

МОСКОВСКИЙ ФИЗИКО-ТЕХНИЧЕСКИЙ ИНСТИТУТ

ФАКУЛЬТЕТ ИННОВАЦИЙ И ВЫСОКИХ ТЕХНОЛОГИЙ

Математическая статистика

Лектор: М.Е. Жуковский

КОНСПЕКТ ЛЕКЦИЙ

автор: АЛЕКСАНДР МАРКОВ

25 мая 2017 г.

Содержание

1	Сходимость случайных векторов	3
2	Вероятностно-статистическая модель	6
3	Статистики. Непараметрические статистики	8
3.1	Определение статистики. Примеры	8
3.2	Непараметрические статистики	8
3.3	Ядерные оценки плотности	10
4	Параметрические распределения. Оценки параметров	11
4.1	Определение и свойства оценок	11
4.2	Методы нахождения оценок	12
5	Способы сравнения статистик	16
5.1	Сравнения произвольных оценок	16
5.2	Поиск наилучшей оценки в классе несмещенных оценок	16
6	Оценка максимального правдоподобия	19
7	Условное математическое ожидание	22
7.1	Определение и свойства	22
7.2	Поиск УМО в абсолютно непрерывном случае	25
7.3	Поиск наилучшей оценки в классе несмещенных оценок	26
8	Доверительные интервалы	30
8.1	Построение доверительных интервалов методом центральной статистики	30
8.2	Асимптотические доверительные интервалы	31
9	Байесовские методы	32
9.1	Введение	32
9.2	Математическое описание байесовских методов. Сравнение подходов	32
10	Линейная регрессия	36
10.1	Линейная модель	36
10.2	Гауссовская линейная модель	38
11	Проверка гипотез	40
11.1	Построение критериев	40
11.2	Гипотезы в линейной регрессии	43

11.3	Критерии согласия	44
11.4	Байесовские критерии	47

1 Сходимость случайных векторов

Определение 1.1. Пусть ξ, ξ_1, \dots, ξ_n — k -мерные случайные вектора. Как и в случае случайных величин, существуют следующие виды сходимости:

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

Утверждение 1.0.1. Пусть ξ, ξ_1, \dots — случайные k -мерные вектора. Тогда верны следующие взаимосвязи между сходимостью векторов и их компонент:

$$\left. \begin{array}{l} \xi_n \xrightarrow{\text{п.н.}} \xi \\ \xi_n \xrightarrow{P} \xi \\ \xi_n \xrightarrow{L_p} \xi \end{array} \right\} \iff \forall i \in \{1, \dots, k\} \left\{ \begin{array}{l} \xi_n^{(i)} \xrightarrow{\text{п.н.}} \xi^{(i)} \\ \xi_n^{(i)} \xrightarrow{P} \xi^{(i)} \\ \xi_n^{(i)} \xrightarrow{L_p} \xi^{(i)} \end{array} \right.$$

Доказательство. 1. *сходимость почти наверное.* \Rightarrow : $\{\xi_n^{(i)} \rightarrow \xi^{(i)}\} \supset \{\xi_n \rightarrow \xi\}$ и вероятность события справа равна 1.

\Leftarrow : $\{\xi_n \rightarrow \xi\} = \bigcap_{j=1}^k \{\xi_n^{(j)} \rightarrow \xi^{(j)}\}$ (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 : $\bigcup_{i=1}^k \{|\xi_n^{(i)} - \xi^{(i)}| > \frac{\varepsilon}{k}\} \supset \{\|\xi_n - \xi\| > \varepsilon\}$

3. *сходимость в L_p .* Очевидна цепочка неравенств

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

Тогда \Leftarrow следует из линейности мат.ожидания, а \Rightarrow из свойства мат.ожидания $f \leq g \Rightarrow Ef \leq Eg$. □

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

Теорема 1.1. (теорема о наследовании сходимостей)

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

2. Пусть $\xi_n \xrightarrow{d} \xi$, $h : \mathbb{R}^k \rightarrow \mathbb{R}^m$ и непрерывна (Замечание: это не то же 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. $\xi_n \xrightarrow{\text{п.н.}} \xi$:

$$P(h(\xi_n) \rightarrow h(\xi)) \geq P(h(\xi_n) \rightarrow h(\xi), \xi \in B) \geq P(\xi_n \rightarrow \xi, \xi \in B) = 1$$

где $B = \{h \text{ непрерывна}\}$, $P(\xi \in B) = 1$.

2. $\xi_n \xrightarrow{P} \xi$: Предположим, что $h(\xi_n) \not\xrightarrow{P} h(\xi)$. Это означает, что

$$\exists \varepsilon > 0 \exists \delta > 0 : P(\|h(\xi_n) - h(\xi)\| > \varepsilon) > \delta \quad (1)$$

для бесконечно многих n . Пусть $\{n_j\}$ это те номера, при которых верно неравенство выше. Из условия $\xi_{n_j} \xrightarrow{P} \xi$. 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 $f : \mathbb{R}^m \rightarrow \mathbb{R}$. Тогда $f(h) = f \circ h : \mathbb{R}^k \rightarrow \mathbb{R}$ — непрерывная и ограниченная функция, а значит

$$Ef(h(\xi_n)) = E(f \circ h)(\xi_n) \rightarrow E(f \circ h)(\xi) = Ef(h(\xi))$$

и $h(\xi_n) \xrightarrow{d} h(\xi)$.

□

Теорема 1.2. (лемма Slutsky)

1. Если $\xi_n \xrightarrow{d} \xi$, а $\eta_n \xrightarrow{d} \eta = c = \text{const}$ — случайные величины. Тогда $\xi_n + \eta_n \xrightarrow{d} \xi + c$, $\xi_n \eta_n \xrightarrow{d} c\xi$
2. Если $\xi_n \xrightarrow{d} \xi = \text{const}$ — случайные вектора, то $\xi_n \xrightarrow{P} \xi$.

