерывно дифференцируема по $\theta\, \forall x \in A$.

R6: Интеграл $\int\limits_A p_\theta(x)\mu(dx)$ трижды дифференцируемый по θ под знаком интеграла.

R7: $\mathsf{E}_\theta \left[\frac{\partial}{\partial\theta} \ln p_\theta(X_1)\right]^2 = i(\theta) \in (0, +\infty)$.

R8: $\forall\theta_0 \in \Theta\ \exists c > 0\ \exists H(x): \forall\theta \in (\theta_0 - c,\ \theta_0 + c): \left|\frac{\partial^3}{\partial\theta^3} \ln p_\theta(x)\right| < H(x)$ и $\mathsf{E}_\theta H(X_1) < +\infty$

Теорема 6.1. *Пусть выполнены условия регулярности* **R1-R2**. *Тогда* $\forall\theta_0 \neq \theta \in \Theta:$

$$\mathsf{P}_{\theta_0}\left(f_{\theta_0}(X_1, \ldots, X_n) > f_\theta(X_1, \ldots, X_n)\right) \xrightarrow{n\to\infty} 1$$

Доказательство.

$$\begin{aligned}
\{f_{\theta_0}(X_1, \ldots, X_n) > f_\theta(X_1, \ldots, X_n)\} &= \left\{\frac{f_\theta(X_1, \ldots, X_n)}{f_{\theta_0}(X_1, \ldots, X_n)} < 1\right\} \\
&= \left\{\ln \frac{\prod p_\theta(X_1, \ldots, X_n)}{\prod p_{\theta_0}(X_1, \ldots, X_n)} < 0\right\} \\
&= \left\{\frac{1}{n}\sum \ln \frac{p_\theta(X_i)}{p_{\theta_0}(X_i)} < 0\right\}
\end{aligned}$$

По УЗБЧ, $\frac{1}{n}\sum \ln \frac{p_\theta(X_i)}{p_{\theta_0}(X_i)} \xrightarrow{\mathsf{P}_\theta \text{ п.н.}} \mathsf{E}_{\mathsf{P}_{\theta_0}} \ln \frac{p_\theta(X_1)}{p_{\theta_0}(X_1)}$

$$\begin{aligned}
\mathsf{E}_{\theta_0} \ln \frac{p_\theta(x)}{p_{\theta_0}(x)} &= \int\limits_A \ln \frac{p_\theta(x)}{p_{\theta_0}} p_{\theta_0}(x)\mu(dx) \\
&\leqslant \int\limits_A \left(\frac{p_\theta(x)}{p_{\theta_0}(x)} - 1\right) p_{\theta_0}\mu(dx) \\
&= \int\limits_A \left(p_\theta(x) - p_{\theta_0}(x)\right)\mu(dx) = 1 - 1 = 0
\end{aligned}$$

где мы воспользовались неравенством $\ln(1+x) \leqslant x$. Равенство в оценке достигается при $\frac{p_\theta(x)}{p_{\theta_0}(x)} - 1 = 0 \Rightarrow p_\theta(x) = p_{\theta_0}(x)$ равенство при всех x или при x из множества меры 0, очевидно, противоречит условию $\mathsf{P}_\theta \neq \mathsf{P}_{\theta_0}$, а значит $\mathsf{E}_{\theta_0} \ln \frac{p_\theta(x)}{p_{\theta_0}(x)} < 0$ — если оно существует.

Рассмотрим $f = \ln \frac{p_\theta(x)}{p_{\theta_0}(x)} \cdot p_{\theta_0}(x);\ \ g = p_\theta(x) - p_{\theta_0}(x)$. В доказательстве мы показали, что $f \leqslant g \Rightarrow f^+ \leqslant g^+,\ f^- \geqslant g^-$, причем $\mathsf{E}g = 0 \Rightarrow \mathsf{E}f = \mathsf{E}f^+ - \mathsf{E}f^- \leqslant \mathsf{E}g^+ - \mathsf{E}g^- = 0$, а значит рассматриваемое мат.ожидание действительно существует и либо конечно, либо равно $-\infty$. В конечном случае применяем УЗБЧ, а случай, когда мат.ожидание равно $-\infty$, примем без доказательства. □

Теорема 6.2. *Пусть выполнены* **R1-R4** *и* $\forall n\ \forall x_1, \ldots, x_n$ *существует единственное решение* θ^* *уравнения* $\frac{\partial}{\partial\theta} \ln f_\theta(x_1, \ldots, x_n) = 0$. *Тогда* θ^* *это состоятельная оценка параметра* θ *и* $\forall\theta \in \Theta:$ $\mathsf{P}_\theta(\theta^*$ *– О.М.П.*$) \to 1$ *при* $n \to \infty$.

Утверждение 6.0.1. Если $|\Theta| < \infty$ и при фиксированных $X_1, \ldots, X_n$ найдется $\arg\max\limits_{\theta\in\Theta} f_\theta(X_1, \ldots, X_n)$ то существует оценка максимального правдоподобия θ^*.

Доказательство. Предположим, что θ^* не является состоятельной. Тогда $\exists \varepsilon > 0 \exists \delta > 0 \exists \theta \in \Theta$ такие, что

$$\begin{cases} \mathsf{P}_\theta(|\theta^* - \theta| > \varepsilon) > \delta \quad \forall n \in \mathbb{N} \\ \forall \theta_0 \in \Theta: \ \mathsf{P}_{\theta_0}(f_{\theta_0} > f_\theta) \to 1 \end{cases} \quad \Rightarrow \text{ противоречие.}$$

□

Теорема 6.3. *(б/д)*

Пусть выполнены **R1-R8** *и* $\forall n\ \forall x_1, \ldots, x_n$ *существует единственное решение* θ^* *уравнения* $\frac{\partial}{\partial \theta} \ln f_\theta(x_1, \ldots, x_n) = 0$*. Тогда* θ^* *является асимптотически нормальной оценкой* θ *с асимптотической дисперсией* $\frac{1}{i(\theta)}$ *— непрерывной в силу* **R5** *и* **R7**.

Более того, если $\widehat{\theta}$ *— асимптотически нормальная оценка* θ *с асимптотической дисперсией* $\sigma^2(\theta)$ *и* σ^2 *непрерывна на* Θ*, то* $\sigma^2(\theta) \geqslant \frac{1}{i(\theta)}$ *— непрерывной в силу* **R5** *и* **R7**.

Определение 6.3. Оценка θ^* называется *асимптотически эффективной*, если она является наилучшей в асимптотическом подходе в классе асимптотически нормальных оценок с непрерывной асимптотической дисперсией.

Теорема 6.4. *Пусть выполнены условия из неравенства Рао-Крамера. Тогда эффективная оценка является оценкой максимального правдоподобия.*

Доказательство. Пусть θ^* — эффективная оценка $\Rightarrow$ $\theta^* - \theta = \frac{1}{i(\theta)} \frac{\partial}{\partial \theta} \ln f_\theta$. Поскольку $i(\theta) > 0$ по определнию, имеем

$$\frac{\partial}{\partial \theta} \ln f_\theta > 0 \iff \theta^* > \theta$$
$$\frac{\partial}{\partial \theta} \ln f_\theta < 0 \iff \theta^* < \theta$$

что и означает, что θ^* это о.м.п. □

7 Условное математическое ожидание

7.1 Определение и свойства

Пусть ξ — случайная величина на $(\Omega,\, \mathscr{F},\, \mathsf{P})$, а $\mathcal{G} \subset \mathscr{F}$ — некоторая σ-алгебра.

Определение 7.1. *Условным математическим ожиданием* ξ при условии $\mathcal{G}$ называется случайная величина $\mathsf{E}(\xi|\mathcal{G}) = \eta$, для которой выполнены следующие свойства:

1. (измеримость) η является $\mathcal{G}$ измеримой случайной величиной.

2. (интегральное условие) $\forall A \in \mathcal{G}\ \mathsf{E}\xi I_A = \mathsf{E}\eta I_A$

Определение 7.2. Функция ν называется *зарядом* на $(\Omega,\, \mathscr{F},\, \mathsf{P})$, если $\nu : \mathscr{F} \to \mathbb{R}$ — счетно-аддитивная функция и $\forall A \in \mathscr{F} :\ |\nu(A)| < +\infty$.

Определение 7.3. Заряд ν называется *абсолютно непрерывным* относительно меры P, если

$$\mathsf{P}(A) = 0 \Rightarrow \nu(A) = 0$$

Замечание. Понятие абсолютной непрерывности как свойства функции или меры носит гораздо более общий характер. Например, распределение вероятностей в абсолютно непрерывном случае является абсолютно непрерывным относительно меры Лебега, поскольку $\mathsf{P}(B) = \int\limits_B g(x)dx$, где g — это плотность распределения P.

Теорема 7.1. *(б/д, Радона-Никодима)*

Если ν — заряд, абсолютно непрерывный относительно меры P, то существует единственная P-п.н. случайная величина η на $(\Omega,\, \mathscr{F},\, \mathsf{P})$, такая что

$$\forall A \in \mathscr{F} :\ \nu(A) = \mathsf{E}\eta I_A = \int\limits_A \eta(\omega)\mathsf{P}(d\omega)$$

Теорема 7.2. *Если ξ — случайная величина на $(\Omega, \mathscr{F}, \mathsf{P})$, такая что $\mathsf{E}|\xi| < +\infty$, а $\mathcal{G} \subset \mathscr{F}$ — некоторая σ-алгебра, то существует $\mathsf{E}(\xi|\mathcal{G})$ единственное P-п.н.*

Доказательство. Рассмотрим функцию $\nu(A) = \mathsf{E}\xi I_A$ для любого множества $A \in \mathcal{G}$. По определению это заряд, абсолютно непрерывный относительно меры P, а значит, по теореме Радона-Никодима, $\exists!$ случайная величина η на $(\Omega,\, \mathcal{G})$, такая что $\nu(A) = \mathsf{E}\eta I_A$.

Осталось заметить, что η является $\mathcal{G}$-измеримой случайной величиной, поскольку задана на $(\Omega,\, \mathcal{G})$, а это значит, что η — искомое условное математическое ожидание. $\square$

Утверждение 7.1.1. Пусть $\mathcal{G} = \sigma(D_1,\, \dots,\, D_n, \dots)$, $\bigsqcup\limits_{i=1}^{+\infty} D_i = \Omega$ и $\forall i\ \mathsf{P}(D_i) > 0$. Тогда верна формула

$$\mathsf{E}(\xi|\mathcal{G}) = \sum_{i=1}^{+\infty} \frac{\mathsf{E}\xi I_{D_i}}{\mathsf{P}(D_i)} I_{D_i}$$

Доказательство. Обозначим $\eta := \mathsf{E}(\xi|\mathcal{G})$. Покажем сначала, что на любом множестве из разбиения η равна константе.

Предположим противное. Тогда, без ограничения общности, $\exists \omega_1, \omega_2 \in D_1 : \eta(\omega_1) = c_1 \neq c_2 = \eta(\omega_2)$. Рассмотрим множество $\eta^{-1}(\{c_1\}) \cap D_1 = A$. Оно лежит в $\mathcal{G}$ поскольку η — $\mathcal{G}$-измеримая величина, и оно отлично от D_1 и $\varnothing$ поскольку в нем лежит ω_1 и не лежит ω_2. Однако, так как $\mathcal{G} = \sigma(D_1, \ldots)$ — объединение конечного и бесконечного числа множеств D_i, то A не может лежать в $\mathcal{G}$ — противоречие, т.е. $\mathsf{E}(\xi|\mathcal{G}) = \sum\limits_{i=1}^{\infty} c_i I_{D_i}$.

