МЕХАНИКО-
МАТЕМАТИЧЕСКИЙ
ФАКУЛЬТЕТ
МГУ ИМЕНИ
М.В. ЛОМОНОСОВА

teach-in
ЛЕКЦИИ УЧЕНЫХ МГУ

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА. СЕМИНАРЫ

РЯДНОВА
ЕКАТЕРИНА МИХАЙЛОВНА

МЕХМАТ МГУ

КОНСПЕКТ ПОДГОТОВЛЕН СТУДЕНТАМИ, НЕ ПРОХОДИЛ ПРОФ. РЕДАКТУРУ И МОЖЕТ СОДЕРЖАТЬ ОШИБКИ. СЛЕДИТЕ ЗА ОБНОВЛЕНИЯМИ НА VK.COM/TEACHINMSU.

ЕСЛИ ВЫ ОБНАРУЖИЛИ ОШИБКИ ИЛИ ОПЕЧАТКИ, ТО СООБЩИТЕ ОБ ЭТОМ, НАПИСАВ СООБЩЕСТВУ VK.COM/TEACHINMSU.

БЛАГОДАРИМ ЗА ПОДГОТОВКУ КОНСПЕКТА
ВЫПУСКНИКА ФИЗИЧЕСКОГО ФАКУЛЬТЕТА МГУ
ФИЛИППОВА ВЛАДИСЛАВА ИГОРЕВИЧА

Содержание

Семинар 1

Сходимости случайных величин

Пусть ξ_n, $\xi \in \mathbb{R}^1$ – случайные величины.

Определение 1.1. Говорят, что ξ_n *сходится к* ξ *почти наверное* (*с вероятностью 1*), если

$$P(\omega\colon\ \xi_n(\omega) \to \xi(\omega)) = 1.$$

Обозначение: $\xi_n \overset{\text{п. н.}}{\to} \xi$.

Определение 1.2. Говорят, что ξ_n *сходится к* ξ *по вероятности*, если

$$P(\omega\colon\ |\xi_n(\omega) - \xi(\omega)| > \varepsilon) \to 0 \quad \forall \varepsilon > 0.$$

Обозначение: $\xi_n \overset{P}{\to} \xi$.

Определение 1.3. $F_\xi(x) \coloneqq P(\xi \leqslant x)$ – *функция распределения* ξ.

Определение 1.4. $\varphi_\xi(t) \coloneqq Ee^{it\xi}$ – *характеристическая функция* ξ.

Определение 1.5. Говорят, что ξ_n *сходится к* ξ *по распределению* (*слабая сходимость*), если

1) $F_{\xi_n}(x) \to F_\xi(x) \quad \forall x \in C(F_\xi)$, или
2) $\varphi_{\xi_n}(t) \to \varphi_\xi(t) \quad \forall t \in \mathbb{R}$, или
3) $Eg(\xi_n) \to Eg(\xi) \quad \forall$ ограниченных непрерывнвых функций g.

Эти три определения сходимости по распределению эквивалентны.

Обозначение: $\xi_n \overset{d}{\to} \xi$.

Утверждение 1.1.

1) Если $\xi_n \overset{\text{п. н.}}{\to} \xi$, *то* $\xi_n \overset{P}{\to} \xi$.

2) Если $\xi_n \overset{P}{\to} \xi$, *то* $\xi_n \overset{d}{\to} \xi$.

Утверждение 1.2. *Предел сходимости почти наверное определяется однозначно.*

Доказательство:

Мера ω, в которых есть сходимость, равна 1. При каждом элементарном исходе ω числовая последовательность $\xi_n(\omega)$ имеет единственный предел. Значит, на множестве вероятности 1 случайная величина $\xi(\omega)$ определяется однозначно. Если существует другой предел последовательности ξ_n, сходящейся почти наверное, то он должен отличаться от ξ только на множестве меры 0. ■

Задача 1.1. Пусть $\xi_n \overset{P}{\to} \xi$ и $\xi_n \overset{P}{\to} \eta$. Доказать, что $P(\xi = \eta) = 1$.

МЕХАНИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ МГУ ИМЕНИ М.В. ЛОМОНОСОВА
teach-in ЛЕКЦИИ УЧЕНЫХ МГУ

Решение:

$$\forall\varepsilon>0 \qquad P(|\xi-\eta|>\varepsilon)=P(|(\xi_n-\eta)-(\xi_n-\xi)|>\varepsilon) \leqslant$$

Так как $|\xi_n-\eta|+|\xi_n-\xi|\geqslant|(\xi_n-\eta)-(\xi_n-\xi)|>\varepsilon$, то из $|(\xi_n-\eta)-(\xi_n-\xi)|>\varepsilon$ следует, что $|\xi_n-\eta|>\dfrac{\varepsilon}{2}$ или $|\xi_n-\xi|>\dfrac{\varepsilon}{2}$.

$$\leqslant P\left(|\xi_n-\eta|>\frac{\varepsilon}{2}\right)+P\left(|\xi_n-\xi|>\frac{\varepsilon}{2}\right)\to 0 \quad \text{при } n\to+\infty,$$

так как $\xi_n \xrightarrow{P} \xi$ и $\xi_n \xrightarrow{P} \eta$.

Так как $P(|\xi-\eta|>\varepsilon)$ не зависит от n, то $P(|\xi-\eta|>\varepsilon)=0$. В силу произвольности выбора $\varepsilon>0$ получаем, что

$$P(\xi\neq\eta)=P\left(\bigcup_{k\geqslant1}\left\{|\xi-\eta|>\frac{1}{k}\right\}\right)=0.$$

Значит, $P(\xi=\eta)=1$. □

Задача 1.2. Доказать, что $\xi_n \xrightarrow{d} C \iff \xi_n \xrightarrow{P} C$, где C – константа.

Решение:

Утверждение $\Longleftarrow$ считаем известным из утверждения (1.1).

Докажем утверждение $\Longrightarrow$.

$$\forall\varepsilon>0 \qquad P(|\xi_n-C|>\varepsilon)=1-P(|\xi_n-C|\leqslant\varepsilon)=1-P(C-\varepsilon\leqslant\xi_n\leqslant C+\varepsilon)\leqslant$$
$$\leqslant 1-P(C-\varepsilon<\xi_n\leqslant C+\varepsilon)=1-F_{\xi_n}(C+\varepsilon)+F_{\xi_n}(C-\varepsilon) \ominus$$

Так как $F_C(x)=0\ \ \forall x<C$, $F_C(x)=1\ \ \forall x\geqslant C$ и $\xi_n \xrightarrow{d} C$, то $F_{\xi_n}(x)\to F_C(x)\ \ \forall x\neq C$. Значит, $F_{\xi_n}(C-\varepsilon)\to 0$ и $F_{\xi_n}(C+\varepsilon)\to 1$.

$$\ominus\ 1-1+0=0.$$

Значит, $\xi_n \xrightarrow{P} C$. □

Задача 1.3. Пусть $E\xi_n\to C$ и $D\xi_n\to 0$. Доказать, что $\xi_n \xrightarrow{P} C$.

Решение:

Используя неравенство Маркова, получаем:

$$\forall\varepsilon>0 \qquad P(|\xi_n-C|>\varepsilon)\leqslant\frac{E|\xi_n-C|^2}{\varepsilon^2}=\frac{D(\xi_n-C)+(E(\xi_n-C))^2}{\varepsilon^2}\to 0,$$

так как $D(\xi_n-C)=D\xi_n\to 0$ и $E(\xi_n-C)=E\xi_n-EC=E\xi_n-C\to C-C=0$.

Таким образом, $\xi_n \xrightarrow{P} C$. □

Сходимости случайных векторов

Пусть ξ_n, $\xi \in \mathbb{R}^k$ – случайные векторы.

Определение 1.6. Говорят, что ξ_n *сходится к* ξ *почти наверное* (*с вероятностью 1*), если

$$P(\xi_n \to \xi) = P(|\xi_n - \xi| \to 0) = 1.$$

где $|\cdot|$ – евклидова норма.

Обозначение: $\xi_n \overset{\text{п. н.}}{\to} \xi$.

Задача 1.4. Доказать, что

$$\xi_n \overset{\text{п. н.}}{\to} \xi \iff \text{покомпонентной сходимости почти наверное.}$$

Решение:

Пусть $\xi_n = \begin{pmatrix} \xi_n^{(1)} \\ \vdots \\ \xi_n^{(k)} \end{pmatrix}$ и $\xi = \begin{pmatrix} \xi^{(1)} \\ \vdots \\ \xi^{(k)} \end{pmatrix}$.

Докажем утверждение $\implies$. Если длина вектора $\xi_n - \xi$ стремится к 0 на множестве вероятности 1, то на том же множестве элементарных исходов разница между любыми соответствующими компонентами ξ_n и ξ тоже будет стремиться к 0.

Теперь докажем утверждение $\impliedby$. Так как $\xi_n^{(i)} \underset{n\to+\infty}{\to} \xi^{(i)}$ $\forall i = \overline{1, k}$, то $P\left(\Omega^{(i)}\right) = 1$, где $\Omega^{(i)}$ – множество сходимости i-й компоненты. Тогда $P\left(\bigcap\limits_{i=1}^{k} \Omega^{(i)}\right) = 1$. Значит, $P(|\xi_n - \xi| \to 0) = 1$. □

Определение 1.7. Говорят, что ξ_n *сходится к* ξ *по вероятности*, если

$$P(|\xi_n - \xi| > \varepsilon) \to 0 \quad \forall \varepsilon > 0.$$

Обозначение: $\xi_n \overset{P}{\to} \xi$.

Задача 1.5. Доказать, что

$$\xi_n \overset{P}{\to} \xi \iff \text{покомпонентной сходимости по вероятности.}$$

Решение:

Пусть $\xi_n = \begin{pmatrix} \xi_n^{(1)} \\ \vdots \\ \xi_n^{(k)} \end{pmatrix}$ и $\xi = \begin{pmatrix} \xi^{(1)} \\ \vdots \\ \xi^{(k)} \end{pmatrix}$.