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

$$\xi_n^{(i)} \xrightarrow{d} C^{(i)} \Rightarrow \xi_n^{(i)} \xrightarrow{P} C^{(i)} \Rightarrow \xi_n \xrightarrow{P} \xi$$

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 $m \geq 1$, $h : \mathbb{R}^m \rightarrow \mathbb{R}$ — функция, дифференцируемая в точке $a \in \mathbb{R}^m$. Пусть $b_n \rightarrow 0$, $b_n \neq 0$. Тогда

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

Доказательство. $b_n \rightarrow 0 \Rightarrow b_n \xi_n \xrightarrow{P} 0$ по лемме Слущкого. 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 $\frac{\varphi(x)}{\|x\|} \rightarrow 0$, то по теореме о наследовании сходимости $\frac{\varphi(\xi_n b_n)}{\|b_n \xi_n\|} \xrightarrow{P} 0$.

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 Слущкого $\frac{\varphi(\xi_n b_n)}{b_n} = \frac{\varphi(\xi_n b_n)}{b_n \|\xi_n\|} \cdot \|\xi_n\| \xrightarrow{P} 0$, 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

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

□

2 Вероятностно-статистическая модель

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

Определение 2.1. Множество \mathcal{X} называется выборочным пространством.

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

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

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

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

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

Определение 2.3. Функция $X : \mathcal{X}^n \rightarrow \mathcal{X}^n; X(x) = x$ называется *наблюдением*. Аналогично одномерному случаю, $P_X = P^n$.

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 :

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

Теперь установим независимость:

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

□

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

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

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

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

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

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

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

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

Будем в дальнейшем для простоты обозначений писать $(\mathcal{X}, \mathcal{B}(\mathcal{X}), P)$ вместо $(\mathcal{X}^\infty, \mathcal{B}(\mathcal{X}^\infty), P^\infty)$ и называть выборку наблюдением и наоборот.

Определение 2.8. Тройка

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

где

- а) \mathcal{X} — выборочное пространство,
- б) $\mathcal{B}(\mathcal{X})$ — σ -алгебра на \mathcal{X} ,
- с) P — множество вероятностных мер на измеримом пространстве $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$

называется *вероятностно-статистической моделью*.

3 Статистики. Непараметрические статистики

3.1 Определение статистики. Примеры

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

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

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

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

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

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

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

□

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

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

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

3.2 Непараметрические статистики

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

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

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

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

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

$$\forall B \in \mathcal{B}(\mathcal{X}) : P_n^*(B) \rightarrow P_X(B) \text{ при } n \rightarrow \infty$$

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

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

Определение 3.3. Функция $F_n^*(x) = P_n^*((-\infty, x]) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x)$ называется *эмпирической функцией распределения*, построенной по выборке X_1, \dots, 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

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

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

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

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

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

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

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

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

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

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.

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

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, \dots, 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. $q(x) = \frac{1}{2} I(|x| \leq 1)$ — прямоугольное ядро.

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

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

Определение 3.5. Рассмотрим выборку X_1, \dots, X_n из неизвестного распределения P . Вероятностная мера \tilde{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 $\frac{B - X_i}{h_n} = \left\{ \frac{x - X_i}{h_n} \mid x \in B \right\}$ и $h_n \rightarrow 0$, $h_n > 0$ называется *сглаженным эмпирическим распределением*.

Сглаженное эмпирическое распределение обладает следующим набором свойств:

1. \tilde{P}_n имеет плотность $\tilde{p}_n(x) = \frac{1}{nh_n} \sum_{i=1}^n q\left(\frac{x - x_i}{h_n}\right)$
2. \tilde{P}_n — свертка распределений P_n^* и $Q(\frac{B}{h_n})$
3. Пусть $\alpha = \int_{\mathbb{R}} q^2(x) dx < +\infty$, $h_n \rightarrow 0$, $nh_n \rightarrow +\infty$ и $p(x)$ — непрерывна и ограничена. Тогда $\tilde{p}_n(x) = p_n(x) + \frac{\xi_n}{\sqrt{nh_n}}$, где $p_n(x) = E\tilde{p}_n(x) = \frac{1}{h_n} \int_{\mathbb{R}} q\left(\frac{x-y}{h_n}\right) p(y) dy$ и $\xi_n(x) \xrightarrow{d} \mathcal{N}(0, \alpha p(x))$

4 Параметрические распределения. Оценки параметров

4.1 Определение и свойства оценок

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

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

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

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

$$E_\theta \theta^* = \theta$$

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

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

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

$$\forall \theta \in \Theta : \theta_n^* \xrightarrow{P_\theta} \theta \quad (\theta_n^* \xrightarrow{P_\theta \text{ п.н.}} \theta)$$

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

Пример 4.2. Оценка \bar{X} является состоятельной оценкой по ЗБЧ для $E_\theta X_1$, и даже сильно состоятельной оценкой для $E_\theta X_1$ по УЗБЧ

Факт. *Сильно состоятельные оценки являются состоятельными.*

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

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

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

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

Пример 4.3. $\sqrt{n}(\bar{X} - E_\theta X_1) \xrightarrow{d_\theta} \mathcal{N}(0, D_\theta X_1)$ по ЦПТ

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

Доказательство.

$$\left. \begin{array}{l} \sqrt{n}(\theta^* - \theta) \xrightarrow{d_\theta} \xi \sim \mathcal{N}(0, \sigma^2(\theta)) \\ \frac{1}{\sqrt{n}} \rightarrow 0 \end{array} \right\} \Rightarrow \text{по лемме Служцкого } \theta^* - \theta \xrightarrow{d_\theta} 0 \Rightarrow \theta^* - \theta \xrightarrow{P_\theta} 0$$

□

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

Доказательство. Прямое следствие теоремы о наследовании сходимости.

□

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

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:

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

□

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

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

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

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

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

4.2 Методы нахождения оценок

1) Метод подстановки

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

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

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

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

Однако, зачастую такой метод неприменим в виду сложности функции G , поэтому 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

$$m(\theta) = (E_\theta g_1(X_1), \dots, E_\theta g_k(X_1))$$

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

Найдем $m^{-1} \left(\begin{pmatrix} \overline{g_1(X)} \\ \dots \\ \overline{g_k(X)} \end{pmatrix} \right) = \theta^*$ — 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 (α, λ) , $g_1(x) = x$, $g_2(x) = x^2$. В таком случае

$$\begin{aligned} E_\theta X_1 &= \int_0^{+\infty} x \frac{\alpha^\lambda e^{-\alpha x}}{\Gamma(\lambda)} dx = \frac{\Gamma(\lambda + 1)}{\Gamma(\lambda) \alpha} \int_0^{+\infty} \frac{x^\lambda \alpha^{\lambda+1} e^{-\alpha x}}{\Gamma(\lambda)} dx = \frac{\lambda}{\alpha} \\ E_\theta X_1^2 &= \int_0^{+\infty} x^2 \frac{\alpha^\lambda e^{-\alpha x}}{\Gamma(\lambda)} dx = \frac{\Gamma(\lambda + 2)}{\Gamma(\lambda)} \frac{1}{\alpha^2} = \frac{\lambda(\lambda + 1)}{\alpha^2} \end{aligned}$$