Воспользуемся интегральным свойством у.м.о. для $A = D_i$. Имеем

$$\mathsf{E}\xi I_A = \mathsf{E}\eta I_A = \mathsf{E}\left(\sum_{j=1}^{\infty} c_j I_{D_j}\right) I_{D_i} = \mathsf{E} c_i I_{D_i} = c_i \mathsf{P}(D_i)$$

откуда следует требуемое утверждение. □

Пример 7.1. Предположим, что мы бросаем кубик и ξ — количество очков, выпавшее на кубике. Пусть $D_1 = \{1, 3, 5\}$ и $D_2 = \{2, 4, 6\}$ — разбиение Ω. Тогда $\mathsf{E}(\xi|\sigma(D_1, D_2)) = \frac{\mathsf{E}\xi I_{D_1}}{\mathsf{P}(D_1)} I_{D_1} + \frac{\mathsf{E}\xi I_{D_2}}{\mathsf{P}(D_2)} I_{D_2} = \frac{3}{2} I_{D_1} + \frac{4}{2} I_{D_2}$.

Докажем некоторые свойства условных математических ожиданий.

Утверждение 7.1.2. $\mathsf{E}(\mathsf{E}(\xi|\mathcal{G})) = \mathsf{E}\xi$

Доказательство. Воспользуемся интегральным свойством для $A = \Omega$: $\mathsf{E}\xi = \mathsf{E}\xi I_A = \mathsf{E}\big(\mathsf{E}(\xi|\mathcal{G}) I_A\big) = \mathsf{E}\left(\mathsf{E}(\xi|\mathcal{G})\right)$ □

Утверждение 7.1.3. Если ξ — $\mathcal{G}$-измеримая, то $\mathsf{E}(\xi|\mathcal{G}) = \xi$ почти наверное — очевидно из определения.

Утверждение 7.1.4. Если $\mathscr{F}_\xi \perp\!\!\!\perp \mathcal{G}$, то $\mathsf{E}(\xi|\mathcal{G}) = \mathsf{E}\xi$ п.н.

Доказательство. Пусть $A \in \mathcal{G}$. Тогда $I_A \perp\!\!\!\perp \xi \Rightarrow \mathsf{E}\xi I_A = \mathsf{P}(A)\mathsf{E}\xi$. Поскольку $\mathsf{E}\xi$ — число, то оно измеримо относительно любой σ-алгебры. Тогда, по интегральному свойству для $\eta = \mathsf{E}\xi$ имеем

$$\mathsf{E}\eta I_A = \mathsf{P}(A)\mathsf{E}\xi = \mathsf{E}\xi I_A$$

□

Утверждение 7.1.5. $\mathsf{E}(a\xi + b\eta|\mathcal{G}) = a\mathsf{E}(\xi|\mathcal{G}) + b\mathsf{E}(\eta|\mathcal{G})$ п.н.

Доказательство. Пусть $\zeta := a\mathsf{E}(\xi|\mathcal{G}) + b\mathsf{E}(\eta|\mathcal{G})$ — $\mathcal{G}$-измеримая случайная величина. Проверим для нее интегральное свойство для $A \in \mathcal{G}$:

$$\mathsf{E}\zeta I_A = a\mathsf{E}\left(\mathsf{E}(\xi|\mathcal{G}) I_A\right) + b\mathsf{E}\left(\mathsf{E}(\eta|\mathcal{G}) I_A\right) = a\mathsf{E}\xi I_A + b\mathsf{E}\xi I_A = \mathsf{E}(a\xi + b\eta) I_A$$

□

Утверждение 7.1.6. Если $\xi \leqslant \eta$, то $\mathsf{E}(\xi|\mathcal{G}) \leqslant \mathsf{E}(\eta|\mathcal{G})$ п.н.

Доказательство. Пусть $A \in \mathcal{G}$. По интегральному свойству имеем:

$$\mathsf{E}\left(\mathsf{E}(\eta - \xi|\mathcal{G})\right) = \mathsf{E}(\eta - \xi)I_A \geqslant 0$$

откуда, поскольку это верно для любого $A \in \mathcal{G}$, из курса теории вероятностей, следует, что $\mathsf{E}(\eta-\xi|\mathcal{G}) \geqslant 0$ п.н. □

Утверждение 7.1.7. (Телескопическое свойство)

Пусть $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathscr{F}$. Тогда

$$\mathsf{E}\left(\mathsf{E}(\xi|\mathcal{G}_1)|\mathcal{G}_2\right) = \mathsf{E}(\xi|\mathcal{G}_1) \text{ п.н.} \tag{1}$$

$$\mathsf{E}\left(\mathsf{E}(\xi|\mathcal{G}_2)|\mathcal{G}_1\right) = \mathsf{E}(\xi|\mathcal{G}_1) \text{ п.н.} \tag{2}$$

Доказательство. Поскольку $\mathsf{E}(\xi|\mathcal{G}_1)$ является $\mathcal{G}_2$-измеримой, то равенство один выполнено по утверждению 7.1.3.

Пусть $\eta := \mathsf{E}(\xi|\mathcal{G}_1)$ — $\mathcal{G}_1$-измерима по определнию. По интегральному свойству, для любого $A \in \mathcal{G}_1 \subset \mathcal{G}_2$ выполнено равенство

$$\mathsf{E}\eta I_A = \mathsf{E}\xi I_A = \mathsf{E}\left(\mathsf{E}(\xi|\mathcal{G}_2)I_A\right)$$

откуда $\eta = \mathsf{E}\left(\mathsf{E}(\xi|\mathcal{G}_2)|\mathcal{G}_1\right)$ п.н. по определению. □

Утверждение 7.1.8. (б/д, аналог теоремы Лебега)

Пусть $\xi_n \xrightarrow{\text{п.н.}} \xi$ — случайные величины на $(\Omega, \mathscr{F}, \mathsf{P})$ и $|\xi_n| \leqslant \eta$ для любого $n \in \mathbb{N}$ и $\mathsf{E}\eta < +\infty$. Тогда для любой σ-алгебры $\mathcal{G} \subset \mathscr{F}$ выполнена сходимость $\mathsf{E}(\xi_n|\mathcal{G}) \xrightarrow{\text{п.н.}} \mathsf{E}(\xi|\mathcal{G})$.

Утверждение 7.1.9. Пусть ξ, η — случайные величины, такие что $\mathsf{E}|\xi\eta| < +\infty$, $\mathsf{E}|\eta| < +\infty$ и η является $\mathcal{G}$-измеримой. Тогда $\mathsf{E}(\xi\eta|\mathcal{G}) = \eta\mathsf{E}(\xi|\mathcal{G})$.

Доказательство. Пусть сначала $\eta = I_A,\ A \in \mathcal{G}$. Тогда для любого $B \in \mathcal{G}$ по интегральному свойству выполнено:

$$\mathsf{E}\eta\mathsf{E}(\xi|\mathcal{G})I_B = \mathsf{E}\left(\mathsf{E}(\xi|\mathcal{G})I_{A\cap B}\right) = \mathsf{E}\xi I_{A\cap B} = \mathsf{E}\xi I_A I_B = \mathsf{E}\xi\eta I_B$$

откуда по линейности получаем требоемое равенство для простых случайных величин.

Пусть η — произвольная случайная величина, и $\eta_n \xrightarrow{\text{п.н.}} \eta$ — последовательность простых, такая что $|\eta_n| < |\eta|$. Тогда $\xi\eta_n \xrightarrow{\text{п.н.}} \xi\eta$, $|\xi\eta_n| < |\xi\eta|$ и $\mathsf{E}|\xi\eta| < +\infty$. По свойству 7.1.8 имеем

$$\mathsf{E}(\xi\eta_n|\mathcal{G}) \xrightarrow{\text{п.н.}} \mathsf{E}(\xi\eta|\mathcal{G})$$

$$\mathsf{E}(\xi\eta_n|\mathcal{G}) = \eta_n\mathsf{E}(\xi|\mathcal{G}) \xrightarrow{\text{п.н.}} \eta\mathsf{E}(\xi|\mathcal{G})$$

□

Теорема 7.3. *(б/д, о наилучшем среднеквадратичном прогнозе)*

Пусть $\mathcal{G} \subset \mathscr{F}$ и ξ — случайная величина на $(\Omega, \mathscr{F}, \mathsf{P})$, $\mathcal{L} = \{\mathcal{G}$-измеримые с.в. с конечным мат. ожиданием $\}$. Тогда выполнено равенство:

$$\arg\min_{\eta\in\mathcal{L}} \mathsf{E}(\xi - \eta)^2 = \mathsf{E}(\xi|\mathcal{G}) \text{ п.н.}$$

7.2 Поиск УМО в абсолютно непрерывном случае

Обозначим

$$\mathsf{E}(\xi|\eta) = \mathsf{E}\left(\xi \mid \mathscr{F}_\eta\right)$$

$$\mathsf{P}(A|\mathcal{G}) = \mathsf{E}(I_A|\mathcal{G})$$

$$\mathsf{P}(A|\eta) = \mathsf{E}(I_A|\mathscr{F}_\eta)$$

Определение 7.4. $\mathsf{E}(\xi|\eta = y) = \varphi(y)$, где φ — борелевская функця, такая что $\forall B \in \mathscr{B}(\mathbb{R})$:

$$\mathsf{E}\xi I(\eta \in B) = \int\limits_B \varphi(y)\mathsf{P}_\eta(dy) = \mathsf{E}\varphi(\eta)I(\eta \in B) = \int\limits_{\omega:\eta(\omega)\in B} \varphi(\eta(\omega))\mathsf{P}(d\omega).$$

Лемма 7.1. *(б/д)*

$\mathsf{E}(\xi|\eta = y) = \varphi(y)$ *тогда и только тогда, когда* $\mathsf{E}(\xi|\eta) = \varphi(\eta)$.

Из теоремы Радона-Никодима следует, что $\mathsf{E}(\xi|\eta = y)$ существует и единственно почти наверное.

В случае, когда ξ и η обе дискретные и $\mathsf{P}(\eta = y) \neq 0$, имеем

$$\mathsf{E}(\xi \mid \eta = y) = \sum x\mathsf{P}(\xi = x|\eta = y) = \int xp_{(\xi|\eta)}(x \mid y)dx.$$

Определение 7.5. Условным распределением ξ при условии η называется $\mathsf{P}_\xi(B \mid \eta) = \mathsf{E}(I(\xi \in B) \mid \eta)$.

Определение 7.6. Функция $p_{(\xi|\eta)}(x \mid y) \geqslant 0$ называется условной плотностью ξ при условии η, если для любых $B \in \mathscr{B}(\mathbb{R})$, $y \in \mathbb{R}$ выполнено равенство

$$\mathsf{P}_\xi(B|\eta = y) = \int\limits_B p_{(\xi|\eta)}(x \mid y)dx.$$

Утверждение 7.2.1. Пусть g — борелевская функция, ξ, η — случайные величины на $(\Omega, \mathscr{F}, \mathsf{P})$, $\mathsf{E}|g(\xi)| < \infty$ и $p_{(\xi|\eta)}(x \mid y)$ — условная плотность. Тогда $\mathsf{E}(g(\xi) \mid \eta = y) = \int\limits_{\mathbb{R}} g(x)p_{(\xi|\eta)}(x \mid y)dx$

Доказательство. Достаточно доказать, что

$$\forall B \in \mathscr{B}(\mathbb{R}): \ \mathsf{E}g(\xi)I(\eta = B) = \int\limits_B \left(\int\limits_{\mathbb{R}} g(x)p_{(\xi|\eta)}(x \mid y)dx\right)\mathsf{P}_\eta(dy).$$

Пусть $g = I_A$ и $A \in \mathscr{B}(\mathbb{R})$.