Докажем утверждение $\implies$. Так как $|\xi_n^{(i)} - \xi^{(i)}| \leqslant |\xi_n - \xi|$, то

$$\forall \varepsilon > 0 \qquad P(|\xi_n^{(i)} - \xi^{(i)}| > \varepsilon) \leqslant P(|\xi_n - \xi| > \varepsilon) \to 0 \qquad \forall i = \overline{1, k}.$$

Теперь докажем утверждение $\Longleftarrow$. Так как

$$|\xi_n - \xi| = \sqrt{\left(\xi_n^{(1)} - \xi^{(1)}\right)^2 + \ldots + \left(\xi_n^{(k)} - \xi^{(k)}\right)^2},$$

то при $|\xi_n - \xi| > \varepsilon$ имеем:

$$\left(\xi_n^{(1)} - \xi^{(1)}\right)^2 + \ldots + \left(\xi_n^{(k)} - \xi^{(k)}\right)^2 > \varepsilon^2.$$

Тогда хотя бы для одного значения i верно, что

$$\left(\xi_n^{(i)} - \xi^{(i)}\right)^2 > \frac{\varepsilon^2}{k}, \qquad \text{то есть} \qquad |\xi_n^{(i)} - \xi^{(i)}| > \frac{\varepsilon}{\sqrt{k}}.$$

Значит, получаем:

$$P(|\xi_n - \xi| > \varepsilon) \leqslant P\left(|\xi_n^{(1)} - \xi^{(1)}| > \frac{\varepsilon}{\sqrt{k}}\right) + \ldots + P\left(|\xi_n^{(k)} - \xi^{(k)}| > \frac{\varepsilon}{\sqrt{k}}\right) \to 0,$$

так как $\xi_n^{(i)} \xrightarrow{P} \xi^{(i)} \quad \forall i = \overline{1, k}$. □

Определение 1.8. $F_\xi(x) \coloneqq P(\xi \leqslant x)$ – *функция распределения* ξ.

Определение 1.9. $\varphi_\xi(t) \coloneqq Ee^{it^T\xi}$ – *характеристическая функция* ξ.

Определение 1.10. Говорят, что ξ_n *сходится к* ξ *по распределению* (*слабая сходимость*), если

1) $F_{\xi_n}(x) \to F_\xi(x) \quad \forall x \in C(F_\xi)$, или
2) $\varphi_{\xi_n}(t) \to \varphi_\xi(t) \quad \forall t \in \mathbb{R}^k$, или
3) $Eg(\xi_n) \to Eg(\xi) \quad \forall$ ограниченных непрерынвных функций g.

Эти три определения сходимости по распределению эквивалентны.
Обозначение: $\xi_n \xrightarrow{d} \xi$.

Задача 1.6. Доказать, что

$$\xi_n \xrightarrow{d} \xi \implies \text{покомпонентная слабая сходимость.}$$

Решение:

$$\varphi_{\xi_n}(t) = Ee^{it^T\xi_n} \to \varphi_\xi(t) \qquad \forall t \in \mathbb{R}^k.$$

Рассмотрим векторы t вида $t^T = \begin{pmatrix} 0 & \ldots & 0 & t^{(i)} & 0 & \ldots & 0 \end{pmatrix}$. Тогда

$$\varphi_{\xi_n}(t) = Ee^{it^T\xi_n} = Ee^{it^{(i)}\xi^{(i)}} = \varphi_{\xi^{(i)}}\left(t^{(i)}\right) \to \varphi_{\xi^{(i)}}\left(t^{(i)}\right) \qquad \forall i = \overline{1, k}.$$

□

Замечание 1.1. Обратное (то есть $\Longleftarrow$), вообще говоря, неверно.

Задача 1.7 (Приём Крамера-Уолда)**.** Доказать, что

$$\xi_n \xrightarrow{d} \xi \iff \lambda^T\xi_n \xrightarrow{d} \lambda^T\xi \quad \forall\lambda \in \mathbb{R}^k.$$

Решение:

$$\begin{aligned}
&\forall t \in \mathbb{R} \qquad \varphi_{\lambda^T\xi_n}(t) = Ee^{it\lambda^T\xi_n} = \varphi_{\xi_n}(t\lambda);\\
&\forall t \in \mathbb{R} \qquad \varphi_{\lambda^T\xi}(t) = Ee^{it\lambda^T\xi} = \varphi_{\xi}(t\lambda).
\end{aligned}$$

Докажем утверждение $\Longleftarrow$. Так как $\lambda^T\xi_n \xrightarrow{d} \lambda^T\xi$, то $\varphi_{\lambda^T\xi_n}(t) \to \varphi_{\lambda^T\xi}(t)$. Тогда $\varphi_{\xi_n}(t\lambda) \to \varphi_{\xi}(t\lambda)$. В частности, при $t = 1$ получаем: $\varphi_{\xi_n}(\lambda) \to \varphi_{\xi}(\lambda) \ \ \forall\lambda \in \mathbb{R}^k$.

Докажем утверждение $\Longrightarrow$. Так как $\xi_n \xrightarrow{d} \xi$, то $\varphi_{\xi_n}(t\lambda) \to \varphi_{\xi}(t\lambda)$. Тогда $\varphi_{\lambda^T\xi_n}(t) \to \varphi_{\lambda^T\xi}(t) \ \ \forall t \in \mathbb{R}$. □

Задача 1.8. Привести пример, что $\xi_n \xrightarrow{d} \xi$ и $\eta_n \xrightarrow{d} \eta$, но $\begin{pmatrix}\xi_n\\ \eta_n\end{pmatrix} \overset{d}{\not\to} \begin{pmatrix}\xi\\ \eta\end{pmatrix}$.

Решение:

Рассмотрим случайную величину, имеющую стандартное нормальное распределение: $X \sim \mathcal{N}(0,1)$. Любые линейные комбинации нормально распределённой случайной величины будут иметь нормальное распределение. Тогда $-X \sim \mathcal{N}(0,1)$.

Пусть $\xi_n \coloneqq X$, тогда $\xi_n \xrightarrow{d} X$. Пусть $\eta_n \coloneqq -X$, тогда $\eta_n \xrightarrow{d} X$.

Но $\begin{pmatrix}X\\ -X\end{pmatrix} \overset{d}{\not\to} \begin{pmatrix}X\\ X\end{pmatrix}$. Это очевидно, так как эти векторы могут лежать только в разных парах квадрантов. Но покажем это, используя приём Крамера-Уолда из задачи (1.7).

Предположим, что $\begin{pmatrix}X\\ -X\end{pmatrix} \xrightarrow{d} \begin{pmatrix}X\\ X\end{pmatrix}$. Тогда возьмём следующую линейную комбинацию: $X - X \xrightarrow{d} X + X$, то есть $0 \xrightarrow{d} 2X$. Получили противоречие, так как $2X \sim \mathcal{N}(0,4)$. Таким образом, $\begin{pmatrix}X\\ -X\end{pmatrix} \overset{d}{\not\to} \begin{pmatrix}X\\ X\end{pmatrix}$. □

Закон больших чисел и центральная предельная теорема

Утверждение 1.3 (Закон больших чисел)**.** *Пусть X_1, X_2, ... – последовательность независимых одинаково распределённых случайных (векторных) величин, причём $E|X_1| < +\infty$. Тогда*

$$\overline{X}_n \coloneqq \frac{X_1 + \ldots + X_n}{n} \xrightarrow{P} EX_1.$$

Утверждение 1.4 (Усиленный закон больших чисел)**.** *Пусть* $X_1,\ X_2,\ \ldots$ *– последовательность независимых одинаково распределённых случайных (векторных) величин, причём* $E|X_1| < +\infty$*. Тогда*

$$\overline{X}_n \coloneqq \frac{X_1 + \ldots + X_n}{n} \overset{п.\,н.}{\to} EX_1.$$

Замечание 1.2. Закон больших чисел и усиленный закон больших чисел следуют из соответствующих законов в одномерном случае и равносильности сходимости случайных векторов по вероятности и сходимости почти наверное соответствующим покомпонентным сходимостям.

Определение 1.11.

$$DX \coloneqq E(X - EX)(X - EX)^T$$

– *ковариационная матрица* случайного вектора X.

Утверждение 1.5 (Центральная предельная теорема)**.** *Пусть* $X_1,\ X_2,\ \ldots$ *– последовательность независимых одинаково распределённых случайных (векторных) величин из* $\mathbb{R}^k$*, причём* $E|X_1|^2 < +\infty$*. Тогда*

$$\sqrt{n}\left(\overline{X}_n - EX_1\right) \overset{d}{\to} \mathcal{N}_k(0,\ DX_1).$$

Доказательство:

$$\forall \lambda \in \mathbb{R}^k \qquad \lambda^T \sqrt{n}\left(\overline{X}_n - EX_1\right) = \sqrt{n}\left(\lambda^T\, \overline{X}_n - \lambda^T EX_1\right).$$

Рассмотрим последовательность $\{\lambda^T X_i\}$ – независимо распределённых случайных (одномерных) величин. $E(\lambda^T X_1) = \lambda^T EX_1$.

$$\begin{aligned}&D(\lambda^T X_1) = E(\lambda^T X_1 - \lambda^T EX_1)^2 = E(\lambda^T X_1 - \lambda^T EX_1)(X_1^T \lambda - EX_1^T \cdot \lambda) = \\ &= \lambda^T DX_1 \lambda < +\infty,\end{aligned}$$

так как $E|X_1|^2 < +\infty$.