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

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

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 $\overline{g_i(X)} \xrightarrow{P_\theta \text{ п.н.}} E_\theta g_i(X)$, откуда $\begin{pmatrix} \overline{g_1(X)} \\ \dots \\ \overline{g_k(X)} \end{pmatrix} \xrightarrow{P_\theta \text{ п.н.}} m(\theta)$, а значит, по теореме о наследовании сходимости $m^{-1} \begin{pmatrix} \overline{g_1(X)} \\ \dots \\ \overline{g_k(X)} \end{pmatrix} \xrightarrow{P_\theta \text{ п.н.}} \theta$. \square

Утверждение 4.2.2. (б/д) Пусть $\Theta \subset \mathbb{R}$, m^{-1} дифференцируема на $m(\Theta)$ и существует $E_\theta [(g_1(X_1))^2]$. 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

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

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, \dots, X_n — выборка из $\mathcal{N}(\theta, 3)$. Найдем оценки для θ по методу моментов и методу квантилей: по методу моментов это \bar{X} , а по методу квантилей: μ^* . Для $\theta^* = \bar{X}$ 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 $g : \Theta^2 \rightarrow \mathbb{R}_+$ называется *функцией потерь*, а $E_{\theta}g(\theta^*, \theta)$ — *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 $\hat{\theta}$, если $\forall \theta \in \Theta : E_{\theta}g(\theta^*, \theta) \leq E_{\theta}g(\hat{\theta}, \theta)$ 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 $\theta_1 \neq \theta_0$. \square

2) Байесовский подход

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

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 $\hat{\theta}$, если $\sup_{\theta \in \Theta} g(\theta^*, \theta) < \sup_{\theta \in \Theta} g(\hat{\theta}, \theta)$

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 $P(B) = \sum_{k \in B \cap \mathbb{Z}_+} P(\{k\}) = \sum_{k \in B \cap \mathbb{Z}_+} p(k) =: \int_B p(x) \mu(dx)$, где $\mu(dx)$ — считающая мера, т.е. $\mu : \mathcal{B}(\mathbb{R}) \rightarrow \mathbb{Z}_+ \cup \{+\infty\}$ 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 $P(B) = \int_B p(x) \mu(dx)$.

Далее считаем, что имеющееся семейство распределений \mathcal{P} — доминируемо относительно некоторой меры μ и $X_1, \dots, X_n = X$ — выборка из исследуемого распределения $P_\theta \in \mathcal{P}$ 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 $I_X(\theta) = E_\theta [u_\theta(X)]^2$ — *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:

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

обосновать которое можно так:

$$\begin{aligned} \frac{\partial}{\partial \theta} \int_{\mathbb{R}} S(x) p_\theta(x) \mu(dx) &= \int_A S(x) \frac{\partial}{\partial \theta} p_\theta(x) \cdot \frac{1}{p_\theta(x)} p_\theta(x) dx = \\ &= \int_{\mathbb{R}} S(x) \left[\frac{\partial}{\partial \theta} \ln p_\theta(x) \right] p_\theta(x) dx = E_\theta \left[S(x) \frac{\partial}{\partial \theta} \ln p_\theta(x) \right] \end{aligned}$$

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 $\hat{\tau}$ — несмещенная оценка параметра $\tau(\theta)$. Тогда выполнено неравенство

$$D_\theta \hat{\tau} \geq \frac{[\tau'(\theta)]^2}{I_X(\theta)} \quad \forall \theta \in \Theta$$

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

$$0 = \frac{\partial}{\partial \theta} E_\theta 1 = E_\theta \frac{\partial}{\partial \theta} \ln p_\theta(x) \Rightarrow E_\theta u_\theta(X) = 0 \quad (1)$$

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

$$\tau'(\theta) = \mathbb{E}_\theta \hat{\theta} u_\theta(X) \quad (2)$$

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

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

возведем в квадрат и воспользуемся неравенством Коши-Буняковского:

$$(\tau'(\theta))^2 = \left(\mathbb{E}_\theta \left[\hat{\theta} - \tau(\theta) \right] u_\theta(X) \right)^2 \leq \mathbb{E}_\theta \left[\hat{\theta} - \tau(\theta) \right]^2 \mathbb{E}_\theta u_\theta(X)^2 = D_\theta \hat{\theta} I_X(\theta)$$

откуда следует требуемое неравенство. \square

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

Определение 5.9. Если $\forall \theta \in \Theta$ для несмещенной оценки $\hat{\theta}$ параметра $\tau(\theta)$ в неравенстве Рао-Крамера достигается равенство, то оценка $\hat{\theta}$ 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 $\forall \theta \in \Theta : 0 = \mathbb{E}_\theta a(\theta) = a(\theta) = 0$.

Имеем теперь

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

Посчитав матожидание 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 $c(\theta) = \frac{\tau'(\theta)}{I_X(\theta)}$ \square

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

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 $X = (X_1, \dots, X_n)$. Тогда

$$\begin{aligned} I_{(X_1, \dots, X_n)}(\theta) &= \mathbb{E}_\theta u_\theta(X_1, \dots, X_n)^2 = D_\theta u_\theta(X_1, \dots, X_n) = \\ &= D_\theta \sum_{i=1}^n u_\theta(X_i) = \sum_{i=1}^n D_\theta u_\theta(X_i) = n D_\theta u_\theta(X_1) = n I_{X_1}(\theta) = n i(\theta) \end{aligned}$$

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

6 Оценка максимального правдоподобия

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

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

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

Пример 6.1. Рассмотрим $\Theta = \mathbb{N}$ и $P_\theta = U\{1, \dots, \theta\}$. Тогда функция правдоподобия равна

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

Откуда $\theta^* = X_{(n)}$.