$\mathsf{E}I(\xi \in A)I(\eta \in B) = \mathsf{P}(\xi \in A,\, \eta \in B)$. Перепишем интеграл

$$\begin{aligned}
\int\limits_B \left(\int\limits_{\mathbb{R}} I(x \in A)p_{(\xi|\eta)}(x \mid y)dx\right)\mathsf{P}_\eta(dy) &= \int\limits_B \left(\int\limits_A p_{(\xi|\eta)}(x \mid y)dx\right)\mathsf{P}_\eta(dy) \\
&= \int\limits_B \mathsf{P}_\xi(A \mid \eta = y)\mathsf{P}_\eta(dy) \\
&= \int\limits_B \mathsf{E}\left(I(\xi \in A) \mid \eta = y\right)\mathsf{P}_\eta(dy) \\
&= \mathsf{E}I(\xi \in A)I(\eta \in B)
\end{aligned}$$

Для простых случайных величин утверждение следует из линейности мат.ожидания. Произвольную случайную величину можно приблизить простыми и воспользоваться теоремой Лебега о мажорируемой сходимости. $\square$

Теорема 7.4. *Если существует плотность $p_{(\xi,\eta)}(x, y)$, то существует и условная плотность*

$$p_{(\xi|\eta)}(x \mid y) = \begin{cases} \frac{p_{(\xi,\eta)}(x,y)}{p_\eta(y)} & p_\eta(y) \neq 0, \\ 0 & p_\eta(y) = 0 \end{cases}.$$

Доказательство. Достаточно доказать, что $\forall B \in \mathscr{B}(\mathbb{R})$:

$$\mathsf{P}_\xi(B \mid \eta = y) = \int\limits_B \frac{p_{(\xi,\eta)}(x, y)}{p_\eta(y)} I(p_\eta(y) \neq 0) dx$$

Рассмотрим $A, B \in \mathscr{B}(\mathbb{R})$. С одной стороны

$$\mathsf{P}(\xi \in B, \eta \in A) = \mathsf{E}I(\xi \in B,\ \eta \in A) = \mathsf{E}I(\xi \in B) I(\eta \in A) =$$
$$= \int\limits_A \mathsf{P}_\xi(B \mid \eta = y) \mathsf{P}_\eta(dy) = \int\limits_A \mathsf{P}_\xi(B \mid \eta = A) p_\eta(y) dy$$

А с другой

$$\mathsf{P}(\xi \in B, \eta \in A) = \int\limits_{B\times A} p_{(\xi,\eta)}(x, y) dx dy =$$
$$= \int\limits_A \left[\int\limits_B p_{(\xi,\eta)}(x, y) dx \right] dy = \int\limits_A \left[\int\limits_B \frac{p_{(\xi,\eta)}(x, y)}{p_\eta(y)} I(p_\eta(y) \neq 0) dx \right] p_\eta(y) dy,$$

откуда следует утверждение теоремы. $\square$

Алгоритм вычисления УМО в абсолютно непрерывном случае.

Пусть даны случайные величины ξ, η с совместной плотностью $p_{(\xi,\eta)}(x, y)$. Мы хотим найти значение $\mathsf{E}(g(\xi) \mid \eta)$.

1. Считаем условную плотность $p_{(\xi|\eta)}(x \mid y)$.

2. Находим функцию φ, для которой $\varphi(y) = \mathsf{E}\left(g(\xi) \mid \eta = y\right) = \int\limits_{\mathbb{R}} g(x) p_{(\xi|\eta)}(x \mid y) dx$.

3. $\mathsf{E}(g(\xi) \mid \eta) = \varphi(\eta)$.

7.3 Поиск наилучшей оценки в классе несмещенных оценок

Определение 7.7. Пусть зафиксирован класс распределений $\mathcal{P} = \{\mathsf{P}_\theta \mid \theta \in \Theta\}$ на $(\Omega,\ \mathscr{F},\ \mathsf{P})$. Тогда σ-алгебра $\mathcal{G} \subset \mathscr{F}$ называется *достаточной σ-алгеброй*, если $\forall A \in \mathscr{F}$ величина $\mathsf{P}_\theta(A \mid \mathcal{G})$ не зависит от θ.

Определение 7.8. Пусть X — наблюдение из распределения $\mathsf{P} \in \mathcal{P} = \{\mathsf{P}_\theta\}$. Тогда статистика $S(X)$ называется *достаточной*, если $\forall B \in \mathscr{B}(\mathbb{R}^n)$ величина $\mathsf{P}_\theta(X \in B \mid S(X))$ не зависит от θ.

Теорема 7.5. *(Критерий факторизации Неймана-Фишера)*

Пусть $\mathcal{P} = \{\mathsf{P}_\theta \mid \theta \in \Theta\}$ — класс распределений, доминируемых относительно μ с плотностью f_θ. Тогда $S(X)$ является достаточной статистикой тогда и только тогда, когда $f_\theta(x) = h(x)g_\theta(S(X))$ для некоторый функций h и g.

Доказательство. Рассмотрим только дискретный случай.

Пусть статистика $S(X)$ — достаточная. Тогда

$$f_\theta(x) = p_\theta(X = x) = \mathsf{P}_\theta(X = x, S(X) = S(x)) = \underbrace{\mathsf{P}_\theta(X = x \mid S(X) = S(x)}_{h(x)}\underbrace{\mathsf{P}_\theta(S(X) = S(x))}_{g(\theta, S(X))}.$$

Пусть наоборот, $\mathsf{P}_\theta(X = x) = h(x)g_\theta(S(x))$. Тогда

$$\mathsf{P}_\theta(X = x \mid S(X) = 1) = \begin{cases} 0 & S(x) \neq 1, \\ \mathsf{P}_\theta(X = x \mid S(X) = S(x)) & S(X) = 1 \end{cases},$$

откуда

$$\mathsf{P}_\theta(X = x \mid S(X) = S(x)) = \frac{\mathsf{P}_\theta(X = x, S(X) = S(x))}{\mathsf{P}(S(X) = S(x))} = \frac{\mathsf{P}_\theta(X = x, S(X) = S(x))}{\sum\limits_{y: S(y) = S(x)} \mathsf{P}_\theta(S(X) = S(x), X = y)} =$$

$$= \frac{\mathsf{P}_\theta(X = x)}{\sum\limits_{y: S(y) = S(x)} \mathsf{P}_\theta(x = y)} = \frac{h(x)g_\theta(S(x))}{\sum\limits_{y: S(y) = S(x)} h(y)g_\theta(S(x))} = \frac{h(x)}{\sum\limits_{y: S(y) = S(x)} h(y)}$$

□

Теорема 7.6. *(Рао-Блэквелла-Колмогорова)*

Пусть $\widehat{\theta}$ — несмещенная оценка θ и $\forall \theta : \mathsf{D}_\theta \widehat{\theta} < +\infty$, а $S(X)$ — достаточная статистика. Тогда для оценки $\theta^* = \mathsf{E}\left(\widehat{\theta} | S(X)\right)$ верно:

1. *θ^* не зависит от θ (как функция)*

2. *θ^* — несмещенная оценка θ*

3. *$\mathsf{D}_\theta \theta^* \leqslant \mathsf{D}_\theta \widehat{\theta}$, причем равенство $\forall \theta \in \Theta \iff \theta^* = \widehat{\theta}$ P_θ почти наверное*

Доказательство. 1. Следствие из определения достаточной статистики.

2. $\mathsf{E}_\theta \theta^* = \mathsf{E}_\theta\left(\mathsf{E}_\theta\left(\widehat{\theta} | S(X)\right)\right) = \mathsf{E}_\theta \widehat{\theta} = \theta$

3.

$$\mathsf{D}_\theta(\widehat{\theta}) = \mathsf{E}_\theta(\widehat{\theta} - \theta)^2 = \mathsf{E}_\theta\big[(\widehat{\theta} - \theta^*) + (\theta^* - \theta)\big]^2 = \mathsf{E}_\theta(\widehat{\theta} - \theta^*)^2 + \mathsf{D}_\theta \theta^* + 2\mathsf{E}_\theta\Big[(\widehat{\theta} - \theta^*)(\theta^* - \theta)\Big] = \mathsf{D}_\theta \theta^* + \underbrace{\mathsf{E}_\theta(\widehat{\theta} - \theta^*)^2}_{\geqslant 0}$$

поскольку

$$\mathsf{E}_\theta\Big[(\widehat{\theta}-\theta^*)(\theta^*-\theta)\Big] = \mathsf{E}_\theta\left[\mathsf{E}\left((\widehat{\theta}-\theta^*)(\theta^*-\theta) \mid S(X)\right)\right] = \mathsf{E}_\theta\Big[(\theta^*-\theta)\mathsf{E}\left(\widehat{\theta}-\theta^* \mid S(X)\right)\Big] = 0$$

причем $\mathsf{D}_\theta\widehat{\theta} = \mathsf{D}_\theta\theta^* \Leftrightarrow \mathsf{E}_\theta(\widehat{\theta}-\theta^*)^2 = 0 \Leftrightarrow \widehat{\theta} = \theta^*$ P_θ п.н. □

Определение 7.9. Статистика $S(X)$ называется *полной*, если для любой борелевской функции f из условия, что $\forall\theta \in \Theta:\ \mathsf{E}_\theta f(S(X)) = 0$ следует, что $f(S(X)) = 0$ P_θ п.н. $\forall\theta \in \Theta$.

Лемма 7.2. *Если $S(X)$ — полная достаточная статистика и для некоторой функции φ верно равенство $\mathsf{E}_\theta\varphi(S(X)) = \theta\ \forall\theta \in \Theta$, то тогда $\varphi(S(X))$ — оптимальная оценка θ.*

Доказательство. В силу теоремы БКР достаточно доказать, что $\varphi(S(X))$ — единственная $S(X)$-измеримая несмещенная оценка θ.

Пусть существует другая $S(X)$-измеримая несмещенная оценка $\psi(S(X))$. Тогда $\forall\theta \in \Theta$:

$$\mathsf{E}_\theta\varphi(S(X)) = \mathsf{E}_\theta\psi(S(X)) = \theta$$
$$\mathsf{E}_\theta\left(\varphi(S(X)) - \psi(S(X))\right) = 0$$
$$\mathsf{E}_\theta(\varphi-\psi)(S(X)) = 0$$

откуда, из определения полноты статистики, следует, что $\varphi - \psi = 0$ почти наверное. □

Алгоритм нахождения оптимальной оценки

1. Находим достаточную оценку $S(X)$
2. Проверяем ее на полноту
3. Если статистика полная, то решаем для φ уравнение $\mathsf{E}_\theta\varphi(S(X)) = \theta\ \ \forall\theta \in \Theta$

Определение 7.10. Пусть $\theta \in \Theta \subseteq \mathbb{R}^k$, X — наблюдение с плотностью p_θ из распределения $\mathsf{P} \in \mathcal{P}$, доминируемого относительно некоторой меры. Пусть $p_\theta(X)$ имеет вид

$$p_\theta(x) = h(x)\exp\left(\sum_{i=1}^{k} a_i(\theta)u_i(X) + b(\theta)\right)$$

где $u_1,\ \ldots,\ u_k$ — борелевские функции. Тогда семейство распределений $\mathcal{P}$ принадлежит *экспоненциальному классу распределений.*

Теорема 7.7. *(б/д)*

Пусть X — наблюдение из $\mathsf{P} \in \{P_\theta \mid \theta \in \Theta\}$, принадлежащего экспоненциальному классу распределений. Пусть кроме того множество $\{(a_1(\theta),\ \ldots,\ a_k(\theta))\}$ содержит k-мерный параллелепипед. Тогда статистика $(u_1(X),\ \ldots,\ u_k(X))$ — полная достаточная статистика.