Тогда в силу центральной предельной теоремы в одномерном случае получаем:

$$\sqrt{n}\left(\lambda^T\, \overline{X}_n - \lambda^T EX_1\right) \overset{d}{\to} \mathcal{N}_1(0,\ \lambda^T DX_1 \lambda) \overset{d}{=} \lambda^T Y,$$

где $Y \sim \mathcal{N}_k(0,\ DX_1)$. Тогда по приёму Крамера-Уолда получаем:

$$\sqrt{n}\left(\overline{X}_n - EX_1\right) \overset{d}{\to} \mathcal{N}_k(0,\ DX_1).$$

■

Утверждения для дальнейшего использования

Теорема 1.1 (Теорема о наследовании сходимости)**.**

Пусть ξ_n, $\xi \in \mathbb{R}^k$ – случайные векторы, $h : \mathbb{R}^k \to \mathbb{R}^m$ – функция, непрерывная почти всюду относительно распределения ξ (то есть на носителе случайного вектора ξ (на множестве, в которое ξ попадает с вероятностью 1)). Тогда:

1) $\xi_n \xrightarrow{п.н.} \xi \implies h(\xi_n) \xrightarrow{п.н.} h(\xi)$;

2) $\xi_n \xrightarrow{P} \xi \implies h(\xi_n) \xrightarrow{P} h(\xi)$;

3) $\xi_n \xrightarrow{d} \xi \implies h(\xi_n) \xrightarrow{d} h(\xi)$.

Лемма 1.1 (Лемма Слуцкого)**.** *Пусть ξ_n, ξ, η_n – случайные векторы, $\xi_n \xrightarrow{d} \xi$ и $\eta_n \xrightarrow{d} C$, где C – константа. Тогда* $\begin{pmatrix} \xi_n \\ \eta_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \xi \\ C \end{pmatrix}$.

Сформулируем следствие.

Утверждение 1.6. *В условиях леммы Слуцкого (1.1) с использованием теоремы (1.1) о наследовании сходимости получаем:*

$$\xi_n + \eta_n \xrightarrow{d} \xi + C;$$
$$\xi_n \eta_n \xrightarrow{d} \xi C.$$

Замечание 1.3. Мы будем использовать утверждение (1.6) в одномерном случае, хотя оно справедливо и в многомерном.

Решение задач на сходимость статистик

Задача 1.9. Пусть X_1, X_2, … – последовательность независимых одинаково распределённых случайных величин. Пусть $S_n^2 := \dfrac{1}{n-1} \sum\limits_{i=1}^{n} \left(X_i - \overline{X}\right)^2$. Определить, к чему S_n^2 сходится по вероятности.

Решение:

$$S_n^2 = \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^{n} \left(X_i^2 - 2\overline{X} X_i + \overline{X}^2 \right) \right) = \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^{n} X_i^2 - 2\overline{X}^2 + \overline{X}^2 \right) =$$
$$= \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^{n} X_i^2 - \overline{X}^2 \right).$$

По закону больших чисел (утверждение (1.3)), если $EX_1^2 < +\infty$ (тогда и $E|X_1| < < +\infty$), то $\dfrac{1}{n} \sum\limits_{i=1}^{n} X_i^2 \xrightarrow{P} EX_1^2$ и $\overline{X} \xrightarrow{P} EX_1$. По теореме (1.1) о наследовании сходимости $-\overline{X}^2 \xrightarrow{P} -E^2 X_1$. Таким образом, $\begin{pmatrix} \frac{1}{n} \sum\limits_{i=1}^{n} X_i^2 \\ -\overline{X}^2 \end{pmatrix} \xrightarrow{P} \begin{pmatrix} EX_1^2 \\ -E^2 X_1 \end{pmatrix}$ в силу равносильности векторной сходимости и покомпонентной по вероятности (задача (1.5)). Тогда по теореме (1.1) о наследовании сходимости $\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 \xrightarrow{P} EX_1^2 - E^2X_1$. Так как $\frac{n}{n-1} \to 1$, то аналогично получаем:

$$S_n^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2\right) \xrightarrow{P} 1 \cdot (EX_1^2 - E^2X_1) = DX_1.$$

□

Замечание 1.4. При использовании усиленного закона больших чисел (утверждение (1.4)) можно показать, что $S_n^2 \xrightarrow{\text{п.н.}} DX_1$.

Утверждение 1.7 (Стохастическая формула Тейлора)**.** *Пусть ξ_n, ξ – случайные (одномерные) величины, $\xi_n \xrightarrow{d} \xi$. Пусть $h : \mathbb{R} \to \mathbb{R}$ – функция, дифференцируемая в точке a; $b_n \to 0$ – числовая последовательность, причём $b_n \neq 0$. Тогда*

$$\frac{h(a + b_n\xi_n) - h(a)}{b_n} \xrightarrow{d} h'(a) \cdot \xi.$$

Замечание 1.5. Если ξ_n и $\xi \in \mathbb{R}^k$ – случайные векторы, то:

1) если $h : \mathbb{R}^k \to \mathbb{R}$, то $h'(a)$ – вектор частных производных, а $h'(a) \cdot \xi$ – скалярное произведение;

2) если $h : \mathbb{R}^k \to \mathbb{R}^m$, то $h'(a)$ – матрица частных производных, а $h'(a) \cdot \xi$ – умножение матрицы размера $m \times k$ на k-мерный вектор.

Сформулируем следствие, которое будем далее использовать.

Утверждение 1.8 (Дельта-метод)**.** *Пусть X_1, X_2, … – последовательность независимых одинаково распределённых случайных (одномерных) величин с конечным вторым моментом. Пусть $EX_1 = a$, $DX_1 = \Sigma$, h – функция, дифференцируемая в точке a. Тогда в силу центральной предельной теоремы и стохастической формулы Тейлора получаем:*

$$\sqrt{n}(h(\overline{X}_n) - h(a)) \xrightarrow{d} h'(a) \cdot \mathcal{N}(0, \Sigma).$$

Замечание 1.6. Если X_1, X_2, … – k-мерные векторы, то DX_1 – ковариационная матрица. Тогда

$$\sqrt{n}(h(\overline{X}_n) - h(a)) \xrightarrow{d} h'(a) \cdot \mathcal{N}_k(0, \Sigma).$$

Задача 1.10. Пусть X_1, X_2, … – последовательность независимых одинаково распределённых случайных (одномерных) величин, $EX_1 = a$, $DX_1 = 1$. Найти такие числовые последовательности a_n и b_n, чтобы последовательность $a_n(\overline{X}_n^2 - b_n)$ имела невырожденный слабый предел (по распределению), и найти этот предел.

Решение:

Пусть $h(x) = x^2$, тогда $h'(a) = 2a$. Используя дельта-метод (утверждение (1.8)), получаем:

$$\sqrt{n}\left(\overline{X}_n^2 - a^2\right) \xrightarrow{d} 2a \cdot \mathcal{N}(0,1) \overset{d}{=} \mathcal{N}(0,\, 4a^2).$$

Если $a \neq 0$, то положим $a_n = \sqrt{n}$ и $b_n = a^2$. В этом случае искомый слабый предел равен $\mathcal{N}(0,\, 4a^2)$.

Если $a = 0$, то $\mathcal{N}(0,0)$ – вырожденный случай. По центральной предельной теореме (утверждение (1.5)) получаем:

$$\sqrt{n}\left(\overline{X}_n - a\right) = \sqrt{n} \cdot \overline{X} \xrightarrow{d} \mathcal{N}(0,1).$$

По теореме (1.1) о наследовании сходимости получаем:

$$n\overline{X}_n^2 \xrightarrow{d} \mathcal{N}^2(0,1) \overset{d}{=} \chi^2(1).$$

Таким образом, в случае $a = 0$ положим $a_n = n$ и $b_n = 0$. Тогда искомый слабый предел равен $\mathcal{N}^2(0,1)$. □

Семинар 2

Оценки параметров

Определение 2.1. *Выборка* – совокупность $X = (X_1, X_2, \ldots, X_n)$ независимых одинаково распределённых случайных величин. Количество n случайных величин в выборке называется *объёмом выборки.*

Определение 2.2. *Параметрическая статистика* – случай, когда для величин выборки известен класс распределения, но неизвестен параметр: $X_1 \sim F(x, \theta)$, где $\theta \in \Theta$ (параметр принадлежит известному параметрическому множеству).

Замечание 2.1. Для простоты будем сначала подразумевать, что $\Theta \subseteq \mathbb{R}^1$, но иногда будем рассматривать многомерные случаи.

Задача: оценить параметр θ или некоторую функцию от него по имеющимся наблюдениям X.

Определение 2.3. *Статистика* – некоторая функция от наблюдений для оценки параметра: $\hat{\theta}_n(X) = \varphi(X_1, \ldots, X_n)$.

Замечание 2.2. Статистика должна быть случайной величиной и не должна содержать неизвестный параметр, поэтому φ – борелевская функция, не зависящая от θ. Таким образом, статистика является измеримой функцией от наблюдений.

Определение 2.4. *Оценка параметра* θ – такая статистика для параметра θ, что $\hat{\theta}_n(X) \in \Theta$.

Замечание 2.3. Будем использовать слова «статистика» и «оценка» как синонимы.

Метод подстановки

Определение 2.5. $\hat{F}_n(x) \coloneqq \dfrac{1}{n}\sum_{i=1}^{n} I(X_i \leqslant x)$, где $x \in \mathbb{R}$, – *эмпирическая функция распределения*, где I – индикатор.

Замечание 2.4. Так как X_i – независимые одинаково распределённые случайные величины, то $I(X_i)$ – тоже независимые одинаково распределённые случайные величины. Значит, в силу усиленного закона больших чисел имеем:

$$\hat{F}_n(x) \overset{\text{п. н.}}{\to} EI(X_1 \leqslant x) = P(X_1 \leqslant x) = F_{X_1}(x).$$

Теорема 2.1 (Теорема Гливенко). *Пусть X_1, X_2, ... – независимые одинаково распределённые случайные величины, $X_1 \sim F(x)$. Тогда*

$$\sup_x |\hat{F}_n(x) - F(x)| \overset{\text{п. н.}}{\to} 0.$$

Метод подстановки заключается в следующем. Допустим, параметр $\theta(x) = G(F(x))$ – функционал от функции распределения. Тогда можно предложить оценку $\hat{\theta}_n(x) = G(\hat{F}_n(x))$. Она будет хорошо оценивать θ, если функционал G будет в некотором смысле непрерывен в точке F, так как функции $F(x)$ и $\hat{F}_n(x)$ близки в смысле теоремы Гливенко (2.1).