Пример 6.2. $\Theta = \mathbb{R} \times (0, +\infty)$ и X_1, \dots, 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

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

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

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

С этого момента считаем, что $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ — произвольное семейство распределений, доминируемое относительно меры μ , плотность P_θ равна p_θ и если $\theta_1 \neq \theta_2$ то $P_{\theta_1} \neq P_{\theta_2}$. 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, \dots, X_n — выборка из $P \in \mathcal{P}$.

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 $\int_A p_\theta(x) \mu(dx)$ трижды дифференцируемый по θ под знаком интеграла.

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

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 $E_\theta H(X_1) < +\infty$

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

$$P_{\theta_0}(f_{\theta_0}(X_1, \dots, X_n) > f_\theta(X_1, \dots, X_n)) \xrightarrow{n \rightarrow \infty} 1$$

Доказательство.

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

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

$$\begin{aligned} E_{\theta_0} \ln \frac{p_\theta(x)}{p_{\theta_0}(x)} &= \int_A \ln \frac{p_\theta(x)}{p_{\theta_0}(x)} p_{\theta_0}(x) \mu(dx) \\ &\leq \int_A \left(\frac{p_\theta(x)}{p_{\theta_0}(x)} - 1 \right) p_{\theta_0}(x) \mu(dx) \\ &= \int_A (p_\theta(x) - p_{\theta_0}(x)) \mu(dx) = 1 - 1 = 0 \end{aligned}$$

где мы воспользовались неравенством $\ln(1+x) \leq x$. 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 $P_\theta \neq P_{\theta_0}$, а значит $E_{\theta_0} \ln \frac{p_\theta(x)}{p_{\theta_0}(x)} < 0$ — 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 $f \leq g \Rightarrow f^+ \leq g^+$, $f^- \geq g^-$, причем $Eg = 0 \Rightarrow Ef = Ef^+ - Ef^- \leq Eg^+ - Eg^- = 0$, 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. \square

Теорема 6.2. Пусть выполнены **R1-R4** и $\forall n \forall x_1, \dots, x_n$ существует единственное решение θ^* уравнения $\frac{\partial}{\partial \theta} \ln f_\theta(x_1, \dots, x_n) = 0$. Тогда θ^* это состоятельная оценка параметра θ и $\forall \theta \in \Theta : P_\theta(\theta^* - O.M.P.) \rightarrow 1$ при $n \rightarrow \infty$.

Утверждение 6.0.1. Если $|\Theta| < \infty$ и при фиксированных X_1, \dots, X_n найдется $\arg \max_{\theta \in \Theta} f_\theta(X_1, \dots, X_n)$ 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

$$\begin{cases} P_\theta(|\theta^* - \theta| > \varepsilon) > \delta & \forall n \in \mathbb{N} \\ \forall \theta_0 \in \Theta : P_{\theta_0}(f_{\theta_0} > f_\theta) \rightarrow 1 \end{cases} \Rightarrow \text{противоречие.}$$

□

Теорема 6.3. (б/д)

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

Более того, если $\hat{\theta}$ — асимптотически нормальная оценка θ с асимптотической дисперсией $\sigma^2(\theta)$ и σ^2 непрерывна на Θ , то $\sigma^2(\theta) \geq \frac{1}{i(\theta)}$ — 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 определению, имеем

$$\begin{aligned} \frac{\partial}{\partial \theta} \ln f_\theta > 0 &\iff \theta^* > \theta \\ \frac{\partial}{\partial \theta} \ln f_\theta < 0 &\iff \theta^* < \theta \end{aligned}$$

что и означает, что θ^* это о.м.п.

□

7 Условное математическое ожидание

7.1 Определение и свойства

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

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

1. (измеримость) η является \mathcal{G} измеримой случайной величиной.
2. (интегральное условие) $\forall A \in \mathcal{G} \ E\xi I_A = E\eta I_A$

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

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

$$P(A) = 0 \Rightarrow \nu(A) = 0$$

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 $P(B) = \int_B g(x)dx$, 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 -н.н. случайная величина η на (Ω, \mathcal{F}, P) , такая что

$$\forall A \in \mathcal{F} : \nu(A) = E\eta I_A = \int_A \eta(\omega)P(d\omega)$$

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

Доказательство. Рассмотрим функцию $\nu(A) = E\xi I_A$ 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 (Ω, \mathcal{G}) , такая что $\nu(A) = E\eta I_A$.

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

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

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

Доказательство. Обозначим $\eta := E(\xi|\mathcal{G})$. 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 \emptyset поскольку в нем лежит ω_1 и не лежит ω_2 . Однако, так как $\mathcal{G} = \sigma(D_1, \dots)$ — объединение конечного и бесконечного числа множеств D_i , то A не может лежать в \mathcal{G} — противоречие, т.е. $E(\xi|\mathcal{G}) = \sum_{i=1}^{\infty} c_i I_{D_i}$.

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

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

откуда следует требуемое утверждение. \square

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 $E(\xi|\sigma(D_1, D_2)) = \frac{E\xi I_{D_1}}{P(D_1)} I_{D_1} + \frac{E\xi I_{D_2}}{P(D_2)} I_{D_2} = \frac{3}{2} I_{D_1} + \frac{4}{2} I_{D_2}$.

Докажем некоторые свойства условных математических ожиданий.

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

Доказательство. Воспользуемся интегральным свойством для $A = \Omega$: $E\xi = E\xi I_A = E(E(\xi|\mathcal{G})I_A) = E(E(\xi|\mathcal{G}))$ \square

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

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

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

$$E\eta I_A = P(A)E\xi = E\xi I_A$$

\square

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

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

$$E\zeta I_A = aE(E(\xi|\mathcal{G})I_A) + bE(E(\eta|\mathcal{G})I_A) = aE\xi I_A + bE\eta I_A = E(a\xi + b\eta)I_A$$

\square

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

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

$$E(E(\eta - \xi|\mathcal{G})) = E(\eta - \xi)I_A \geq 0$$

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

Утверждение 7.1.7. (Телескопическое свойство)

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

$$E(E(\xi|\mathcal{G}_1)|\mathcal{G}_2) = E(\xi|\mathcal{G}_1) \text{ п.н.} \quad (1)$$

$$E(E(\xi|\mathcal{G}_2)|\mathcal{G}_1) = E(\xi|\mathcal{G}_1) \text{ п.н.} \quad (2)$$

Доказательство. Поскольку $E(\xi|\mathcal{G}_1)$ является \mathcal{G}_2 -измеримой, то равенство (1) выполнено по утверждению 7.1.3.