Замечание. Зачастую достаточно проверить, чтобы функции $a_1, \ldots a_k$ были л.н.з. и Θ содержало в себе открытое множество.

Пример 7.2. Пусть $X_1, \ldots, X_n \sim \mathcal{N}(a, \sigma^2)$. Найдем оптимальную оценку для $a^2 + \sigma^2$.

Статистика $S(X) = (\sum X_i^2, \sum X_i)$ является достаточной, причем $\mathsf{E}_\theta \sum X_i^2 = n(a^2 + \sigma^2)$, откуда получаем, что $\overline{X^2}$ — оптимальная оценка для $a^2 + \sigma^2$.

8 Доверительные интервалы

8.1 Построение доверительных интервалов методом центральной статистики

Определение 8.1. Пусть X — наблюдение из $\mathsf{P} \in \{\mathsf{P}_\theta \mid \theta \in \Theta\}$ и $\Theta \subset \mathbb{R}$. *Доверительным интервалом уровня* $\gamma \in (0, 1)$ называется такая пара статистик $(T_1(X),\, T_2(X))$, что

$$\forall \theta \in \Theta \;\; \mathsf{P}_\theta\left(\theta \in (T_1(X),\, T_2(X))\right) \geqslant \gamma$$

если $\forall \theta$ достигается равенство, то интервал называется *точным.*

Замечание. Обычно рассматриваются д.и. уровня $\gamma = 0.9,\ 0.95,\ 0.98,\ 0.99$.

Приведем один из методов построения доверительных интервалов: *метод центральной статистики.*

Определение 8.2. Пусть X — наблюдение из P. Случайная величина $G(X,\, \theta)$, распределение которой не зависит от θ, называется *центральной статистикой.*

Зафиксируем числа $1 > \gamma_2 > \gamma_1 > 0$ и $\gamma_2 - \gamma_1 = \gamma$. Пусть $z_{\gamma_1},\ z_{\gamma_2}$ — квантили уровней γ_1 и γ_2 распределения $G(X,\, \theta)$ соответственно. Тогда выполнено неравенство

$$\mathsf{P}_\theta(z_{\gamma_1} \leqslant G(X,\, \theta) \leqslant z_{\gamma_2}) \geqslant \gamma_2 - \gamma_1 = \gamma$$

причем равенство достигается в случае, когда для функции распределения G существуют две точки непрерывности $x_1,\ x_2$, такие что $F_G(x_1) = \gamma_1,\ F_G(x_2) = \gamma_2$.

Пусть $T_i(X)$ — решения уравнений $G(X,\, T_i(X)) = z_{\gamma_i}$ для $i = 1,\ 2$. Тогда

$$\mathsf{P}_\theta(T_1(X) < \theta < T_2(X)) = \mathsf{P}_\theta(z_{\gamma_1} < G(X,\, \theta) < z_{\gamma_2}) \geqslant \gamma$$

Пример 8.1. Пусть $X_1,\, \ldots,\, X_n \sim \mathcal{N}(b,\, \sigma^2)$. Тогда $\frac{X_1 - b}{\sigma} \sim \mathcal{N}(0, 1)$ и $\frac{1}{\sqrt{n}} \sum \frac{X_i - b}{\sigma} \sim \mathcal{N}(0, 1)$.

Пусть $z_{\frac{1-\gamma}{2}},\ z_{\frac{1+\gamma}{2}}$ — квантили уровней $\frac{1-\gamma}{2},\ \frac{1+\gamma}{2}$ распределения $\mathcal{N}(0,\ 1)$. Тогда

$$\mathsf{P}\left(z_{\frac{1-\gamma}{2}} \leqslant \sqrt{n}\frac{\overline{X} - b}{\sigma} \leqslant z_{\frac{1+\gamma}{2}}\right) = \gamma$$

Выражая отсюда b или σ, получаем доверительный интервал для этих параметров уровня γ.

Лемма 8.1. *Пусть у случайной величины X непрерывная функция распределения F и $X_1,\, \ldots,\, X_n$ — н.о.р. случайные величины. Тогда*

$$-\sum \ln F(X_i) \sim \Gamma(1,\ n)$$

Доказательство. $\mathsf{P}(F(X_1) \leqslant y) = \mathsf{P}(X_1 \leqslant F^{-1}(y)) = F(F^{-1}(y)) \Rightarrow F(X_i) \sim U[0,\ 1]$. Тогда $-\ln F(X_i) \sim Exp(1) \Rightarrow -\sum \ln F(X_i) \sim \Gamma(1, n)$ □

Следствие 8.0.1. *Пусть* $\mathcal{P} = \{\mathsf{P}_\theta \mid \theta \in \Theta\}$ *такое семейство распределений, что* $\forall\theta$ P_θ *имеет непрерывную функцию распределения. Тогда* $-\sum \ln F(X_i)$ *— центральная статистика с распределением* $\Gamma(1,\, n)$

8.2 Асимптотические доверительные интервалы

Определение 8.3. Пусть $X_1, \ldots, X_n$ — выборка из распределения $\mathsf{P} \in \{\mathsf{P}_\theta \mid \theta \in \Theta \subset \mathbb{R}\}$. Тогда последовательность пар статистик $\left(T_1^n(X), T_2^n(X)\right)$ называется *асимптотическим доверительным интервалом уровня* γ, если

$$\varliminf_{n\to+\infty} \mathsf{P}_\theta\left(\theta \in \left[T_1^n(X),\, T_2^n(X)\right]\right) \geqslant \gamma$$

Асимптотический доверительным интервал называется *точным*, если равенство обращается в равенство, а $\varliminf$ превращается в lim.

Пример 8.2. Пусть $X_1, \ldots, X_n$ имеет распределение P_θ с $\mathsf{E}_\theta X = \theta$, $\mathsf{D}_\theta X = \sigma^2(\Theta) > 0$ — непрерывная функция. По ЦПТ выполнена сходимость

$$\sqrt{n}\frac{\overline{X}-\theta}{\sigma(\theta)} \xrightarrow{d_\theta} \mathcal{N}(0,1)\ \forall\theta$$

По ЗБЧ $\overline{X} \xrightarrow{\mathsf{P}_\theta \text{ п.н.}} \theta$ откуда, по теореме о наследовании сходимости, $\sigma(\overline{X}) \xrightarrow{\mathsf{P}_\theta \text{ п.н.}} \sigma(\theta)$ Тогда

$$\sqrt{n}\frac{\overline{X}-\theta}{\sigma(\overline{X})} = \underbrace{\frac{\sqrt{n}(\overline{X}-\theta)}{\sigma(\theta)}}_{\to\mathcal{N}(0,1)} \cdot \underbrace{\frac{\sigma(\theta)}{\sigma(\overline{X})}}_{\to 1} \xrightarrow{d_\theta} \mathcal{N}(0,\,1) \text{ по л. Слуцкого}$$

Тогда для $z_{\frac{1+\gamma}{2}}$ — квантиль $\mathcal{N}(0,1)$ уровня $\frac{1+\gamma}{2}$ верно

$$\mathsf{P}_\theta\left(-z_{\frac{1+\gamma}{2}} \leqslant \sqrt{n}\frac{\overline{X}-\theta}{\sigma(\overline{X})} \leqslant z_{\frac{1+\gamma}{2}}\right) = \mathsf{P}_\theta\left(\underbrace{\overline{X} - z_{\frac{1+\gamma}{2}}\frac{\sigma(\overline{X})}{\sqrt{n}}}_{T_1} \leqslant \theta \leqslant \underbrace{\overline{X} + z_{\frac{1+\gamma}{2}}\frac{\sigma(\overline{X})}{\sqrt{n}}}_{T_2}\right) \to \gamma\ \forall\theta$$

Замечание. $T_2 - T_1 \to 0$

9 Байесовские методы

9.1 Введение

Напоминание: Пусть $(\Omega,\ \mathscr{F},\ \mathsf{P})$ — вероятностное пространство, $\{D_n\}$ — разбиение $\Omega, A \in \mathscr{F}$. Тогда формула Байеса имеет вид

$$\mathsf{P}(D_n \mid A) = \frac{\mathsf{P}(A \mid D_n)\mathsf{P}(D_n)}{\sum\limits_{i=0}^{\infty} \mathsf{P}(A \mid D_i)\mathsf{P}(D_i)} \tag{3}$$

Определение 9.1. Назовем A — *результатом эксперимента*, $\mathsf{P}(D_n)$ — *априорная* вероятность — известная *до* эксперимента. $P(D_n \mid A)$ — *апостериорная* вероятность — *после* эксперимента.

Пусть $\xi,\ \eta$ — случайные величины на $(\Omega,\ \mathscr{F},\ \mathsf{P})$. Тогда формула Байеса в общем виде:

$$p_{\eta|\xi}(y \mid x) = \frac{p_{\xi|\eta}(x \mid y)p_\eta(y)}{\int\limits_{\mathbb{R}^n} p_{\xi|\eta}(x \mid y)p_\eta(y)} \tag{4}$$

История становления Байесовских методов в статистике:

1763: опубликована работа Байеса с формулой 3.

1812: получена современная формула Байеса 4.

1920: Фишер нашел оптимальную оценку ОМП, после чего байесовские методы оказались забыты.

1990: Возраждение байесовских методов.

2010: Начало активного использования баесовских методов в BigData.

2017: Лекция по байесовскиим методам на ПМИ ФИВТ.

Замечание. Баесовские методы в BigData используются, например, в задаче распознования лиц на фотографии или работе со словами, имеющими несколько смысловых значений, в word2vec.

9.2 Математическое описание байесовских методов. Сравнение подходов

Пусть θ — случайный вектор, имеющий распределение Q, доминируемое относительно некоторой меры, с плотностью $q(t)$ и $\theta \in \Theta \subset \mathbb{R}^d$. Пусть X — наблюдение из распределения $\mathsf{P} \in \mathcal{P} = \{\mathsf{P}_t : t \in \Theta\}$, где t — значение случайного вектора θ и P_t имеет плотность $p_t(x)$. Тогда функция

$$f(t,x) = q(t)p_t(x)$$

есть плотность вектора $(\theta,\ X)$.

Определение 9.2. Плотность $q(t)$ называется *априорной плотностью*, а $q(t \mid x)$, определяемая по формуле

$$q(t \mid x) = \frac{q(t)p_t(x)}{\int\limits_{\Theta} q(s)p_s(x)dx},$$

называется *апостериорной* плотностью.

Способы оценивания θ.

1. Апостреорное распределение — это оценивания θ целым распределением вероятностей, откуда получаются последующие оценки.