Определение 2.6. $\alpha_k \coloneqq EX_1^k$ – *начальный момент степени* k.

Определение 2.7. $\mu_k \coloneqq E(X_1 - EX_1)^k$ – *центральный момент степени* k.

Задача 2.1. Пусть $X_1 \sim F(x, \theta)$. Оценить α_1 и μ_2.

Решение:

$$\alpha_1 = \int\limits_{\mathbb{R}} x\, dF_\theta(x); \qquad \hat{\alpha}_1 = \int\limits_{\mathbb{R}} x\, d\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X};$$

$$\alpha_2 = \int\limits_{\mathbb{R}} x^2\, dF_\theta(x); \qquad \hat{\alpha}_2 = \int\limits_{\mathbb{R}} x^2\, d\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \overline{X^2};$$

$$\mu_2 = \alpha_2 - \alpha_1^2; \qquad \hat{\mu}_2 = \hat{\alpha}_2 - \hat{\alpha}_1^2 = \overline{X^2} - \overline{X}^2.$$

□

Свойства оценок

Будем оценивать не только сам параметр θ, но и некоторую функцию $\tau(\theta)$.

Определение 2.8. $\hat{\tau}_n(x)$ – *несмещённая оценка* для $\tau(\theta)$, если

$$E_\theta \hat{\tau}_n(x) = \tau(\theta) \qquad \forall \theta \in \Theta.$$

Определение 2.9. $\hat{\tau}_n(x)$ – *состоятельная оценка* для $\tau(\theta)$, если

$$\hat{\tau}_n(x) \xrightarrow{P_\theta} \tau(\theta) \qquad \forall \theta \in \Theta.$$

Определение 2.10. $\hat{\tau}_n(x)$ – *сильно состоятельная оценка* для $\tau(\theta)$, если

$$\hat{\tau}_n(x) \xrightarrow{\text{п. н.}\theta} \tau(\theta) \qquad \forall \theta \in \Theta.$$

Замечание 2.5. С прикладной точки зрения просто состоятельности вполне достаточно, поэтому к сильной состоятельности будем обращаться редко.

Определение 2.11. $\hat{\tau}_n(x)$ – $\sqrt{n}$-*асимптотически гауссовская (нормальная) оценка* для $\tau(\theta)$, если

$$\sqrt{n}(\hat{\tau}_n(x) - \tau(\theta)) \xrightarrow{d_\theta} \mathcal{N}(0,\, \sigma^2(\theta)) \qquad \forall \theta \in \Theta,$$

где $\sigma^2(\theta) > 0$.

Замечание 2.6. Индекс θ указывает, что распределение зависит от параметра θ. Можно его не писать, а только подразумевать.

Замечание 2.7. В отличие от состоятельности и асимптотической нормальности несмещённость определена для любого объёма выборки.

Утверждение 2.1. *Из $\sqrt{n}$-асимптотической нормальности следует состоятельность оценки.*

Доказательство:
В силу асимптотической нормальности имеем:

$$\sqrt{n}(\hat{\tau}_n(x) - \tau(\theta)) \xrightarrow{d} \mathcal{N}(0,\, \sigma^2(\theta)) \qquad \forall \theta \in \Theta.$$

Домножим эту последовательность на последовательность $\dfrac{1}{\sqrt{n}} \to 0$. В силу утверждения (1.6) получим:

$$\hat{\tau}_n(x) - \tau(\theta) \xrightarrow{d} 0.$$

Так как сходимость к константе по распределению и по вероятности эквивалентны, то $\hat{\tau}_n(x) - \tau(\theta) \xrightarrow{P} 0$. Добавим к обеим частям постоянную последовательность $\tau(\theta)$ (последовательности, сходящиеся по вероятности, можно складывать), получим: $\hat{\tau}_n(x) \xrightarrow{P} \tau(\theta)$. Таким образом, оценка состоятельна. ■

Утверждение 2.2. *Из несмещённости не следует состоятельность оценки.*

Доказательство:
Пусть $X_1, \ldots, X_n$ – выборка, $\theta = EX_1 \in \mathbb{R}$ – параметр. В качестве оценок можно взять $\overline{X}$ и X_1.

$$\begin{aligned} &E\overline{X} = E\frac{X_1 + \ldots + X_n}{n} = \frac{1}{n} \cdot n\theta = \theta; \\ &EX_1 = \theta \quad \text{по определению.} \end{aligned}$$

Таким образом, обе оценки несмещённые.

По закону больших чисел (утверждение (1.3)) $\overline{X} \xrightarrow{P} \theta \;\; \forall \theta \in \mathbb{R}$. Но $X_1 \to X_1$, то есть $X_1 \overset{P}{\not\to} \theta$. Значит, оценка $\overline{X}$ состоятельна, а оценка X_1 несостоятельна. ■

Утверждение 2.3. *Из состоятельности не следует несмещённость.*

Доказательство:
Пусть $X_1, \ldots, X_n$ – выборка, $\theta = DX_1 > 0$ – параметр. В качестве оценки возьмём выборочную дисперсию $\hat{\theta}_n = \hat{\mu}_2 = \overline{X^2} - \overline{X}^2$.

Как было показано при решении задачи (1.9), $\hat{\theta}_n \xrightarrow{P} \theta$. Значит, оценка $\hat{\theta}_n$ состоятельна.

МЕХАНИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ МГУ ИМЕНИ М.В. ЛОМОНОСОВА

teach-in
ЛЕКЦИИ УЧЕНЫХ МГУ

Теперь исследуем оценку на несмещённость. Удобнее работать с центрированными случайными величинами (у которых математическое ожидание равно 0). Для этого сделаем преобразования. Пусть $Y_i = X_i - EX_i$, тогда

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n}(X_i - EX_1) = \overline{X} - EX_i; \qquad E\overline{Y} = EY_i = 0;$$

$$\hat{\theta}_n = \overline{X^2} - \overline{X}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left((X_i - EX_i) - (\overline{X} - EX_i)\right)^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 = \overline{Y^2} - \overline{Y}^2;$$

$$E\hat{\theta}_n = E\overline{Y^2} - E\overline{Y}^2 = EY_1^2 - D\overline{Y} - (E\overline{Y})^2 = \theta - \frac{1}{n^2}\cdot n \cdot DY_1 = \theta - \frac{1}{n}\theta = \frac{n-1}{n}\theta.$$

Таким образом, $E\hat{\theta}_n \neq \theta$, значит, оценка смещённая. ■

Замечание 2.8. Несмещённой оценкой является

$$S_n^2 \coloneqq \frac{n}{n-1}\cdot\hat{\mu}_2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2.$$

Её тоже называют выборочной дисперсией. Далее будем обозначать S_n^2 именно несмещённую выборочную дисперсию.

Решение задач на свойства оценок

Задача 2.2. Исследовать выборочную дисперсию на $\sqrt{n}$-асимптотическую нормальность.

Решение:

Будем использовать представление выборочной дисперсии через центрированные случайные величины: $\hat{\theta}_n = \overline{Y^2} - \overline{Y}^2$.

1-й способ.

По центральной предельной теореме

$$\sqrt{n}\left(\begin{pmatrix}\overline{Y}\\ \overline{Y^2}\end{pmatrix} - \begin{pmatrix}0\\ \theta\end{pmatrix}\right) \xrightarrow{d} \mathcal{N}_2\left(\begin{pmatrix}0\\ 0\end{pmatrix}, D\begin{pmatrix}Y_1\\ Y_1^2\end{pmatrix}\right),$$

где

$$D\begin{pmatrix}Y_1\\ Y_1^2\end{pmatrix} = \begin{pmatrix}DY_1 & \mathrm{Cov}(Y_1, Y_1^2)\\ \mathrm{Cov}(Y_1, Y_1^2) & D(Y_1^2)\end{pmatrix} = \begin{pmatrix}\mu_2 & \mu_3\\ \mu_3 & \mu_4 - \mu_2^2\end{pmatrix},$$

где μ_2, μ_3, μ_4 – центральные моменты величины X_1, причём $\theta = \mu_2$.

Используя дельта-метод (утверждение (1.8)) для векторов $\begin{pmatrix}Y_i\\ Y_i^2\end{pmatrix}$, получаем:

$$\hat{\theta}_n = h(\overline{Y}, \overline{Y^2}) = \overline{Y^2} - \overline{Y}^2; \qquad h(x, y) = y - x^2; \qquad h(0, \theta) = \theta;$$
$$\nabla h = (-2x,\, 1); \qquad \nabla h(0, \theta) = (0, 1);$$
$$\sqrt{n}\left(h(\overline{Y}, \overline{Y^2}) - h(0, \theta)\right) = \sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix},$$

где $\begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} \stackrel{d}{=} \mathcal{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix},\, \begin{pmatrix} \mu_2 & \mu_3 \\ \mu_3 & \mu_4 - \mu_2^2 \end{pmatrix}\right)$. Тогда получаем:

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} \eta_2 \sim \mathcal{N}(0,\, \mu_4 - \mu_2^2).$$

2-й способ.