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

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

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

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

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

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

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

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

откуда по линейности получаем требуемое 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 $E|\xi\eta| < +\infty$. По свойству 7.1.8 имеем

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

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

\square

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

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

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

7.2 Поиск УМО в абсолютно непрерывном случае

Обозначим

$$\mathbb{E}(\xi|\eta) = \mathbb{E}(\xi | \mathcal{F}_\eta)$$

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

$$\mathbb{P}(A|\eta) = \mathbb{E}(I_A|\mathcal{F}_\eta)$$

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

$$\mathbb{E}\xi I(\eta \in B) = \int_B \varphi(y) \mathbb{P}_\eta(dy) = \mathbb{E}\varphi(\eta) I(\eta \in B) = \int_{\omega: \eta(\omega) \in B} \varphi(\eta(\omega)) \mathbb{P}(d\omega).$$

Лемма 7.1. (6/д)

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

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

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

$$\mathbb{E}(\xi | \eta = y) = \sum x \mathbb{P}(\xi = x | \eta = y) = \int x p_{(\xi|\eta)}(x | y) dx.$$

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

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

$$\mathbb{P}_\xi(B|\eta = y) = \int_B p_{(\xi|\eta)}(x | y) dx.$$

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

Доказательство. Достаточно доказать, что

$$\forall B \in \mathcal{B}(\mathbb{R}) : \mathbb{E}g(\xi) I(\eta = B) = \int_B \left(\int_{\mathbb{R}} g(x) p_{(\xi|\eta)}(x | y) dx \right) \mathbb{P}_\eta(dy).$$

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

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

$$\begin{aligned} \int_B \left(\int_{\mathbb{R}} I(x \in A) p_{(\xi|\eta)}(x | y) dx \right) \mathbb{P}_\eta(dy) &= \int_B \left(\int_A p_{(\xi|\eta)}(x | y) dx \right) \mathbb{P}_\eta(dy) \\ &= \int_B \mathbb{P}_\xi(A | \eta = y) \mathbb{P}_\eta(dy) \\ &= \int_B \mathbb{E}(I(\xi \in A) | \eta = y) \mathbb{P}_\eta(dy) \\ &= \mathbb{E}I(\xi \in A) I(\eta \in B) \end{aligned}$$

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

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

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

$$P_{\xi}(B | \eta = y) = \int_B \frac{p_{(\xi,\eta)}(x, y)}{p_{\eta}(y)} I(p_{\eta}(y) \neq 0) dx$$

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

$$\begin{aligned} P(\xi \in B, \eta \in A) &= EI(\xi \in B, \eta \in A) = EI(\xi \in B)I(\eta \in A) = \\ &= \int_A P_{\xi}(B | \eta = y) P_{\eta}(dy) = \int_A P_{\xi}(B | \eta = A) p_{\eta}(y) dy \end{aligned}$$

А с другой

$$\begin{aligned} P(\xi \in B, \eta \in A) &= \int_{B \times A} p_{(\xi,\eta)}(x, y) dx dy = \\ &= \int_A \left[\int_B p_{(\xi,\eta)}(x, y) dx \right] dy = \int_A \left[\int_B \frac{p_{(\xi,\eta)}(x, y)}{p_{\eta}(y)} I(p_{\eta}(y) \neq 0) dx \right] p_{\eta}(y) dy, \end{aligned}$$

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 $E(g(\xi) | \eta)$.

1. Считаём условную плотность $p_{(\xi|\eta)}(x | y)$.
2. Находим функцию φ , для которой $\varphi(y) = E(g(\xi) | \eta = y) = \int_{\mathbb{R}} g(x) p_{(\xi|\eta)}(x | y) dx$.
3. $E(g(\xi) | \eta) = \varphi(\eta)$.

7.3 Поиск наилучшей оценки в классе несмещенных оценок

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

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

Теорема 7.5. (Критерий факторизации Неймана-Фишера)

Пусть $\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$ — 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 некоторых функций h и g .

Доказательство. Рассмотрим только дискретный случай.

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

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

Пусть наоборот, $P_\theta(X = x) = h(x)g_\theta(S(x))$. Тогда

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

откуда

$$\begin{aligned} P_\theta(X = x \mid S(X) = S(x)) &= \frac{P_\theta(X = x, S(X) = S(x))}{P(S(X) = S(x))} = \frac{P_\theta(X = x, S(X) = S(x))}{\sum_{y: S(y)=S(x)} P_\theta(S(X) = S(x), X = y)} = \\ &= \frac{P_\theta(X = x)}{\sum_{y: S(y)=S(x)} P_\theta(X = y)} = \frac{h(x)g_\theta(S(x))}{\sum_{y: S(y)=S(x)} h(y)g_\theta(S(x))} = \frac{h(x)}{\sum_{y: S(y)=S(x)} h(y)} \end{aligned}$$

□

Теорема 7.6. (Рао-Блэквелла-Колмогорова)

Пусть $\hat{\theta}$ — несмещенная оценка θ и $\forall \theta : D_\theta \hat{\theta} < +\infty$, а $S(X)$ — достаточная статистика. Тогда для оценки $\theta^* = E(\hat{\theta} \mid S(X))$ верно:

1. θ^* не зависит от θ (как функция)
2. θ^* — несмещенная оценка θ
3. $D_\theta \theta^* \leq D_\theta \hat{\theta}$, причем равенство $\forall \theta \in \Theta \iff \theta^* = \hat{\theta}$ почти наверное

Доказательство. 1. Следствие из определения достаточной статистики.

$$2. E_\theta \theta^* = E_\theta (E_\theta(\hat{\theta} \mid S(X))) = E_\theta \hat{\theta} = \theta$$

3.

$$D_\theta(\hat{\theta}) = E_\theta(\hat{\theta} - \theta)^2 = E_\theta[(\hat{\theta} - \theta^*) + (\theta^* - \theta)]^2 = E_\theta(\hat{\theta} - \theta^*)^2 + D_\theta \theta^* + 2E_\theta[(\hat{\theta} - \theta^*)(\theta^* - \theta)] = D_\theta \theta^* + \underbrace{E_\theta(\hat{\theta} - \theta^*)^2}_{\geq 0}$$