2. Интервальные оценки: пусть u_p — квантиль апостериорного распределения. Тогда доверительный интервал для θ есть $(u_{(1-\alpha)/2}, u_{(1+\alpha)/2})$.

3. Точечные оценки:

   (a) $\mathsf{E}(\theta \mid X)$ — математическое ожидание по апостериорному распределению.

   (b) $\arg\max\limits_{t \in \Theta} q(t \mid x)$ — мода априорного распределения.

Подходы	Частотный	Байесовский
Интерпритация случайности	Никакая случайная величина никем не прогнозируема (объективная неопределенность)	Любая случайная величина — детерминированный процесс, но часть фактов скрыта от нас (субъективное незнание)
Величины	Четкое деление на случайные величины и параметры	Все случайно (в понимании выше)
Основной метод вывода	Оценка максимального правдоподобия	Формула Байеса
Типы оценок	Точечные и интервальные	Апостериорное распределение
Корректность методов	Верны при $n \to +\infty$	Верны при $n \geqslant 0$.

Теорема 9.1. *Оценка $\mathsf{E}(\theta|X)$ — наилучшая оценка параметра θ в баесовским подходе с квадратичной функцией потерь.*

Доказательство. Нам необходимо найти оценку $\widehat{\theta}$, для которой

$$\int\limits_{\Theta} R(\widehat{\theta},\ t)q(t)dt \to \max .$$

Перепишем интеграл

$$\int\limits_{\Theta} \mathsf{E}_t \left(\widehat{\theta} - t\right)^2 q(t)dt = \int\limits_{\Theta} \int\limits_{\mathcal{X}} \left(\widehat{\theta}(x) - t\right)^2 f(t,x)dxdt = \mathsf{E}\left(\widehat{\theta}(x) - \theta\right)^2 \to \max_{\widehat{\theta}} .$$

Применяя теорему о наилучшем приближении

$$\widehat{\theta} = \mathsf{E}(\theta|X).$$

$\square$

У байесовсокго метода в статистике имеются свои недостатки. Вот самые существенные из них:

1. Предполагается, что распределение $q(t)$ задано, поскольку иначе не существует конструктивных способов выбрать его.

2. Большые вычислительные затраты.

Пример 9.1. Пусть $X_1, \ldots, X_n \sim \mathcal{N}(\theta, 1)$, $\theta \sim Cauchy$. Тогда

$$\int\limits_{\mathbb{R}} q(t)p_t(x)dt = \int\limits_{\mathbb{R}} \frac{1}{(2\pi)^{n/2}} \exp\left[-\frac{1}{2}\sum(x_i - t)^2\right] \frac{1}{\pi(1+t^2)} dt.$$

Такой интеграл достаточно тяжело посчитать аналитически, а значит нет знаменателя в формуле Байеса, что означает, что из оценок байесовским методом можно посчитать только моду.

Определение 9.3. Пусть $X_1, \ldots, X_n$ — выборка из распределения $\mathsf{P} \in \mathcal{P} = \{\mathsf{P}_t \mid t \in \Theta\}$ — некоторый класс распределений. Пусть на Θ задан класс распределений $\mathcal{Q} = \{Q_\alpha \mid \alpha \in \mathcal{A}\}$. Класс $\mathcal{Q}$ называется *сопряженным* к классу $\mathcal{P}$, если при взятии априорного распределения из класса $\mathcal{Q}$ соответствующее ему апостериорное распределение тоже лежит в классе $\mathcal{Q}$.

Пример 9.2. Пусть $X_1, \ldots, X_n \sim Exp(\theta)$. Найдем сопряженный класс распределений и байесовскую оценку. Плотность выборки $p_t(X)$ равна

$$p_t(x) = t^n e^{-t\sum X_i}.$$

Возьмем $q(t)$ пропорциональную выражению выше, где коэффицент пропорциональности не зависит от θ.

$$q(t) \propto t^{\beta-1}e^{-\alpha t} \Rightarrow q(t) = \frac{\alpha^\beta}{\Gamma(\beta)} t^{\beta-1}e^{-\alpha t}.$$

Покажем, что гамма распределение действительно сопряжено экспоненциальному:

$$q(\theta|x) \propto q(t)p_t(x) \propto t^{\beta+n-1}e^{-t(\alpha+\sum X_i)} \Rightarrow q(\theta|x) \sim \Gamma(\alpha + \sum X_i,\ \beta+n).$$

Тогда точечная байесовская оценка есть

$$\mathsf{E}(\theta \mid X) = \frac{\beta+n}{\alpha+\sum X_i}.$$

Пример 9.3. Пусть $X_1, \ldots, X_n \sim \mathcal{N}(\theta, 1)$. Тогда

$$p_t(x) \propto \exp(-poly_2(t)) \Rightarrow q(t) \propto \exp(-poly_2(t)) \Rightarrow q(t) \sim \mathcal{N}(a,\ \sigma^2).$$

В качестве упражнения можно доказать, что

$$q(t \mid X) = \mathcal{N}\left(\frac{\sum X_i + \frac{a}{\sigma^2}}{n+\frac{1}{\sigma^2}}; \frac{1}{n+\frac{1}{\sigma^2}}\right).$$

Пример 9.4. Найдем класс распределений, сопряженный экспоненциальному классу, т.е. $p_t(x) = \frac{g(x)}{h(x)} e^{-t^T u(x)}$. Для выборки имеем

$$p_t(X) \propto \frac{1}{h^n(x)} e^{-t^T \sum u(X_i)} \Rightarrow q(t) \propto h^{-\beta}(t) e^{-t\alpha} = \frac{h^{-\beta}(t)}{f(\alpha, \beta)} e^{-t^T \alpha}$$

и

$$q(t \mid X) \propto q(t) p_t(X) \propto \frac{1}{h^{\beta+n}} \exp\left[-t^T(\alpha + \sum X_i\right].$$

То есть экспоненциальный класс распределений сопряжен сам себе.

10 Линейная регрессия

10.1 Линейная модель

Начнем с некоторых примеров.

Пример 10.1. Рассмотрим следующую задачу. Пусть имеется 2 груза неизвестной массы и весы. Мы взвешиваем грузы с целью узнать их массу. Пусть мы три раза взвесили первый груз и получили веса $\{x_1,\ x_2,\ x_3\}$, пять раз взвесили второй груз с показаниями весов $\{y_1,\ \ldots,\ y_5\}$ и десять раз оба груза вместе с весами $\{z_1,\ \ldots,\ z_{10}\}$. Причем из-за погрешности измерений все числа $x_i,\ y_i,\ z_i$ различны. Условие задачи можно представить следующим образом:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ y_1 \\ \vdots \\ y_5 \\ z_1 \\ \vdots \\ z_{10} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ & \vdots & \\ 0 & 1 \\ 1 & 1 \\ & \vdots & \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} + \vec{\varepsilon},$$

где a, b — неизвестные веса грузов, а $\vec{\varepsilon}$ — вектор ошибок измерений.

Пример 10.2. Пусть случайная величина X зависит от времени по закону $a_3t^3 + a_2t^2 + a_1t + a_0$, где a_i неизвестны и необходимо найти их оценку. В разные моменты времени t_i были проведены измерения величины X и получены результаты $X_i = a_3t_i^3 + a_2t_i^2 + a_1t_i + a_0 + \varepsilon_i$. Тогда задачу можно сформулировать так:

$$\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} t_1^3 & t_1^2 & t_1 & 1 \\ & & \vdots & \\ t_n^3 & t_n^2 & t_n & 1 \end{pmatrix} \begin{pmatrix} a_3 \\ a_2 \\ a_1 \\ a_0 \end{pmatrix} + \vec{\varepsilon}.$$

Поставим задачу линейной регрессии.

Пусть $X \in \mathbb{R}^n$ — случайный вектор. Известно, что $X = l + \varepsilon$, где $l \in \mathbb{R}^n$ — не случайный вектор, а $\varepsilon \in \mathbb{R}^n$ — случайный, причем $l \in L \subset \mathbb{R}^n$, где $L = \langle z_1,\ \ldots,\ z_k \rangle$ — k-мерное подпространство в $\mathbb{R}^n$. Пусть $Z = (z_1,\ \ldots,\ z_k)$ — известная матрица и $l = Z\theta$, $\theta \in \Theta \subset \mathbb{R}^k$ — неизвестный вектор-параметр. При этом ε — это вектор-столбец из независимых одинаково распределенных случайных величин с $\mathsf{E}\varepsilon_i = 0$ и $\mathsf{D}\varepsilon_i = \sigma^2 > 0$, где σ^2 неизвестно. Задача линейной регрессии заключается в нахождении оценок θ и σ^2.

Определение 10.1. Оценка вектора l значением $\widehat{l} = \mathrm{proj}_L X$ называется оценкой методом наименьших квадратов.

$$\widehat{l} = \arg\min_{l \in L} ||X - l||^2.$$

Попробуем найти оценку для θ. Для этого преобразуем выражение выше.

$$\begin{aligned} ||X - l||^2 =& ||X - Z\theta||^2 = (X - Z\theta)^T (X - Z\theta) = \\ &= X^T X - (Z\theta)^T X - X^T (Z\theta) + (Z\theta)^T (Z\theta) = X^T X - 2X^T Z\theta + \theta^T Z^T Z\theta. \end{aligned}$$

Поскольку для $\widehat{l}$ достигается минимум, а норма это гладкая функция, то

$$\frac{\partial}{\partial \theta_i} ||X - l||^2 = -2(X^T Z)_i + 2(Z^T Z\theta)_i = 0.$$

так как равенство верно для любого i, то

$$\begin{aligned} Z^T Z\theta = (X^T Z)^T &\Rightarrow Z^T Z\theta = Z^T X \\ &\Rightarrow \widehat{\theta} = (Z^T Z)^{-1} Z^T X \text{ — оценка } \theta \text{ по методу наименьших квадратов} \\ &\Rightarrow \widehat{l} = Z\widehat{\theta}. \end{aligned}$$

Утверждение 10.1.1. Оценка $\widehat{\theta}$ несмещенная.

Доказательство.

$$\mathsf{E}_\theta \widehat{\theta} = \mathsf{E}_\theta (Z^T Z)^{-1} Z^T X = (Z^T Z)^{-1} Z^T \mathsf{E}_\theta X = (Z^T Z)^{-1} Z^T Z\theta = \theta.$$

□

Найдем дисперсию $\mathsf{D}_\theta \widehat{\theta}$:

$$\begin{aligned} \mathsf{D}_\theta \widehat{\theta} = \mathsf{D}_\theta (Z^T Z)^{-1} Z^T X = (Z^T Z)^{-1} Z^T \mathsf{D}_\theta (X) ((Z^T Z)^{-1} Z^T)^T = \\ = \sigma^2 (Z^T Z)^{-1} Z^T ((Z^T Z)^{-1} Z^T)^T = \sigma^2 (Z^T Z)^{-1} Z^T Z (Z^T Z)^{-1} = \sigma^2 (Z^T Z)^{-1} \end{aligned}$$

Утверждение 10.1.2. $\frac{1}{n-k} \mathsf{E}_\theta ||X - Z\widehat{\theta}||^2 = \sigma^2$.

Доказательство. Будем использовать следующую формулу: $\mathrm{tr}\, AB = \mathrm{tr}\, BA$. Обозначим $A := Z(Z^T Z)^{-1} Z^T$. Тогда $\mathrm{tr}\, A = \mathrm{tr}\, Z(Z^T Z)^{-1} Z^T = \mathrm{tr}(Z^T Z)^{-1} (Z^T Z) = k$.