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) = \sqrt{n}\left(\overline{Y^2} - \theta\right) - \sqrt{n}\,\overline{Y}^2. \tag{2.1}$$

По центральной предельной теореме (утверждение (1.5)) получаем:

$$\sqrt{n}\left(\overline{Y^2} - \theta\right) \xrightarrow{d} \mathcal{N}(0,\, D(Y_1^2)), \quad \text{где} \quad D(Y_1^2) = \mu_4 - \mu_2^2;$$
$$\sqrt{n}\,\overline{Y} \xrightarrow{d} \mathcal{N}(0,\, DY_1); \qquad \left(\sqrt{n}\,\overline{Y}\right)^2 \xrightarrow{d} \mathcal{N}^2(0,\, DY_1).$$

По лемме Слуцкого (1.1) получаем:

$$\frac{1}{\sqrt{n}}\left(\sqrt{n}\,\overline{Y}\right)^2 \xrightarrow{d} 0 \cdot \mathcal{N}^2(0,\, DY_1); \qquad \sqrt{n}\,\overline{Y}^2 \xrightarrow{d} 0.$$

Тогда по лемме Слуцкого (1.1) для выражения (2.1) получаем:

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} \mathcal{N}(0,\, \mu_4 - \mu_2^2).$$

Таким образом, выборочная дисперсия является $\sqrt{n}$-асимптотически нормальной оценкой с асимптотической дисперсией $\mu_4 - \mu_2^2$. □

Задача 2.3. Пусть $X_1, \ldots, X_n$ – выборка, $X_1 \sim \mathcal{U}(0, \theta)$, где $\theta > 0$. В качестве оценки можно взять $\hat{\theta}_n = \max(X_1, \ldots, X_n) = X_{(n)}$ – n-ю порядковую статистику (подробнее на следующем семинаре). Исследовать свойства этой оценки.

Решение:

$$F_{X_{(n)}}(x) = P(X_{(n)} \leqslant x) = P(X_1 \leqslant x,\ \ldots,\ X_n \leqslant x) =$$
$$= P(X_1 \leqslant x) \cdot \ldots \cdot P(X_n \leqslant x) = F_{X_1}^n(x) = \begin{cases} 0, & x \leqslant 0 \\ \left(\dfrac{x}{\theta}\right)^n, & x \in (0, \theta) \\ 1, & x \geqslant \theta \end{cases};$$
$$p_{X_{(n)}}(x) = \frac{n}{\theta}\left(\frac{x}{\theta}\right)^{n-1} I(x \in (0, \theta));$$

$$EX_{(n)} = \int\limits_{\mathbb{R}} x \cdot p_{X_{(n)}}(x)\,dx = \int\limits_0^\theta x \cdot \frac{n}{\theta}\left(\frac{x}{\theta}\right)^{n-1} dx = \frac{n}{\theta^n} \cdot \frac{x^{n+1}}{n+1}\bigg|_0^\theta = \frac{n}{n+1}\theta.$$

Таким образом, оценка $\hat{\theta}_n = X_{(n)}$ является смещённой для параметра θ. Теперь легко видеть, что оценка $\tilde{\theta}_n = \dfrac{n+1}{n}X_{(n)}$ является несмещённой для параметра θ.

$$EX^2_{(n)} = \int\limits_{\mathbb{R}} x^2 \cdot p_{X_{(n)}}(x)\,dx = \int\limits_0^\theta x^2 \cdot \frac{n}{\theta}\left(\frac{x}{\theta}\right)^{n-1} dx = \frac{n}{\theta^n} \cdot \frac{x^{n+2}}{n+2}\bigg|_0^\theta = \frac{n}{n+2}\theta^2;$$

$$DX_{(n)} = EX^2_{(n)} - E^2X_{(n)} = \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2 = \frac{n}{(n+2)(n+1)^2}\theta^2.$$

Используя неравенство Маркова, получим:

$$\forall \varepsilon > 0 \qquad P\left(\left|\sqrt{n}(X_{(n)} - \theta)\right| > \varepsilon\right) \leqslant \frac{n \cdot E(X_{(n)} - \theta)^2}{\varepsilon^2} =$$

$$= \frac{n}{\varepsilon^2} \cdot E\left((X_{(n)} - EX_{(n)}) - (\theta - EX_{(n)})\right)^2 = \frac{n}{\varepsilon^2}\left(DX_{(n)} + (EX_{(n)} - \theta)^2\right) \textcircled{$\rightarrow$}$$

$$nDX_{(n)} = \frac{n^2}{(n+2)(n+1)^2}\theta^2 \underset{n\to+\infty}{\to} 0;$$

$$n(EX_{(n)} - \theta)^2 = n\left(\frac{n}{n+1} - 1\right)^2\theta^2 = \frac{n}{(n+1)^2}\theta^2 \underset{n\to+\infty}{\to} 0;$$

$$\textcircled{$\rightarrow$}\ 0.$$

Таким образом, $\sqrt{n}(X_{(n)} - \theta) \xrightarrow{P} 0$ и $\sqrt{n}(X_{(n)} - \theta) \xrightarrow{d} 0$. Значит, оценка $\hat{\theta}_n = X_{(n)}$ не является $\sqrt{n}$-асимптотически гауссовской.

Домножая $\sqrt{n}(X_{(n)} - \theta) \xrightarrow{P} 0$ на $\dfrac{1}{\sqrt{n}} \to 0$, получаем $X_{(n)} \xrightarrow{P} \theta$. Значит, оценка $\hat{\theta}_n = X_{(n)}$ является состоятельной. □

Замечание 2.9. Можно показать, что $-n(X_{(n)} - \theta) \xrightarrow{d} E\left(\dfrac{1}{\theta}\right)$.

Задача 2.4. Пусть X_1 – выборка объёма 1, $X_1 \sim \mathrm{Bin}(1, \theta)$ (распределение Бернулли), где $\theta \in (0, 1)$. Показать, что не существует несмещённых оценок для $\tau(\theta) = \dfrac{1}{\theta}$.

Решение:

Предположим, что существует несмещённая оценка $\varphi(X_1)$. Тогда

$$
\begin{aligned}
&E_\theta \varphi(X_1) = \frac{1}{\theta} \qquad \forall \theta \in (0,1);\\
&E_\theta \varphi(X_1) = \varphi(1)\cdot\theta + \varphi(0)\cdot(1-\theta);\\
&\varphi(1)\theta^2 + \varphi(0)\theta - \varphi(0)\theta^2 - 1 = 0;\\
&(\varphi(1) - \varphi(0))\theta^2 + \varphi(0)\theta - 1 = 0 \qquad \forall \theta \in (0,1).
\end{aligned}
$$

Но левая часть последнего равенства может принимать нулевое значение не более чем в двух точках. Получили противоречие. Значит, не существует несмещённых оценок для $\tau(\theta) = \dfrac{1}{\theta}$. $\square$

Задача 2.5. Пусть X_1 – выборка объёма 1, X_1 – число успехов до 1-го неуспеха в схеме Бернулли с вероятностью успеха θ в отдельном испытании, где $\theta \in (0,1)$. Найти несмещённую оценку для параметра θ.

Решение:

$$P_\theta(X_1 = k) = \theta^k(1-\theta), \quad \text{где} \quad k = 0,1,2,\ldots.$$

Пусть $\varphi(X_1)$ – несмещённая оценка θ. Тогда

$$
\begin{aligned}
&E_\theta \varphi(X_1) = \theta \qquad \forall \theta \in (0,1);\\
&E_\theta \varphi(X_1) = \sum_{k=0}^{+\infty} \varphi(k)\cdot\theta^k(1-\theta);\\
&\sum_{k=0}^{+\infty} \varphi(k)\theta^k = \frac{\theta}{1-\theta}.
\end{aligned}
$$

Используя формулу суммы ряда геометрической прогрессии $\dfrac{1}{1-\theta} = \sum\limits_{k=0}^{+\infty} \theta^k$, получим:

$$
\begin{aligned}
&\sum_{k=0}^{+\infty} \varphi(k)\theta^k = \sum_{k=0}^{+\infty} \theta^{k+1}; \qquad \sum_{k=0}^{+\infty} \varphi(k)\theta^k = \sum_{k=1}^{+\infty} \theta^k \qquad \forall \theta \in (0,1);\\
&\varphi(0) = 0, \qquad \varphi(k) = 1 \quad \forall k = 1,2,\ldots.
\end{aligned}
$$

Таким образом, $\varphi(X_1) = I(X_1 \geqslant 1) \notin \Theta = (0,1)$. $\square$

Замечание 2.10. В этой задаче хотя несмещённая оценка и существует, но она бессмысленна, так как её значения не попадают в множество значений параметра.

Семинар 3

Метод моментов

Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim F_\theta(x)$, где параметр $\theta \in \Theta \subseteq \mathbb{R}^k$.

Введём функции $g_i \ \ \forall i = \overline{1,k}$ такие, что $E_\theta g_i(X_1) =: m_i(\theta)$ – конечная величина.

Пусть $\overline{g_i(X)} := \dfrac{1}{n}\sum\limits_{j=1}^{n} g_i(X_j)$ – среднее значение функции g_i по выборке X.

Теперь мы можем записать следующие k уравнений: $\begin{cases} m_1(\theta) = \overline{g_1(X)} \\ \vdots \\ m_k(\theta) = \overline{g_k(X)} \end{cases}$.

Определение 3.1. Пусть $m(\theta) := \begin{pmatrix} m_1(\theta) \\ \vdots \\ m_k(\theta) \end{pmatrix}$. Если система выше для функции $m(\theta)$ имеет единственное решение, то это решение будем называть *оценкой параметра θ по методу моментов*.

Замечание 3.1. Стандартный выбор функций в методе моментов: $g_i(x) = x^i$.

Решение задач на метод моментов

Задача 3.1. Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim \mathrm{Exp}(\theta)$, где параметр $\theta > 0$. С помощью метода моментов оценить θ и изучить свойства этой оценки.

Решение:

1-й способ.

Пусть $g(x) = x$ – пробная функция. Тогда получаем:

$$E_\theta X_1 = \frac{1}{\theta}; \qquad \overline{X} = \frac{1}{\theta}; \qquad \hat{\theta} = \frac{1}{\overline{X}}.$$

По закону больших чисел (утверждение (1.3)) $\overline{X} \xrightarrow{P} \dfrac{1}{\theta}$. При $\theta > 0$ в силу непрерывности преобразования $\dfrac{1}{\overline{X}}$ по теореме (1.1) о наследовании сходимости получаем, что $\hat{\theta} = \dfrac{1}{\overline{X}} \xrightarrow{P} \theta$. Таким образом, оценка состоятельна.

2-й способ.