поскольку

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

причем $D_\theta \hat{\theta} = D_\theta \theta^* \Leftrightarrow \mathbb{E}_\theta (\hat{\theta} - \theta^*)^2 = 0 \Leftrightarrow \hat{\theta} = \theta^* \text{ п.н.}$ □

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

Лемма 7.2. Если $S(X)$ — полная достаточная статистика и для некоторой функции φ верно равенство $\mathbb{E}_\theta \varphi(S(X)) = \theta \forall \theta \in \Theta$, 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 $\forall \theta \in \Theta :$

$$\mathbb{E}_\theta \varphi(S(X)) = \mathbb{E}_\theta \psi(S(X)) = \theta$$

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

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

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 $\mathbb{E}_\theta \varphi(S(X)) = \theta \forall \theta \in \Theta$

Определение 7.10. Пусть $\theta \in \Theta \subseteq \mathbb{R}^k$, X — наблюдение с плотностью p_θ из распределения $P \in \mathcal{P}$, 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, \dots, u_k — борелевские функции. Тогда семейство распределений \mathcal{P} принадлежит *экспоненциальному классу распределений*.

Теорема 7.7. (б/д)

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

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

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

Статистика $S(X) = (\sum X_i^2, \sum X_i)$ является достаточной, причём $E_\theta \sum X_i^2 = n(a^2 + \sigma^2)$, 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 $P \in \{P_\theta \mid \theta \in \Theta\}$ и $\Theta \subset \mathbb{R}$. Доверительным интервалом уровня $\gamma \in (0, 1)$ называется такая пара статистик $(T_1(X), T_2(X))$, что

$$\forall \theta \in \Theta \quad P_\theta(\theta \in (T_1(X), T_2(X))) \geq \gamma$$

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

$$P_\theta(z_{\gamma_1} \leq G(X, \theta) \leq z_{\gamma_2}) \geq \gamma_2 - \gamma_1 = \gamma$$

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

$$P_\theta(T_1(X) < \theta < T_2(X)) = P_\theta(z_{\gamma_1} < G(X, \theta) < z_{\gamma_2}) \geq \gamma$$

Пример 8.1. Пусть $X_1, \dots, X_n \sim \mathcal{N}(b, \sigma^2)$. 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

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

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

Лемма 8.1. Пусть y случайной величины X непрерывная функция распределения F и X_1, \dots, X_n — н.о.р. случайные величины. Тогда

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

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

Следствие 8.0.1. Пусть $\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$ 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, \dots, X_n — выборка из распределения $P \in \{P_\theta \mid \theta \in \Theta \subset \mathbb{R}\}$. Тогда последовательность пар статистик $(T_1^n(X), T_2^n(X))$ называется *асимптотическим доверительным интервалом* уровня γ , если

$$\lim_{n \rightarrow +\infty} P_\theta (\theta \in [T_1^n(X), T_2^n(X)]) \geq \gamma$$

Асимптотический доверительный интервал называется *точным*, если равенство обращается в равенство, а \lim превращается в \lim .

Пример 8.2. Пусть X_1, \dots, X_n имеет распределение P_θ с $E_\theta X = \theta$, $D_\theta X = \sigma^2(\theta) > 0$ — непрерывная функция. По ЦПТ выполнена сходимость

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

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

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

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

$$P_\theta \left(-z_{\frac{1+\gamma}{2}} \leq \sqrt{n} \frac{\bar{X} - \theta}{\sigma(\bar{X})} \leq z_{\frac{1+\gamma}{2}} \right) = P_\theta \left(\underbrace{\bar{X} - z_{\frac{1+\gamma}{2}} \frac{\sigma(\bar{X})}{\sqrt{n}}}_{T_1} \leq \theta \leq \underbrace{\bar{X} + z_{\frac{1+\gamma}{2}} \frac{\sigma(\bar{X})}{\sqrt{n}}}_{T_2} \right) \rightarrow \gamma \quad \forall \theta$$

Замечание. $T_2 - T_1 \rightarrow 0$

9 Байесовские методы

9.1 Введение

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

$$P(D_n | A) = \frac{P(A | D_n)P(D_n)}{\sum_{i=0}^{\infty} P(A | D_i)P(D_i)} \quad (3)$$

Определение 9.1. Назовем A — *результатом эксперимента*, $P(D_n)$ — *априорная* вероятность — известная *до* эксперимента. $P(D_n | A)$ — *апостериорная* вероятность — *после* эксперимента.

Пусть ξ, η — случайные величины на (Ω, \mathcal{F}, P) . Тогда формула Байеса в общем виде:

$$p_{\eta|\xi}(y | x) = \frac{p_{\xi|\eta}(x | y)p_{\eta}(y)}{\int_{\mathbb{R}^n} p_{\xi|\eta}(x | y)p_{\eta}(y)} \quad (4)$$

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 байесовским методам на ПМИ ФИВТ.

Замечание. Баесовские методы в BigData используются, например, в задаче распознавания 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 $P \in \mathcal{P} = \{P_t : t \in \Theta\}$, 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 (θ, X) .

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

$$q(t | x) = \frac{q(t)p_t(x)}{\int_{\Theta} q(s)p_s(x)dx},$$

называется *апостериорной* плотностью.

Способы оценивания θ .

1. Апостериорное 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:
 - (а) $E(\theta | X)$ — математическое ожидание по апостериорному распределению.
 - (б) $\arg \max_{t \in \Theta} q(t | x)$ — 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 $n \rightarrow +\infty$	Верны при $n \geq 0$.

Теорема 9.1. Оценка $E(\theta|X)$ — наилучшая оценка параметра θ в байесовском подходе с квадратичной функцией потерь.

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

$$\int_{\Theta} R(\hat{\theta}, t)q(t)dt \rightarrow \max.$$

Перепишем интеграл

$$\int_{\Theta} E_t (\hat{\theta} - t)^2 q(t)dt = \int_{\Theta} \int_{\mathcal{X}} (\hat{\theta}(x) - t)^2 f(t, x)dxdt = E (\hat{\theta}(x) - \theta)^2 \rightarrow \max_{\hat{\theta}}.$$

Применяя теорему о наилучшем приближении

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

□

У байесовского 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. Большие вычислительные затраты.