Поскольку $\widehat{\theta}$ — несмещенная оценка, то $\mathsf{E}_\theta (X - Z\widehat{\theta}) = 0$, откуда $\mathrm{tr}\, D_\theta (X - Z\widehat{\theta}) = \mathsf{E}_\theta ||X - Z\widehat{\theta}||^2$.

$$\begin{aligned} \mathrm{tr}\, D_\theta (X - Z\widehat{\theta}) = \mathrm{tr} D_\theta (E - A) X = \mathrm{tr} \left[(E - A) \mathsf{D}_\theta X (E - A)^T \right] = \\ = \mathrm{tr} \left[(E - A)\sigma^2 \right] = n\sigma^2 - \sigma^2 \, \mathrm{tr}\, A = n\sigma^2 - \sigma^2 k = (n - k)\sigma^2. \end{aligned}$$

поскольку $A^2 = A$. □

Следствие 10.0.1. $\frac{1}{n-k} ||X - Z\widehat{\theta}||^2 = \widehat{\sigma}^2$ *— несмещенная оценка* σ^2.

10.2 Гауссовская линейная модель

Определение 10.2. Линейная модель называется *гауссовской*, если $X = l + \varepsilon$, где $l = Z\theta$ и $\varepsilon \sim \mathcal{N}(0,\, \sigma^2 E)$.

Теорема 10.1. *(б/д, об ортогональном разложении гауссовского вектора)*

Пусть $X \sim \mathcal{N}(b,\, \sigma^2 E)$. *Пусть* $\mathbb{R}^n = L_1 \oplus \ldots \oplus L_r;\ \dim L_i = k_i;\ l_i = \operatorname{proj}_{L_i} l$ *и* $X_i = \operatorname{proj}_{L_i} X$ — *ортогональные проекции вектора* X.

Тогда $X_1,\, \ldots,\, X_r$ — *независимые случайные вектора и*

$$\frac{1}{\sigma^2}||X_i - l_i||^2 \sim \chi^2_{k_i},$$

где

$$\chi^2_k = \Gamma\left(\frac{1}{2};\ \frac{k}{2}\right) \overset{d}{=} \xi_1^2 + \ldots + \xi_k^2,$$

где $\xi_i \sim \mathcal{N}(0,1)$ — *независимые одинаково распределенные.*

Рассмотрим плотность выборки:

$$p(X) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left[-\frac{\sum(X_i - l_i)^2}{2\sigma^2}\right] = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left[-\frac{||X - l||^2}{2\sigma^2}\right] =$$
$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left[-\frac{||\operatorname{proj}_L X - l||^2 + ||X - \operatorname{proj}_L X||^2}{2\sigma^2}\right]$$

откуда, по критерию Неймана-Фишера, статистика $S(X) = (\operatorname{proj}_L X;\ ||X - \operatorname{proj}_L X||)$ — достаточная.

Теорема 10.2. *(б/д)*

Статистика $(\operatorname{proj}_L X;\ ||X - \operatorname{proj}_L X||)$ — *полная.*

Следствие 10.2.1. *Оценки* $\widehat{\theta}$ *и* $\widehat{\sigma}^2$ — *оптимальные оценки* θ *и* σ^2 *соответственно.*

Доказательство. Достаточно выразить эти оценки как функции от $S(X)$, поскольку они несмещенные.

$$\widehat{\theta} = (Z^T Z)^{-1} Z^T \operatorname{proj}_L X$$
$$\widehat{\sigma}^2 = \frac{1}{n-k}||X - Z\widehat{\theta}||^2 = \frac{1}{n-k}||X - \operatorname{proj}_L X||^2$$

□

Утверждение 10.2.1. $\widehat{\theta} \perp\!\!\!\perp X - Z\widehat{\theta}$, причем $\frac{1}{\sigma^2}||X - Z\widehat{\theta}||^2 \sim \chi^2(n-k)$ и $\frac{1}{\sigma^2}||Z\widehat{\theta} - Z\theta||^2 \sim \chi^2(k)$.

Доказательство. По теореме 10.1:

$$Z\widehat{\theta} = \operatorname{proj}_L X \perp\!\!\!\perp \operatorname{proj}_{L^\perp} X = X - Z\widehat{\theta}.$$

Поскольку $\widehat{\theta} = (Z^T Z)^{-1} Z^T (Z\widehat{\theta}) \Rightarrow \widehat{\theta} \perp\!\!\!\perp X - Z\widehat{\theta}$.

Распределение статистик следует из того, что $\dim L = k$. □

Определение 10.3. Пусть $\xi \sim \mathcal{N}(0,1)$, $\eta \sim \chi^2_k$ и $\xi \perp\!\!\!\perp \eta$. Тогда случайная величина

$$\frac{\xi}{\sqrt{\frac{\eta}{k}}} \sim T_k$$

имеет *распределение Стьюдента* с k степенями свободы.

Пусть $\xi \sim \chi^2_k$, $\eta \sim \chi^2_m$, $\xi \perp\!\!\!\perp \eta$. Тогда случайная величина

$$\frac{\xi/k}{\eta/m} \sim F_{k,m}$$

имеет *распределение Фишера* с параметрами k, m.

Построим доверительные интервалы для параметров в гауссовой линейной модели.

Доверительный интервал для σ^2*:*

Поскольку $\frac{1}{\sigma^2}||X - Z\widehat{\theta}||^2 \sim \chi^2(n-k)$, то достаточно взять квантиль $u_{1-\gamma}$ распределения $\chi^2(n-k)$, а значит

$$\mathsf{P}\left(\frac{1}{\sigma^2}||X - Z\widehat{\theta}||^2 > u_{1-\gamma}\right) = \gamma \Rightarrow \mathsf{P}\left(\sigma^2 \in \left(0; \frac{||X - Z\widehat{\theta}||^2}{u_{1-\gamma}}\right)\right) = \gamma.$$

Доверительный интервал для $\widehat{\theta}_i$*:*

Поскольку $\widehat{\theta} \sim \mathcal{N}(\theta,\, \sigma^2 \underbrace{(Z^T Z)^{-1}}_{A})$, где $A = (a)_{ij}$, то $\widehat{\theta}_i \sim \mathcal{N}(\theta_i,\, \sigma^2 a_{ii})$. Тогда

$$\begin{cases} \frac{\widehat{\theta}_i - \theta_i}{\sqrt{\sigma^2 a_{ii}}} \sim \mathcal{N}(0,1) \\ \frac{1}{\sigma^2}||X - Z\widehat{\theta}||^2 \sim \chi^2(n-k) \end{cases} \Rightarrow \sqrt{\frac{n-k}{a_{ii}}} \frac{\widehat{\theta}_i - \theta_i}{\sqrt{||X - Z\widehat{\theta}||^2}} \sim T_{n-k},$$

откуда

$$\mathsf{P}\left(u_{(1-\gamma)/2} \leqslant \sqrt{\frac{n-k}{a_{ii}}} \frac{\widehat{\theta}_i - \theta_i}{\sqrt{||X - Z\widehat{\theta}||^2}} \leqslant u_{(1+\gamma)/2}\right) = \gamma,$$

где u_p — квантили T_{n-k}.

Доверительная область для θ*:*

$$\begin{cases} \frac{1}{\sigma^2}||Z\widehat{\theta} - Z\theta||^2 \sim \chi^2(k) \\ \frac{1}{\sigma^2}||X - Z\widehat{\theta}||^2 \sim \chi^2(n-k) \end{cases} \Rightarrow \frac{n-k}{k} \frac{||Z\widehat{\theta} - Z\theta||^2}{||X - Z\widehat{\theta}||^2} \sim F_{k,n-k}$$

11 Проверка гипотез

11.1 Построение критериев

Обозначим неизвестное распределение P. Тогда гипотезой назовем любое утверждение относительно P и обозначим $H : \mathsf{P} \in \mathcal{P}$. Пусть $\mathcal{P}_0$ и $\mathcal{P}_1$ — два непересекающихся класса распределений. Мы будем проверять гипотезы вида "наблюдаемая величина имеет распределение из класса $\mathcal{P}_0$ "и обозначать их $H_0 : \mathsf{P} \in \mathcal{P}_0$. Тогда H_0 называется *основной гипотезой*. Противоречащую ей гипотезу $H_1 : \mathsf{P} \in \mathcal{P}_1$ назовем *альтернативной* гипотезой.

Определение 11.1. Гипотеза H_0 называется простой, если $|\mathcal{P}_0| = 1$.

Определение 11.2. Множество S называется *критерием* проверки гипотезы H_0 против альтернативы H_1, если $S \subseteq \mathcal{X}$.

Гипотеза H_0 отвергается в пользу альтернативы H_1 если $X \in S$.

Пример 11.1. Пусть $\mathcal{P} = \{\mathsf{P}_\theta \mid \theta \in \Theta\}$ и $H_0 : \mathsf{P} = \mathsf{P}_{\theta_0}$, $H_1 : \mathsf{P} \neq \mathsf{P}_{\theta_0}$. Построим для θ доверительный интервал $(T_1(X), T_2(X))$ уровня γ. Тогда

$$\mathsf{P}_\theta \left(\underbrace{\theta \in (T_1(X),\, T_2(X))}_{A} \right) \geqslant \gamma.$$

Если событие A не выполнено для θ_0, то гипотеза H_0 отвергается. Заметим, что с вероятностью $\leqslant 1-\gamma$ верная гипотеза будет отвергнута.

Определение 11.3. *Ошибкой первого рода* называется ситуация, когда отвергается верная гипотеза.*Ошибкой второго рода* называется ситуация, когда неверная гипотеза не отвергается.

Определение 11.4. *Мощностью* критерия S называется функция $\beta(Q, S) = Q(X \in S)$, где $Q \in \mathcal{P} = \mathcal{P}_0 \sqcup \mathcal{P}_1$.

Определение 11.5. S — критерий *уровня значимости* $1 > \varepsilon > 0$, если

$$\forall Q \in \mathcal{P}_0 : \;\; \beta(Q, S) \leqslant \varepsilon.$$

Размер критерия S — наименьший из его уровней значимости.

$$\alpha(S) := \sup_{Q \in \mathcal{P}_0} \beta(Q, S).$$

Определение 11.6. Пусть S и R — два критерия уровня значимости ε. Тогда критерий S *мощнее* критерия R, если

$$\forall Q \in \mathcal{P}_1 : \;\; \beta(Q, S) \geqslant \beta(Q, R).$$

Заметим, что рассматриваемая вероятность $\beta(Q, S) = Q(X \in S)$ это вероятность отклонить неверную гипотезу.

Определение 11.7. Критерий S называется *равномерно наиболее мощным критерием* (далее рнмк) уровня значимости ε, если выполнены следующие два свойства:

1. $\alpha(S) \leqslant \varepsilon$.

2. S мощнее любого другого критерия R уровня значимости $\alpha(R) \leqslant \varepsilon$.

Определение 11.8. Критерий S называется *несмещенным*, если

$$\sup_{Q\in\mathcal{P}_0} \beta(Q,S) < \inf_{Q\in\mathcal{P}_1} \beta(Q,S).$$

Критерий S называется *состоятельным*, если

$$\forall Q \in \mathcal{P}_1 \lim_{n\to+\infty} \beta(Q,S) = \lim_{n\to+\infty} Q\left(X \in S\right) \to 1,$$

где n — размер выборки $X = X_1, \ldots, X_n$.