Пусть $g(x) = x^2$ – пробная функция. Тогда получаем:

$$E_\theta X_1^2 = D_\theta X_1 + E_\theta^2 X_1 = \frac{1}{\theta^2} + \left(\frac{1}{\theta}\right)^2 = \frac{2}{\theta^2}; \qquad \overline{X^2} = \frac{2}{\theta^2}; \qquad \hat{\theta} = \sqrt{\frac{2}{\overline{X^2}}}.$$

$\overline{X^2} \xrightarrow{P} \dfrac{2}{\theta^2}$. При $\theta > 0$ в силу непрерывности преобразования $\sqrt{\dfrac{2}{\overline{X^2}}}$ по теореме (1.1) о наследовании сходимости получаем, что $\hat{\theta} = \sqrt{\dfrac{2}{\overline{X^2}}} \xrightarrow{P} \theta$. Таким образом, оценка состоятельна. □

Замечание 3.2. Можно показать, что для $g(x) = x^k$, где $k \in \mathbb{N}$, получим:

$$E_\theta X_1^k = \frac{k!}{\theta^k}; \qquad \hat{\theta} = \sqrt[k]{\frac{k!}{\overline{X^k}}}.$$

По дельта-методу (утверждение (1.8)) можно показать, что оценка $\hat{\theta}$ является $\sqrt{n}$-асимптотически нормальной.

Задача 3.2. Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim \mathcal{U}(\theta_1, \theta_2)$, где $\theta_1 < \theta_2$. С помощью метода моментов оценить параметр $\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}$ и изучить свойства этой оценки.

Решение:

$$\begin{cases} E_\theta X_1 = \overline{X} \\ E_\theta X_1^2 = \overline{X^2} \end{cases} \iff \begin{cases} E_\theta X_1 = \overline{X} \\ E_\theta X_1^2 - E_\theta^2 X_1 = \overline{X^2} - \overline{X}^2 \end{cases} \iff \begin{cases} E_\theta X_1 = \overline{X} \\ D_\theta X_1 = \hat{\mu}_2 \end{cases};$$

$$\begin{cases} \dfrac{\theta_1 + \theta_2}{2} = \overline{X} \\ \dfrac{(\theta_2 - \theta_1)^2}{12} = \hat{\mu}_2 \end{cases}; \qquad \begin{cases} \dfrac{\theta_1 + \theta_2}{2} = \overline{X} \\ \dfrac{\theta_2 - \theta_1}{2} = \sqrt{3\hat{\mu}_2} \end{cases}; \qquad \begin{cases} \hat{\theta}_1 = \overline{X} - \sqrt{3\hat{\mu}_2} \\ \hat{\theta}_2 = \overline{X} + \sqrt{3\hat{\mu}_2} \end{cases}.$$

Так как $\overline{X}$ и $\hat{\mu}_2$ – состоятельные оценки, то в силу непрерывности преобразований оценка параметра $\hat{\theta}$ состоятельна.

Покажем, что оценка $\hat{\theta}_1$ является $\sqrt{n}$-асимптотически нормальной.

$$\hat{\theta}_1 = \overline{X} - \sqrt{3\hat{\mu}_2} = h\left(\overline{X}, \overline{X^2}\right), \quad \text{где} \quad h(x,y) = x - \sqrt{3(y - x^2)}.$$

Будем использовать многомерный дельта-метод (утверждение (1.8)), переходя к центрированным наблюдениям.

$$Y_i \coloneqq X_i - EX_i; \qquad \overline{X} = \overline{Y} + \frac{\theta_1 + \theta_2}{2}; \qquad \hat{\mu}_2 = \overline{Y^2} - \overline{Y}^2;$$

$$\sqrt{n}(\hat{\theta}_1 - \theta_1) = \sqrt{n}\left(\overline{Y} + \frac{\theta_1 + \theta_2}{2} - \sqrt{3\left(\overline{Y^2} - \overline{Y}^2\right)} - \theta_1\right) =$$

$$= \sqrt{n}\left(\overline{Y} - \sqrt{3\left(\overline{Y^2} - \overline{Y}^2\right)} + \frac{\theta_2 - \theta_1}{2}\right) \ominus$$

По центральной предельной теореме (утверждение (1.5)) получаем:

$$\sqrt{n}\left(\begin{pmatrix}\overline{Y}\\ \overline{Y^2}\end{pmatrix} - \begin{pmatrix}0\\ DX_1\end{pmatrix}\right) \xrightarrow{d} \eta \sim \mathcal{N}_2\left(0,\, D\begin{pmatrix}Y_1\\ Y_1^2\end{pmatrix}\right);$$

$$h(0,\, DX_1) = -\sqrt{3DX_1} = -\frac{\theta_2-\theta_1}{2};$$

$$= \sqrt{n}\left(h(\overline{Y},\, \overline{Y^2}) - h(0,\, DX_1)\right) \xrightarrow{d} \nabla h\big|_{(0,\, DX_1)} \cdot \eta =$$

$$\eta = \begin{pmatrix}\eta_1\\ \eta_2\end{pmatrix} \sim \mathcal{N}_2\left(0,\, D\begin{pmatrix}Y_1\\ Y_1^2\end{pmatrix}\right) = \mathcal{N}_2\left(0,\, \begin{pmatrix}DY_1 & 0\\ 0 & DY_1^2\end{pmatrix}\right),$$

значит, гауссовские величины η_1 и η_2 независимы;

$$\nabla h\big|_{(0,\, DX_1)} = \left(1,\ -\frac{\sqrt{3}}{2\sqrt{DX_1}}\right) = \left(1,\ -\frac{3}{\theta_2-\theta_1}\right);$$

$$= \eta_1 - \frac{3}{\theta_2-\theta_1}\eta_2 \sim \mathcal{N}\left(0,\, DY_1 + \frac{9}{(\theta_2-\theta_1)^2}\cdot DY_1^2\right);$$

$$DY_1 = DX_1 = \frac{(\theta_2-\theta_1)^2}{12};$$

$$DY_1^2 = EY_1^4 - E^2Y_1^2 = EY_1^4 - D^2X_1 =$$

$$EY_1^4 = \int\limits_{-\frac{\theta_2-\theta_1}{2}}^{\frac{\theta_2-\theta_1}{2}} \frac{y^4}{\theta_2-\theta_1}\,dy = \frac{2}{\theta_2-\theta_1}\cdot\frac{y^5}{5}\bigg|_0^{\frac{\theta_2-\theta_1}{2}} = \frac{(\theta_2-\theta_1)^4}{80};$$

$$= \frac{(\theta_2-\theta_1)^4}{80} - \frac{(\theta_2-\theta_1)^4}{144};$$

$$DY_1 + \frac{9}{(\theta_2-\theta_1)^2}\cdot DY_1^2 = \left(\frac{1}{12} + 9\left(\frac{1}{80} - \frac{1}{144}\right)\right)(\theta_2-\theta_1)^2 = \frac{2}{15}(\theta_2-\theta_1)^2.$$

$\square$

Замечание 3.3. Оценка может быть не совсем адекватной. Рассмотрим в качестве примера выборку $(0,0,0,0,1)$.

$$\overline{X} = \frac{1}{5};\qquad \overline{X^2} = \frac{1}{5};\qquad \hat{\mu}_2 = \frac{4}{25};\qquad \sqrt{\hat{\mu}_2} = \frac{2}{5};\qquad \hat{\theta}_2 = \frac{1}{5} + \sqrt{3}\cdot\frac{2}{5} < 1.$$

Как видно, оценка правого конца интервала лежит левее значения 1, но это значение было получено при наблюдении, а значит, истинный правый конец интервала должен лежать правее значения 1.

Задача 3.3. Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim \mathrm{Pareto}(\theta, 1)$, то есть плотность вероятности X_1 имеет вид $p_\theta(x) = \dfrac{\theta}{x^{\theta+1}} I(x > 1)$, где параметр $\theta > 0$. С помощью метода моментов оценить параметр θ и изучить свойства этой оценки.

Решение:

Возьмём пробную функцию $g(x) = x$.

При $\theta > 1$ имеем:

$$E_\theta X_1 = \int\limits_1^{+\infty} x \cdot \frac{\theta}{x^{\theta+1}}\,dx = \int\limits_1^{+\infty} \frac{\theta}{x^{\theta}}\,dx = \theta \cdot \left.\frac{x^{-\theta+1}}{-\theta+1}\right|_1^{+\infty} = \frac{\theta}{\theta-1}.$$

При $\theta \in (0,1]$ получаем $E_\theta X_1 = +\infty$. Таким образом, мы не сможем найти оценку для всех возможных значений θ, используя пробную функцию $g(x) = x$.

Возьмём пробную функцию $g(x) = \ln x$.

$$E_\theta \ln X_1 = \int\limits_1^{+\infty} \ln x \cdot \frac{\theta}{x^{\theta+1}}\,dx = \int\limits_1^{+\infty} \ln x \cdot \left(-\frac{1}{x^{\theta}}\right)' dx =$$

$$= \ln x \cdot \left.\left(-\frac{1}{x^{\theta}}\right)\right|_1^{+\infty} + \int\limits_1^{+\infty} \frac{dx}{x^{\theta+1}} = 0 + \left.\frac{x^{-\theta}}{-\theta}\right|_1^{+\infty} = \frac{1}{\theta} \qquad \forall \theta > 0.$$

Тогда по методу моментов получаем:

$$\overline{\ln X} = \frac{1}{\theta}; \qquad \hat{\theta} = \frac{1}{\overline{\ln X}}.$$

Аналогично решениям предыдущих задач можно показать, что эта оценка будет состоятельной, $\sqrt{n}$-асимптотически гауссовской. □

Метод выборочных квантилей

Определение 3.2. Пусть $F(x)$ – функция распределения, $p \in (0,1)$ – число. Тогда $\zeta_p := F^{-1}(p)$ – *квантиль уровня* p.

Замечание 3.4. Так как $F(x)$ является монотонно неубывающей функцией, то проблема с нахождением значения обратной функции может возникнуть только в двух случаях:

1) Когда $F(x)$ разрывна и значение p не принадлежит множеству значений $F(x)$. В этом случае p обязательно находится в некоторой области значений, которые функция $F(x)$ не принимает из-за разрыва в некоторой конкретной точке x^*. Тогда $\zeta_p := x^*$.