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

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

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, \dots, X_n — выборка из распределения $P \in \mathcal{P} = \{P_t \mid t \in \Theta\}$ — 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 $X_1, \dots, X_n \sim Exp(\theta)$. 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 коэффициент 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

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

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

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

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 веса. 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 $\{y_1, \dots, y_5\}$ и десять раз оба груза вместе с весами $\{z_1, \dots, z_{10}\}$. Причем из-за погрешности измерений все числа x_i, y_i, z_i различны. Условие задачи можно представить следующим образом:

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

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

Пример 10.2. Пусть случайная величина X зависит от времени по закону $a_3 t^3 + a_2 t^2 + a_1 t + a_0$, где a_i неизвестны и необходимо найти их оценку. В разные моменты времени t_i были проведены измерения величины X и получены результаты $X_i = a_3 t_i^3 + a_2 t_i^2 + a_1 t_i + a_0 + \varepsilon_i$. 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 $L = \langle z_1, \dots, z_k \rangle$ — k -мерное подпространство в \mathbb{R}^n . Пусть $Z = (z_1, \dots, z_k)$ — 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 $E\varepsilon_i = 0$ и $D\varepsilon_i = \sigma^2 > 0$, где σ^2 неизвестно. Задача линейной регрессии заключается в нахождении оценок θ и σ^2 .

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

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

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 \hat{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

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

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

Доказательство.

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

□

Найдем дисперсию $\mathbf{D}_\theta \hat{\theta}$:

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

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

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

Тогда $\text{tr } A = \text{tr } Z(Z^T Z)^{-1} Z^T = \text{tr } (Z^T Z)^{-1} (Z^T Z) = k$.

Поскольку $\hat{\theta}$ — несмещенная оценка, то $\mathbf{E}_\theta (X - Z\hat{\theta}) = 0$, откуда $\text{tr } \mathbf{D}_\theta (X - Z\hat{\theta}) = \mathbf{E}_\theta \|X - Z\hat{\theta}\|^2$.

$$\begin{aligned} \text{tr } \mathbf{D}_\theta (X - Z\hat{\theta}) &= \text{tr } \mathbf{D}_\theta (E - A)X = \text{tr } [(E - A)\mathbf{D}_\theta X (E - A)^T] = \\ &= \text{tr } [(E - A)\sigma^2] = n\sigma^2 - \sigma^2 \text{tr } A = n\sigma^2 - \sigma^2 k = (n - k)\sigma^2. \end{aligned}$$

поскольку $A^2 = A$.

□

Следствие 10.0.1. $\frac{1}{n-k} \|X - Z\hat{\theta}\|^2 = \hat{\sigma}^2$ — 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 $\mathbb{R}^n = L_1 \oplus \dots \oplus L_r$; $\dim L_i = k_i$; $l_i = \text{proj}_{L_i} l$ и $X_i = \text{proj}_{L_i} X$ — ортогональные проекции вектора X .

Тогда X_1, \dots, X_r — независимые случайные вектора и

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

где

$$\chi_k^2 = \Gamma\left(\frac{1}{2}; \frac{k}{2}\right) \stackrel{d}{=} \xi_1^2 + \dots + \xi_k^2,$$

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

Рассмотрим плотность выборки:

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

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

Теорема 10.2. (*б/д*)

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

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

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

$$\begin{aligned} \hat{\theta} &= (Z^T Z)^{-1} Z^T \text{proj}_L X \\ \hat{\sigma}^2 &= \frac{1}{n-k} \|X - Z\hat{\theta}\|^2 = \frac{1}{n-k} \|X - \text{proj}_L X\|^2 \end{aligned}$$

□

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

Доказательство. По теореме 10.1:

$$Z\hat{\theta} = \text{proj}_L X \perp \text{proj}_{L^\perp} X = X - Z\hat{\theta}.$$

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

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

□

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

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

имеет *распределение Стьюдента* с k степенями свободы.

Пусть $\xi \sim \chi_k^2$, $\eta \sim \chi_m^2$, $\xi \perp \eta$. 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 $\frac{1}{\sigma^2} \|X - Z\hat{\theta}\|^2 \sim \chi^2(n - k)$, то достаточно взять квантиль $u_{1-\gamma}$ распределения $\chi^2(n - k)$,

а значит

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

Доверительный интервал для $\hat{\theta}_i$:

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

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

откуда

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

где u_p — квантили T_{n-k} .

Доверительная область для θ :

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

11 Проверка гипотез

11.1 Построение критериев

Обозначим неизвестное распределение P . Тогда гипотезой назовем любое утверждение относительно P и обозначим $H : P \in \mathcal{P}$. 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 $H_0 : P \in \mathcal{P}_0$. Тогда H_0 называется *основной гипотезой*. Противоречущую ей гипотезу $H_1 : P \in \mathcal{P}_1$ 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 $\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$ и $H_0 : P = P_{\theta_0}$, $H_1 : P \neq P_{\theta_0}$. Построим для θ доверительный интервал $(T_1(X), T_2(X))$ уровня γ . Тогда

$$P_\theta \left(\underbrace{\theta \in (T_1(X), T_2(X))}_A \right) \geq \gamma.$$

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

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

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

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

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

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

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

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

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

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

Теорема 11.1. (*лемма Неймана-Пирсона*)

Пусть R — критерий, такой что

$$P_0(X \in R) \leq P_0(X \in S_\lambda)$$

(т.е. уровня значимости $P_0(X \in S_\lambda)$). Тогда критерий S_λ мощнее критерия R

$$P_1(X \in R) \leq P_1(X \in S_\lambda)$$

и, кроме того, S_λ — несмещенный критерий.