Пусть P_0, P_1 — два распределения, доминируемые относительно меры μ с плотностями p_0 и p_1 соответственно. Рассмотрим гипотезы $H_0 :\ \mathsf{P} = \mathsf{P}_0$ и $H_1 :\ \mathsf{P} = \mathsf{P}_1$. Введем для $\lambda \geqslant 0$ множество

$$S_\lambda := \{x :\ p_1(x) - \lambda p_0(x) \geqslant 0\}.$$

Теорема 11.1. *(лемма Неймана-Пирсона)*

Пусть R — критерий, такой что

$$\mathsf{P}_0(X \in R) \leqslant \mathsf{P}_0(X \in S_\lambda)$$

(т.е. уровня значимости $\mathsf{P}_0(X \in S_\lambda)$. Тогда критерий S_λ мощнее критерия R

$$\mathsf{P}_1(X \in R) \leqslant \mathsf{P}_1(X \in S_\lambda)$$

и, кроме того, S_λ — несмещенный критерий.

Доказательство. По свойствам индикаторов

$$I(X\in R)(p_1(x) - \lambda p_0(x)) \leqslant I(X \in R)I(X \in S_\lambda)(p_1(x) - \lambda p_0(x)) \leqslant I(X \in S_\lambda)(p_1(x) - \lambda p_0(x)).$$

Тогда

$$\begin{aligned}\mathsf{P}_1(x \in R) - \lambda \mathsf{P}_0(x \in R) &= \int \left(p_1(x) - \lambda p_0(x)\right) I(x \in R)\mu(dx) \\ &\leqslant \int I(X \in S_\lambda)(p_1(x) - \lambda p_0(x))\mu(dx) = \mathsf{P}_1(x \in S_\lambda) - \lambda \mathsf{P}_0(x \in S_\lambda),\end{aligned}$$

откуда

$$\mathsf{P}_1(x \in R) - \mathsf{P}_1(x \in S_\lambda) \leqslant \lambda\left(\mathsf{P}_0(x \in R) - \mathsf{P}_0(x \in S_\lambda)\right) \leqslant 0 \Rightarrow \mathsf{P}_1(x \in R) \leqslant \mathsf{P}_1(x \in S_\lambda).$$

Рассмотрим теперь два случая:

1. $\lambda \geqslant 1 \Rightarrow \forall x \in S_\lambda:\ p_1(x) \geqslant p_0(x)$ и

$$\mathsf{P}_0(x \in S_\lambda) = \int I(x \in S_\lambda) p_0(x)\mu(dx) \leqslant \int I(x \in S_\lambda) p_1(x)\mu(dx) = \mathsf{P}_1(x \in S_\lambda).$$

2. $\lambda < 1$. Тогда для $x \notin S_\lambda : p_1(x) < p_0(x)$ и

$$\mathsf{P}_0(x \notin S_\lambda) = \int I_(x \notin S_\lambda) p_0(x)\mu(dx) \geqslant$$
$$\geqslant \int I_(x \notin S_\lambda) p_1(x)\mu(dx) = \mathsf{P}_1(x \notin S_\lambda),$$

откуда

$$\mathsf{P}_0(x \in S_\lambda) = 1 - \mathsf{P}_0(x \notin S_\lambda) \leqslant 1 - \mathsf{P}_1(x \notin S_\lambda) = \mathsf{P}_1(x \in S_\lambda).$$

□

Следствие 11.1.1. *Пусть λ таково, что*

$$\mathsf{P}_0(x \in S_\lambda) = \varepsilon.$$

Тогда S_λ это рнмк уровня значимости ε.

Замечание. Для дискретного пространства не существует рнмк уровня значимости ε для всех ε.

Пример 11.2. Пусть $X_1,\ \ldots,\ X_n \sim Bern(\theta)$. Нужно проверить гипотезу $H_0 : \theta = \frac{1}{4}$ против $H_1 : \theta = \frac{1}{3}$. Тогда

$$S_\lambda = \{x :\ \left(\frac{1}{3}\right)^{\sum X_i} \left(\frac{2}{3}\right)^{n-\sum X_i} - \lambda \left(\frac{1}{4}\right)^{\sum X_i} \left(\frac{3}{4}\right)^{n-\sum X_i} \geqslant 0\} = \{\left(\frac{4}{3}\right)^{\sum X_i} \left(\frac{8}{9}\right)^{n-\sum X_i} \geqslant \lambda\}.$$

Поскольку функция $\left(\frac{4}{3}\right)^{\sum X_i} \left(\frac{8}{9}\right)^{n-\sum X_i}$ возрастает по $T(X) = \sum X_i$, то неравенство верно при $T(X) \geqslant \tilde{\lambda}$. Если H_0 верна, то $\sum X_i \sim Bin(n,\ \frac{1}{4})$ и $\tilde{\lambda}$ — квантиль уровня $1-\varepsilon$ распределения $Bin(n,\ \frac{1}{4})$. Причем все проведенные рассуждения верны только для тех ε, где достигается равенство $\mathsf{P}_0(x \in S_\lambda) = \varepsilon$.

Теорема 11.2. *(б/д, о монотонном отношении правдоподобия)*

Пусть $\Theta \subset \mathbb{R}$, $H_0 : \theta \leqslant \theta_0 (\theta = \theta_0)$, $H_1 : \theta > \theta_0$. Пусть P_θ доминируемо относительно меры μ с плотностью p_θ и

$$\forall \theta_2 > \theta_1 \in \Theta :\ \frac{f_{\theta_2}(X)}{f_{\theta_1}(X)} = g(T(X), \theta_1,\ \theta_2),$$

где g не убывает по $T(X)$.

Тогда рнмк уровня значимости ε имеет вид

$$S_\varepsilon = \{T(X) \geqslant c_\varepsilon\},$$

если

$$\mathsf{P}_0(S_\varepsilon) = \varepsilon.$$

Пример 11.3. $X_1, \ldots, X_n \sim Bern(\theta)$, $H_0 : \theta \geqslant \frac{1}{4}$, $H_1 : \theta < \frac{1}{4}$. Сделаем замену $\tilde{\theta} = -\theta$. Тогда $H_0 : \tilde{\theta} \leqslant -\frac{1}{4}$, $H_1 : \tilde{\theta} > -\frac{1}{4}$. Для $\tilde{\theta}_2 > \tilde{\theta}_1$:

$$\frac{f_{\tilde{\theta}_2}}{f_{\tilde{\theta}_1}} = \left(\underbrace{\frac{-\tilde{\theta}_2}{-\tilde{\theta}_1}}_{<1}\right)^{\sum X_i} \left(\underbrace{\frac{1-(-\tilde{\theta}_2)}{1-(-\tilde{\theta}_1})}_{>1}\right)^{n-\sum X_i} = g(-\sum X_i).$$

Функция g возрастает по $-\sum X_i$, откуда рнмк имеет вид $S = \{\sum X_i \leqslant c\}$, где c — квантиль уровня ε для $Bin(n, \frac{1}{4})$.

Пусть $X_1, \ldots, X_n$ — выборка из $U[0, \theta]$. Проверим гипотезу $H_0 : \theta = \theta_0$ против альтернативы $H_1 : \theta < \theta_0$.

Утверждение 11.1.1. Рнмк имеет вид $S = \{X_{(n)} \leqslant \theta_0 \varepsilon^{\frac{1}{n}}\}$.

Доказательство.

$$\mathsf{P}_{\theta_0}(X_{(n)} \leqslant c) = \varepsilon \Rightarrow \left(\frac{c}{\theta_0}\right)^n = \varepsilon \Rightarrow c = \theta_0 \varepsilon^{\frac{1}{n}}.$$

Пусть теперь R — критерий уровня значимости ε, т.е. $\mathsf{P}_{\theta_0}(x \in R) \leqslant \varepsilon$. Возможны два случая:

1. $\theta \leqslant c = \theta_0 \varepsilon^{\frac{1}{n}}$. Тогда

$$\mathsf{P}_{\theta}(x \in S) = 1 \geqslant \mathsf{P}_{\theta}(x \in R) \Rightarrow \ S \text{ — мощнее.}$$

2. $\theta \in (c, \theta_0)$. Тогда $\mathsf{P}_{\theta}(x \in S) = \left(\frac{\theta_0}{\theta}\right)^n \varepsilon$ и

$$\mathsf{P}_{\theta}(x \in R) = \int\limits_{[0,\theta]^n} \frac{1}{\theta^n} I(X_1, \ldots, X_n \in R) dx_1 \ldots dx_n = \left(\frac{\theta_0}{\theta}\right)^n \int\limits_{[0,\theta]^n} \frac{1}{\theta_0^n} I(X_1, \ldots, X_n \in R) dx_1 \ldots dx_n$$

$$\leqslant \left(\frac{\theta_0}{\theta}\right)^n \int\limits_{[0,\theta_0]^n} \frac{1}{\theta_0^n} I(X_1, \ldots, X_n \in R) dx_1 \ldots dx_n = \left(\frac{\theta_0}{\theta}\right)^n \mathsf{P}_{\theta_0}(X \in R) \leqslant \varepsilon \left(\frac{\theta_0}{\theta}\right)^n$$

□

11.2 Гипотезы в линейной регрессии

Рассмотрим линейную модель $X = Z\theta + \varepsilon$, где $\varepsilon \sim \mathcal{N}(0, \sigma I_n)$, $\theta \in \mathbb{R}^k$. Будем проверять гипотезы вида $H : T\theta = \tau$, где T это матрица размера $(m \times k)$, $m \leqslant k$, $\operatorname{rk} T = m$. Как мы помним,

$$\widehat{\theta} = (Z^T Z)^{-1} Z^T X \sim \mathcal{N}(\theta, \sigma^2 (Z^T Z)^{-1}),$$

а, зная это, имеем

$$T\widehat{\theta} \sim \mathcal{N}(\underbrace{T\theta}_{=\tau}, \sigma^2 \underbrace{T(Z^T Z)^{-1} T^T}_{=B}).$$

Матрица B обратиа как матрица с полным рангом, а значит

$$\sqrt{B^{-1}} \frac{1}{\sigma} (T\widehat{\theta} - \tau) \sim \mathcal{N}(0, I_m) \Rightarrow \frac{1}{\sigma^2} (T\widehat{\theta} - \tau)^T B^{-1} (T\widehat{\theta} - \tau) \sim \chi^2(m)$$

а

$$\frac{1}{\sigma^2}||X - Z\widehat{\theta}||^2 \sim \chi^2(n-k).$$

Зная всё это, получаем, что

$$\widehat{F} = \frac{(T\widehat{\theta} - \tau)^T B^{-1}(T\widehat{\theta} - \tau)}{||X - Z\widehat{\theta}||^2}\frac{n-k}{m} \sim F_{m,n-k}$$

Определение 11.9. Статистика $\widehat{F}$ называется ф-статистикой (эф).

Критерий для проверки H уровня значимости ε имеет вид $\widehat{F} > u_{1-\varepsilon}$, где u — квантиль уровня $1-\varepsilon$ распределения $F_{m,n-k}$.