2) Когда $F(x)$ постоянна на некотором множестве значений x и равна p. Тогда квантилью уровня p можно считать любую точку из этого множества значений x. Мы для определённости будем выбирать самую левую точку из множества значений x.

С учётом замечания можем дать более точное определение.

Определение 3.3. Пусть $F(x)$ – функция распределения, $p \in (0,1)$ – число. Тогда $\zeta_p := \inf\{z\colon\ F(z) \geqslant p\}$ – *квантиль уровня* p.

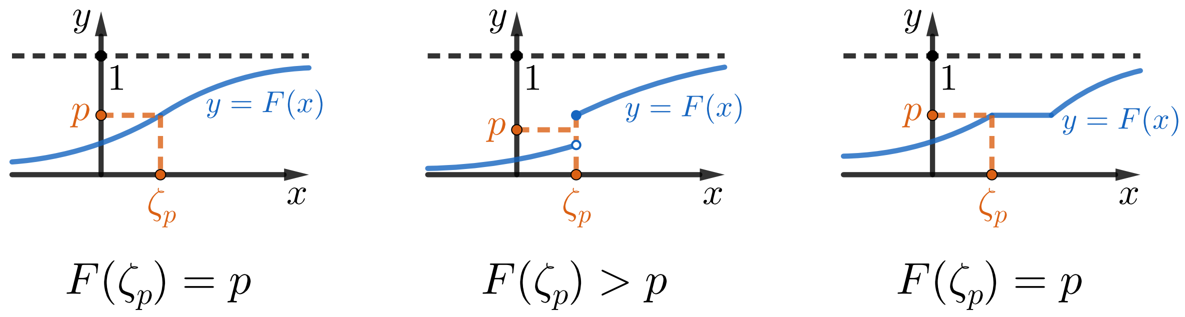

Рис. 3.1: Квантили уровня p для различных функций распределения $F(x)$

Определение 3.4. $Z_{n,p} \coloneqq \inf\{z\colon\ F_n(z) \geqslant p\}$ – *выборочная квантиль уровня* p.

Пусть $X = (X_1, \ldots, X_n)$ – выборка. Тогда $\forall x \in \mathbb{R}$ $\ F_n(x) \coloneqq \dfrac{1}{n}\sum\limits_{i=1}^{n} I(X_i \leqslant x)$ – эмпирическая функция распределения.

Определение 3.5. Пусть $x_i \coloneqq X_i(\omega)$ – реализация случайных величин в эксперименте при элементарном исходе ω. Упорядочим числа x_i (в упорядоченном наборе будем индекс указывать в скобках): $x_{(1)} \leqslant x_{(2)} \leqslant \ldots \leqslant x_{(n)}$. Тогда *$k$-й порядковой статистикой* назовём $X_{(k)}(\omega) = x_{(k)}$.

Например, $X_{(1)} = \min\{X_1, \ldots, X_n\}$, $\quad X_{(n)} = \max\{X_1, \ldots, X_n\}$.

Определение 3.6. $X_{(1)}, \ldots, X_{(n)}$ – *вариационный ряд.*

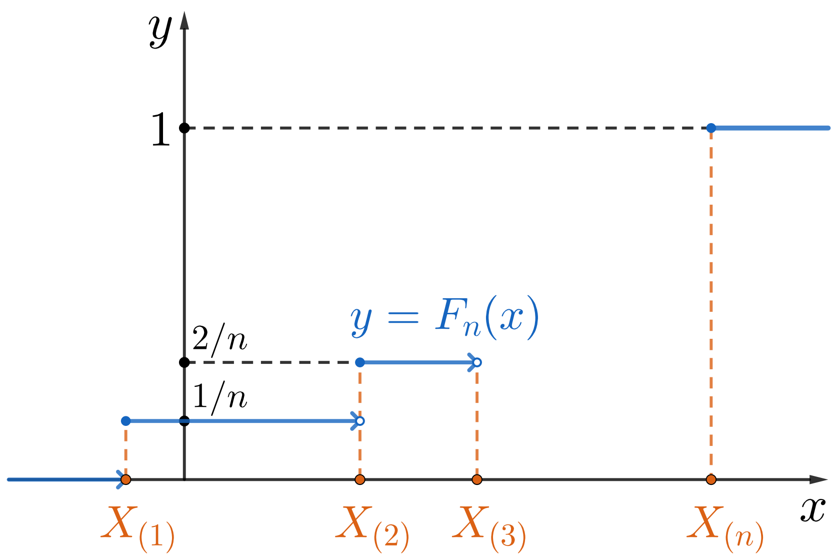

Рис. 3.2: Эмпирическая функция распределения для вариационного ряда

Из графика видно, что выборочную квантиль уровня p можно записать следующим образом:

$$Z_{n,p} = \begin{cases} X_{(pn)}, & pn \in \mathbb{Z} \\ X_{(\lfloor pn \rfloor + 1)}, & pn \notin \mathbb{Z} \end{cases} = X_{(\lceil pn \rceil)}.$$

Определение 3.7. *Медиана* – квантиль уровня $\dfrac{1}{2}$.

Определение 3.8. Пусть $X_1, \ldots, X_n$ – выборка. *Выборочной медианой* называется

$$\hat{m} \coloneqq \begin{cases} X_{(k)}, & n = 2k-1 \\ \dfrac{X_{(k)} + X_{(k+1)}}{2}, & n = 2k \end{cases}$$

где $k \in \mathbb{N}$.

Теорема 3.1. *Пусть $X_1, \ldots, X_n$ – выборка, $p \in (0,1)$. Пусть $f(\zeta_p) \coloneqq F'(\zeta_p) > 0$. Тогда*

$$\sqrt{n}(Z_{n,p} - \zeta_p) \xrightarrow{d} \mathcal{N}\left(0, \frac{p(1-p)}{f^2(\zeta_0)}\right).$$

Замечание 3.5. Так как из $\sqrt{n}$-асимптотической нормальности следует состоятельность, то в условиях теоремы выше выборочная квантиль $Z_{n,p}$ уровня p является состоятельной оценкой для теоретической квантили ζ_p уровня p.

Замечание 3.6. Теорема (3.1) верна, если заменить $Z_{n,p}$ на $X_{(k)}$, где $k = np + O(1)$ (это следует понимать в том смысле, что $k = np$, если $np \in \mathbb{Z}$, и $k = \lceil np \rceil$, если $np \notin \mathbb{Z}$).

Определение 3.9. $\zeta_{1/4}$ и $\zeta_{3/4}$ – соответственно *нижняя* и *верхняя квартили.*

Решение задач на метод выборочных квантилей

Задача 3.4. Пусть $X_1, \ldots, X_n$ – выборка, $X_1 \sim \text{Cauchy}(0, \theta)$, где параметр $\theta > 0$, то есть плотность вероятности X_1 имеет вид $p_\theta(x) = \dfrac{\theta}{\pi(x^2+\theta^2)}$, а плотность распределения – $F_\theta(x) = \dfrac{1}{\pi}\operatorname{arctg}\left(\dfrac{x}{\theta}\right) + \dfrac{1}{2}$. С помощью метода выборочных квантилей оценить θ.

Решение:

Найдём нижнюю квартиль:

$$F_\theta(x) = \frac{1}{4}; \qquad \operatorname{arctg}\left(\frac{x}{\theta}\right) = -\frac{\pi}{4}; \qquad \zeta_{1/4} = -\theta.$$

Найдём верхнюю квартиль:

$$F_\theta(x) = \frac{3}{4}; \qquad \operatorname{arctg}\left(\frac{x}{\theta}\right) = \frac{\pi}{4}; \qquad \zeta_{3/4} = \theta.$$

Таким образом, в качестве оценки θ можно взять $Z_{n,3/4}$, или $-Z_{n,1/4}$, или $\dfrac{1}{2}\left(Z_{n,3/4} - Z_{n,1/4}\right)$. Такие оценки будут состоятельными и $\sqrt{n}$-асимптотически нормальными. □

Семинар 4

Сравнение оценок. Равномерный подход

Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim F_\theta(x)$, где $\theta \in \Theta$. Пусть $\hat{\theta}_n(X)$ – оценка параметра θ.

Определение 4.1. $R(\hat{\theta}_n, \theta) = E_\theta(\hat{\theta}_n - \theta)^2$ – *среднеквадратичный риск.*

Рассмотрим некоторую постоянную оценку $\tilde{\theta}_n = \theta_0 \in \Theta$. Пусть θ_n^* – оптимальная оценка в том смысле, что при всех значениях $\theta \in \Theta$ её функция риска $R(\theta_n^*, \theta)$ меньше, чем для любой другой оценки. Тогда, в частности, необходимо выполнение неравенства $E_{\theta_0}(\theta_n^* - \theta_0)^2 \leqslant E_{\theta_0}(\tilde{\theta}_n - \theta_0)^2 = 0$, то есть $E_{\theta_0}(\theta_n^* - \theta_0)^2 = 0$. В силу произвольности выбора $\theta_0 \in \Theta$ получаем: $E_\theta(\theta_n^* - \theta)^2 = 0 \ \ \forall \theta \in \Theta$. Тогда $\theta_n^* \overset{P_\theta\text{-п. н.}}{=} \theta$. Таким образом, оценка θ_n^* является вырожденной ситуацией.

Задача 4.1. Пусть $X = (X_1)$ – выборка, $X_1 \sim \mathcal{U}(\theta,\ \theta + 1)$, где параметр $\theta \in \mathbb{Z}$. Оценить θ.

Решение:

Так как X_1 может лежать только в промежутке длины 1 с целыми границами, то естественно в качестве оценки взять $\hat{\theta} \coloneqq \lfloor X_1 \rfloor$. С P_θ-вероятностью 1 получаем: $\lfloor X_1 \rfloor = \theta$, то есть $\theta_n^* \overset{P_\theta\text{-п. н.}}{=} \theta$. □

Сформулируем утверждения, которые будем использовать при решении следующей задачи.