Доказательство. По свойствам индикаторов

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

Тогда

$$\begin{aligned} P_1(x \in R) - \lambda P_0(x \in R) &= \int (p_1(x) - \lambda p_0(x)) I(x \in R) \mu(dx) \\ &\leq \int I(X \in S_\lambda)(p_1(x) - \lambda p_0(x)) \mu(dx) = P_1(x \in S_\lambda) - \lambda P_0(x \in S_\lambda), \end{aligned}$$

откуда

$$P_1(x \in R) - P_1(x \in S_\lambda) \leq \lambda (P_0(x \in R) - P_0(x \in S_\lambda)) \leq 0 \Rightarrow P_1(x \in R) \leq P_1(x \in S_\lambda).$$

Рассмотрим теперь два случая:

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

$$P_0(x \in S_\lambda) = \int I(x \in S_\lambda) p_0(x) \mu(dx) \leq \int I(x \in S_\lambda) p_1(x) \mu(dx) = P_1(x \in S_\lambda).$$

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

$$\begin{aligned} P_0(x \notin S_\lambda) &= \int I(x \notin S_\lambda) p_0(x) \mu(dx) \geq \\ &\geq \int I(x \notin S_\lambda) p_1(x) \mu(dx) = P_1(x \notin S_\lambda), \end{aligned}$$

откуда

$$P_0(x \in S_\lambda) = 1 - P_0(x \notin S_\lambda) \leq 1 - P_1(x \notin S_\lambda) = P_1(x \in S_\lambda).$$

□

Следствие 11.1.1. Пусть λ таково, что

$$P_0(x \in S_\lambda) = \varepsilon.$$

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

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

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 $T(X) \geq \tilde{\lambda}$. Если H_0 верна, то $\sum X_i \sim \text{Bin}(n, \frac{1}{4})$ и $\tilde{\lambda}$ — квантиль уровня $1 - \varepsilon$ распределения $\text{Bin}(n, \frac{1}{4})$. Причем все проведенные рассуждения верны только для тех ε , где достигается равенство $P_0(x \in S_\lambda) = \varepsilon$.

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

Пусть $\Theta \subset \mathbb{R}$, $H_0 : \theta \leq \theta_0$ ($\theta = \theta_0$), $H_1 : \theta > \theta_0$. 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

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

если

$$P_0(S_\varepsilon) = \varepsilon.$$

Пример 11.3. $X_1, \dots, X_n \sim \text{Bern}(\theta)$, $H_0 : \theta \geq \frac{1}{4}$, $H_1 : \theta < \frac{1}{4}$. Сделаем замену $\tilde{\theta} = -\theta$. Тогда $H_0 : \tilde{\theta} \leq -\frac{1}{4}$, $H_1 : \tilde{\theta} > -\frac{1}{4}$. 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 $S = \{\sum X_i \leq c\}$, где c — квантиль уровня ε для $\text{Bin}(n, \frac{1}{4})$.

Пусть X_1, \dots, 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 $S = \{X_{(n)} \leq \theta_0 \varepsilon^{\frac{1}{n}}\}$.

Доказательство.

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

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

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

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

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

$$\begin{aligned} \mathbb{P}_{\theta}(x \in R) &= \int_{[0, \theta]^n} \frac{1}{\theta^n} I(X_1, \dots, X_n \in R) dx_1 \dots dx_n = \left(\frac{\theta_0}{\theta} \right)^n \int_{[0, \theta_0]^n} \frac{1}{\theta_0^n} I(X_1, \dots, X_n \in R) dx_1 \dots dx_n \\ &\leq \left(\frac{\theta_0}{\theta} \right)^n \int_{[0, \theta_0]^n} \frac{1}{\theta_0^n} I(X_1, \dots, X_n \in R) dx_1 \dots dx_n = \left(\frac{\theta_0}{\theta} \right)^n \mathbb{P}_{\theta_0}(X \in R) \leq \varepsilon \left(\frac{\theta_0}{\theta} \right)^n \end{aligned}$$

□

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 $(m \times k)$, $m \leq k$, $\text{rk } T = m$. Как мы помним,

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

а, зная это, имеем

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

Матрица B обратна как матрица с полным рангом, а значит

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

а

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

Зная всё это, получаем, что

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

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

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

Пример 11.4. $X_1, \dots, X_n \sim \mathcal{N}(x, \sigma^2)$, $Y_1, \dots, Y_m \sim \mathcal{N}(y, \sigma^2)$. 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

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

и

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

11.3 Критерии согласия

Критерий хи-квадрат Пирсона

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

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

Теорема 11.3. (Пирсон)

При условии верности H_0 , выполнена сходимость

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

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

$$DY_j = B - pp^T, \text{ где } B = \text{diag}(p_1^0, \dots, p_m^0).$$

По ЦПТ

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

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

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

По теореме о наследовании сходимости

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

□

Следствие 11.3.1. Гипотеза H отвергается на уровне значимости ε , если $\hat{\mu} > u_{1-\varepsilon}$, 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 $P(\hat{\mu} > u_{1-\varepsilon}) \rightarrow 1$. По УЗБЧ $\frac{\mu_i}{n} \xrightarrow{\text{п.н.}} p_i$, а значит

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

В частности, при $i = 1$:

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

□

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: $a_1 = \text{"да, да"}$, $a_2 = \text{"да, нет"}$, $a_3 = \text{"нет, нет"}$. 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 Евгеньевич получил следующие результаты: $\mu_1 = 28$, $\mu_2 = 20$, $\mu_3 = 12$. В таком случае,

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

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 $H : P = P_0$ и $\alpha = P_0(S(x) > u)$ — его уровень значимости. Найдем значение $S(x)$ для выборки X_1, \dots, X_n : $S(X_1, \dots, X_n) = t$. Величина $p = P_0(S(x) > t)$ 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

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

Рассмотрим статистику $H : P = P_0$. Пусть $S(X) = \sqrt{n} \sup_x |F(x) - F_n(x)|$ 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.

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

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)

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

где $\xi \sim a1$.

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

Определение 11.11. Все три критерия (Пирсона, Колмогорова-Смирнова, Мизеса-Смирнова) называются *критериями согласия*, поскольку проверяют гипотезу вида $H : P = P_0$.

11.4 Байесовские критерии

Пусть мы хотим проверить гипотезу $H_0 : P = P_0$ против альтернативы $H_1 : P = P_1$, где P_0, P_1 — доминируемые относительно меры μ . Пусть Q — априорное распределение, и $Q(P = P_0) = p_0$, $Q(P = P_1) = p_1$. 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

$$p_0 P_0(X \in S_0) + p_1 P_1(X \in S_1) \rightarrow \min_{S_0, S_1}$$

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

Пусть $S = \begin{cases} S_0 & P = P_0, \\ S_1 & P = P_1 \end{cases}$ — случайное множество. Имеем

$$P_0(X \in S_0) = E I(X \in S) = E (E(I(X \in S) | X)).$$

Найдем условное мат.ожидание

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

Тогда

$$\max_{S_0, S_1} E(I(x \in S_1) q_0 + I(x \in S_0) q_1) \leq E \max\{q_0, q_1\}$$

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

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

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