Пример 11.4. $X_1, \ldots, X_n \sim \mathcal{N}(x, \sigma^2)$, $Y_1, \ldots, Y_m \sim \mathcal{N}(y, \sigma^2)$. Проверим гипотезу $H : x = y$.

$$\begin{pmatrix} X_1 \\ \vdots \\ X_n \\ Y_1 \vdots \\ Y_m \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ & \vdots & \\ 1 & 0 \\ 0 & 1 \\ & \vdots & \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \vec{\varepsilon}.$$

Возьмем $T = \begin{pmatrix} 1 & -1 \end{pmatrix} \Rightarrow \tau = T\theta = 0$. Тогда

$$Z^T Z = \begin{pmatrix} n & 0 \\ 0 & m \end{pmatrix}, \quad \widehat{\theta} = \begin{pmatrix} \overline{X} \\ \overline{Y} \end{pmatrix} \Rightarrow B = \frac{1}{m} + \frac{1}{n}$$

и

$$\widehat{F} = \frac{(\overline{X} - \overline{Y})^2 \frac{nm}{n^2+m^2}}{nS_X^2 + mS_Y^2} \cdot \frac{n+m-2}{1}.$$

11.3 Критерии согласия

Критерий хи-квадрат Пирсона

Пусть $X_1, \ldots, X_n$ — выборка из распределения, у которого $a_1, \ldots, a_m$ — все возможные исходы в одном испытании с вероятностями соответсвенно $p_1, \ldots, p_m$. Проверим гипотезу $H : p_i = p_i^0 \quad \forall i \in \{1, \ldots, m\}$. Положим $\mu_i = \sum\limits_{j=1}^{n} I(X_j = a_i)$ и рассмотрим статистику

$$\widehat{\mu} = \sum_{i=1}^{m} \frac{(\mu_i - np_i^0)^2}{np_i^0}$$

Теорема 11.3. *(Пирсон)*

При условии верности H_0, выполнена сходимость

$$\widehat{\mu} \xrightarrow{d} \chi^2(m-1) \text{ при } n \to \infty.$$

Доказательство. Рассмотрим вектора $Y_j = (I(X_j = a_1),\, \ldots,\, I(X_j = a_m))$, $(Y_j)_i \sim Bern(p_i^0)$. Тогда $\mathsf{E}Y_j = (p_1^0,\, \ldots,\, p_m^0) = p^T$ и

$$\mathsf{D}Y_j = B - pp^T, \text{ где } B = \operatorname{diag}(p_1^0,\, \ldots,\, p_m^0).$$

По ЦПТ

$$\sqrt{B^{-1}}\sqrt{n}\left(\left(\frac{Y_1 + \ldots + Y_n}{n}\right)^T - p\right) \xrightarrow{d} \mathcal{N}(0,\, \sqrt{B^{-1}}(B - pp^T)\sqrt{B^{-1}}).$$

Пусть $Z = \sqrt{B^{-1}}p = \begin{pmatrix} \sqrt{p_1^0} \\ \vdots \\ \sqrt{p_m^0} \end{pmatrix}$ и V — ортогональная матрица, первая строка которой равна Z^T. Тогда

$$V\sqrt{B^{-1}}\sqrt{n}\left(\left(\frac{Y_1 + \ldots + Y_n}{n}\right)^T - p\right) \xrightarrow{d} \mathcal{N}(0; \underbrace{VI_mV^T}_{=I_m} - VZZ^TV^T) = \mathcal{N}(0; \operatorname{diag}(0, \underbrace{1,\, \ldots,\, 1}_{m-1})).$$

По теореме о наследовании сходимости

$$\left\|V\sqrt{B^{-1}}\sqrt{n}\left(\left(\frac{Y_1 + \ldots + Y_n}{n}\right)^T - p\right)\right\|^2 = \sum_{j=1}^{m} \frac{(\mu_j - np_j^0)^2}{np_j^0} \xrightarrow{d} \chi^2(m-1).$$

$\square$

Следствие 11.3.1. *Гипотеза H отвергается на уровне значимости ε, если $\widehat{\mu} > u_{1-\varepsilon}$, где u — квантиль уровня $1 - \varepsilon$ распределения $\chi^2(m-1)$.*

Утверждение 11.3.1. Критерий Пирсона — состоятельный критерий.

Доказательство. Пусть $\exists i : \ p_i \neq p_i^0$. Без ограничения общности $i = 1$. Покажем, что в таком случае $\mathsf{P}(\widehat{\mu} > u_{1-\varepsilon}) \to 1$. По УЗБЧ $\frac{\mu_i}{n} \xrightarrow{\text{п.н.}} p_i$, а значит

$$\widehat{\mu} = \sum_{i=1}^{m} \frac{(\mu_i - np_i^0)^2}{np_i^0} = \sum_{i=1}^{m} \frac{n\left(\frac{\mu_i}{n} - p_i^0\right)^2}{p_i^0}.$$

В частности, при $i = 1$:

$$\frac{n\left(\frac{\mu_1}{n} - p_1^0\right)^2}{p_1^0} \xrightarrow{\text{п.н.}} n\frac{(p_1 - p_1^0)^2}{p_1^0} \xrightarrow{\text{п.н.}} +\infty.$$

$\square$

Пример 11.5. Максим Евгеньевич Жуковский едет на лекцию по математической статистике. Он планирует задать слушателям три вопроса в начале лекции, возможные варианты ответа на которые следующие: a_1 ="да, да", a_2 ="да, нет", a_3 ="нет, нет". В электричке Максим Евгеньевич выдвинул гипотезу $H : p_1^0 = \frac{1}{2},\, p_2^0 = \frac{1}{3},\, p_3^0 = \frac{1}{6}$.

Проведя опрос, Максим Евгеньвич получил следующие результаты: $\mu_1 = 28$, $\mu_2 = 20$, $\mu_3 = 12$. В таком случае,

$$\widehat{\mu} = \frac{(28 - 30)^2}{30} + \frac{(12 - 10)^2}{30} = \frac{8}{15}.$$

Посмотрев на википедии квантили $\chi^2(2)$, Максим Евгеньевич, пользуясь критерием Пирсона, отвергает H на уровне значимости 0.8, но не отвергает на уровне значимости 0.1.

Определение 11.10. Пусть $\{S(x) > u\}$ — критерий проверки гипотезы $H : \mathsf{P} = \mathsf{P}_0$ и $\alpha = \mathsf{P}_0(S(x) > u)$ — его уровень значимости. Найдем значение $S(x)$ для выборки $X_1, \ldots, X_n$: $S(X_1, \ldots, X_n) = t$. Величина $p = \mathsf{P}_0(S(x) > t)$ называется *p-значением (p-value)*. При $t > u \Rightarrow p < \alpha$ гипотеза H отвергается.

Критерий Колмогорова-Смирнова

Теорема 11.4. *(Колмогоров, Смирнов)*

Пусть имеется выборка из распределения с непрерывной функцией распределения. Тогда

$$\sqrt{n}\sup_x |F(x) - F_n(x)| \xrightarrow{d} K,$$

где K — распределение Колмогорова с функцией распределения

$$\begin{cases} \sum\limits_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2x^2} & x \geqslant 0, \\ 0 & x < 0 \end{cases}$$

Рассмотрим статистику $H : \mathsf{P} = \mathsf{P}_0$. Пусть $S(X) = \sqrt{n}\sup\limits_x |F(x) - F_n(x)|$ и $u_{1-\alpha}$ — квантиль распределения K уровня $1 - \alpha$. Тогда $\{S(x) > u_{1-\alpha}\}$ это критерий проверки H уровня значимости α.

Утверждение 11.3.2.

$$S(X) = \sqrt{n} \sup_{0 \leqslant k \leqslant n} \left\{ \left| F(X_{(k)} - \frac{k}{n} \right|, \left| F(X_{(k+1)} - \frac{k}{n} \right| \right\},$$

где $X_{(0)} := -\infty$, $X_{(n+1)} := +\infty$.

Доказательство. Следует из того, что $F_n(x) = \text{const}$ на $[X_{(k)}, X_{(k+1)})$. □

Критерий Мизеса-Смирнова

Теорема 11.5. *(б/д)*

$$\underbrace{n \int\limits_{\mathbb{R}} (F(x) - F_n(x)) dF(x)}_{\omega^2} \xrightarrow{d} \xi,$$

где $\xi \sim a1$.

Упражнение. $\omega^2 = \frac{1}{12n} + \sum\limits_{k=1}^{n} \left(X_{(k)} - \frac{k - \frac{1}{2}}{n} \right)^2$

Определение 11.11. Все три критерия (Пирсона, Колмогорова-Смиронова, Мизеса-Смирнова) называются *критериями согласия*, поскольку проверяют гипотезу вида $H : \mathsf{P} = \mathsf{P}_0$.

11.4 Байесовские критерии

Пусть мы хотим проверить гипотезу $H_0 : \mathsf{P} = \mathsf{P}_0$ против альтернативы $H_1 : \mathsf{P} = \mathsf{P}_1$, где P_0, P_1 — доминируемые относительно меры μ. Пусть Q — априорное распределение, и $Q(\mathsf{P} = \mathsf{P}_0) = p_0$, $Q(\mathsf{P} = \mathsf{P}_1) = p_1$. Для получения критерия разобьем множество $\mathcal{X} = S_0 \sqcup S_1$ на 2, такие что $X \in S_i \Rightarrow$ отклоняем H_i.

Вероятность ошибки первого рода в такой модели равна

$$p_0\mathsf{P}_0(X \in S_0) + p_1\mathsf{P}_1(X \in S_1) \to \min_{S_0,S_1}$$

и задача стоит в том, чтобы найти такое разбиение $\mathcal{X}$, при котором она минимальна.

Пусть $S = \begin{cases} S_0 & \mathsf{P} = \mathsf{P}_0, \\ S_1 & \mathsf{P} = \mathsf{P}_1 \end{cases}$ — случайное множество. Имеем

$$\mathsf{P}_0(X \in S_0) = \mathsf{E}I(X \in S) = \mathsf{E}\left(\mathsf{E}(I(X \in S) \mid X\right).$$

Найдем условное мат.ожидание

$$\begin{aligned}\mathsf{E}\left(I(x \in S) \mid X = x\right) = I(x \in S_0)\underbrace{\frac{p_0 f_0(x)}{p_0 f_0(x) + p_1 f_1(x)}}_{q_0} + I(x \in S_1)\underbrace{\frac{p_1 f_1(x)}{p_0 f_0(x) + p_1 f_1(x)}}_{q_1} = \\ = 1 - I(x \in S_1)q_0 - I(x \in S_0)q_1.\end{aligned}$$

Тогда

$$\max_{S_0,S_1} \mathsf{E}\left(I(x \in S_1)q_0 + I(x \in S_0)q_1\right) \leqslant \mathsf{E}\max\{q_0,\ q_1\}$$

и равенство достигается при $S_1 := \{p_0 f_0 > p_1 f_1\}$, $S_0 := \{p_1 f_1 \geqslant p_0 f_0\}$.

Пример 11.6. $X_1,\ \ldots,\ X_n \sim \mathcal{N}(a, 1)$ и $H_0 : a = a_0$, $H_1 : a = a_1$ и $Q(a = a_i) = p_i$. Тогда S_0 имеет вид

$$\begin{aligned}S_0 &= \left\{p_1\left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2}\sum(X_i - a_0)^2\right] > p_0\left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2}\sum(X_i - a_1)^2\right]\right\} \\ &= \left\{\exp\left[(a_1 - a_0)\sum X_i - \frac{a_1^2 - a_0^2}{2}n\right] > \frac{p_0}{p_1}\right\} \\ &= \left\{(a_1 - a_0)\overline{X} > \frac{a_1^2 - a_0^2}{2} + \frac{1}{n}\ln\frac{p_0}{p_1}\right\}.\end{aligned}$$