Утверждение 4.1. *Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim \mathcal{N}(\theta_1,\ \theta_2^2)$, где параметры $\theta_1 \in \mathbb{R}$, $\theta_2 > 0$. $S^2 \coloneqq \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ – несмещённый вариант выборочной дисперсии. Тогда*

$$\frac{(n-1)S^2}{\theta_2^2} \overset{d}{=} \chi^2(n-1) \overset{d}{=} \xi_1^2 + \ldots + \xi_{n-1}^2,$$

где $\{\xi_i\}$ – независимые одинаково распределённые стандартные гауссовские величины (то есть $\xi_i \sim \mathcal{N}(0,1)$).

Утверждение 4.2. *Пусть $\xi \sim \mathcal{N}(0, \sigma^2)$. Тогда*

$$E\xi^{2n-1} = 0, \qquad E\xi^{2n} = (2n-1)!! \cdot \sigma^{2n} \qquad \forall n \in \mathbb{N}.$$

Задача 4.2. Пусть $X = (X_1, \ldots, X_n)$ – выборка, $n > 1$, $X_1 \sim \mathcal{N}(\theta_1,\ \theta_2^2)$, где параметры $\theta_1 \in \mathbb{R}$, $\theta_2 > 0$. Запишем несмещённый вариант выборочной дисперсии:

$$S^2 \coloneqq \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2.$$

Рассмотрим для дисперсии следущий класс оценок: $\mathcal{K} = \{\lambda S^2\}$, где $\lambda > 0$. Найти наилучшую оценку θ_2^2 в классе $\mathcal{K}$ в среднеквадратичном подходе.

Решение:

Учитывая, что S^2 – несмещённая оценка θ_2^2, получаем:

$$R(\lambda S^2,\, \theta_2^2) = E_\theta(\lambda S^2 - \theta_2^2)^2 = E_\theta \left(\lambda(S^2 - \theta_2^2) + (\lambda-1)\theta_2^2\right)^2 = \lambda^2 DS^2 + (\lambda-1)^2\theta_2^4 \ominus$$

Используя утверждения (4.1) и (4.2), получаем:

$$\begin{aligned}
&D\frac{(n-1)S^2}{\theta_2^2} = \frac{(n-1)^2}{\theta_2^4}DS^2 = D\chi^2(n-1) = (n-1)D\xi_1^2 = \\
&= (n-1)\left(E\xi_1^4 - E^2\xi_1^2\right) = (n-1)(3-1) = 2(n-1);\\
&DS^2 = \frac{2\theta_2^4}{n-1};\\
&\ominus \left(\frac{2\lambda^2}{n-1} + (\lambda-1)^2\right)\theta_2^4.
\end{aligned}$$

Минимизируем $R(\lambda S^2,\, \theta_2^2)$ по λ при $\lambda > 0$. Относительно λ это выражение является параболой с ветвями вверх, значит, наименьшее значение достигается в вершине параболы: $\lambda^* = \dfrac{2}{2\left(\dfrac{2}{n-1}+1\right)} = \dfrac{n-1}{n+1}$. Тогда наилучшей оценкой в среднеквадратичном подходе будет

$$S^{*2} = \lambda^* S^2 = \frac{1}{n+1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2.$$

Эта оценка будет состоятельной, $\sqrt{n}$-асимптотически нормальной и смещённой, но её смещение будет стремиться к 0 при $n \to +\infty$. □

Сравнение в классе несмещённых оценок с конечной дисперсией

Пусть $\hat{\tau}_n(X)$ – несмещённая оценка $\tau(\theta)$ с конечной дисперсией. Тогда $E_\theta\hat{\tau}_n(X) = \tau(\theta)$, значит, получаем:

$$R(\hat{\tau}_n(X), \tau(\theta)) = E_\theta(\hat{\tau}_n(X) - \tau(\theta))^2 = E_\theta(\hat{\tau}_n(X) - E_\theta\hat{\tau}_n(X))^2 = D\hat{\tau}_n(X).$$

Определение 4.2. Несмещённая оценка $\tau_n^*(X)$ для $\tau(\theta)$ с конечной дисперсией называется *оптимальной*, если $D_\theta\tau_n^*(X) \leqslant D_\theta\hat{\tau}_n(X)\ \ \forall\theta \in \Theta$ для любой другой несмещённой оценки $\hat{\tau}_n(X)$ для $\tau(\theta)$.

Утверждение 4.3. *Пусть $\tau_1^*(X)$ и $\tau_2^*(X)$ – две оптимальные оценки для $\tau(\theta)$ (индексы 1 и 2 указывают номер оценки, объём выборки n для краткости не пишем). Тогда $\tau_1^* \overset{P_\theta\text{-п. н.}}{=} \tau_2^*$.*

МЕХАНИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ МГУ ИМЕНИ М.В. ЛОМОНОСОВА
teach-in
ЛЕКЦИИ УЧЕНЫХ МГУ

Доказательство:

Так как оценки оптимальные, то $E_\theta \tau_1^* = E_\theta \tau_2^* = \tau(\theta)$, а также

$$D_\theta \tau_1^* \leqslant D_\theta \tau_2^* \quad \text{и} \quad D_\theta \tau_2^* \leqslant D_\theta \tau_1^* \qquad \forall \theta \in \Theta.$$

Значит, $D_\theta \tau_1^* = D_\theta \tau_2^* =: \sigma^2(\theta)$ – наименьшая дисперсия в классе несмещённых оценок с конечной дисперсией.

Рассмотрим оценку $\hat{\tau} = \dfrac{\tau_1^* + \tau_2^*}{2}$. Для неё получаем:

$$E_\theta \hat{\tau} = \frac{1}{2}(E_\theta \tau_1^* + E_\theta \tau_2^*) = \tau(\theta) \qquad \forall \theta \in \Theta.$$

По определению оптимальной оценки получаем:

$$\sigma^2(\theta) \leqslant D_\theta \hat{\tau} = \frac{1}{4}\left(2\sigma^2(\theta) + 2\operatorname{Cov}(\tau_1^*, \tau_2^*)\right) \leqslant$$

По неравенству Коши-Буняковского получаем:

$$|\operatorname{Cov}(\tau_1^*, \tau_2^*)| \leqslant \sqrt{D\tau_1^* \cdot D\tau_2^*} = \sigma^2(\theta);$$
$$\leqslant \frac{1}{4}\left(2\sigma^2(\theta) + 2\sigma^2(\theta)\right) = \sigma^2(\theta) \qquad \forall \theta \in \Theta.$$

Таким образом, $D_\theta \hat{\tau} = \sigma^2(\theta)$. Значит, оценка $\hat{\tau}$ тоже является оптимальной. Причём в неравенстве Коши-Буняковского достигается равенство. Значит, величины $\tau_1^* - \tau(\theta)$ и $\tau_2^* - \tau(\theta)$ почти наверное пропорциональны:

$$\tau_1^* - \tau(\theta) \overset{P_\theta\text{-п. н.}}{=} k(\theta)\left(\tau_2^* - \tau(\theta)\right).$$

Тогда для дисперсии $\sigma^2(\theta)$ получаем:

$$\sigma^2(\theta) = \operatorname{Cov}(\tau_1^*, \tau_2^*) = E((\tau_1^* - \tau(\theta))(\tau_2^* - \tau(\theta)) =$$
$$= E\left(k(\theta)\left(\tau_2^* - \tau(\theta)\right)^2\right) = k(\theta) \cdot D\tau_2^* = k(\theta) \cdot \sigma^2(\theta) \qquad \forall \theta \in \Theta.$$

Значит, $k(\theta) \equiv 1$. Тогда

$$\tau_1^* - \tau(\theta) \overset{P_\theta\text{-п. н.}}{=} \tau_2^* - \tau(\theta); \qquad \tau_1^* \overset{P_\theta\text{-п. н.}}{=} \tau_2^*.$$

■

Задача 4.3. Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim \mathcal{U}(0, \theta)$, где $\theta > 0$. Из метода моментов можно взять такую оценку: $\hat{\theta}_1 = 2\overline{X}$ (это несмещённая и состоятельная оценка). Также можно взять следующую оценку: $\hat{\theta}_2 = \dfrac{n+1}{n} X_{(n)}$ (нормировочный множитель присутствует, чтобы оценка была несмещённой, также оценка является состоятельной). Индексы 1 и 2 указывают номер оценки, объём выборки n для краткости не пишем. Выяснить, какая из оценок $\hat{\theta}_1$, $\hat{\theta}_2$ лучше.

Решение:

$$\begin{aligned}
&D_\theta\hat{\theta}_1 = 4D_\theta\overline{X} = 4\cdot\frac{n\cdot D_\theta X_1}{n^2} = \frac{4}{n}\cdot\frac{\theta^2}{12} = \frac{\theta^2}{3n};\\
&D_\theta\hat{\theta}_2 = \frac{\theta^2}{n(n+2)};\\
&\frac{\theta^2}{3n} \ \vee \ \frac{\theta^2}{n(n+2)}; \qquad n+2 \ \vee \ 3.
\end{aligned}$$

При $n = 1$ оценки совпадают. При $n > 1$ имеем: $n+2 > 3$, значит, $D_\theta\hat{\theta}_1 > D_\theta\hat{\theta}_2$. Таким образом, оценка $\hat{\theta}_2$ лучше оценки $\hat{\theta}_1$. □

Семинар 5

Неравенство Рао-Крамера и эффективные оценки

Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim P_\theta$, где параметр $\theta \in \Theta \subseteq \mathbb{R}^1$.

Определение 5.1. Пусть у наблюдения X есть плотность $p(x, \theta)$ относительно некоторой меры μ. *Условия регулярности (R)*:

1) носитель распределение $\mathcal{N}_p \coloneqq \{p(x, \theta) > 0\}$ не зависит от неизвестного параметра θ;

2) Θ – открытый интервал, и $\exists \dfrac{\partial}{\partial\theta} \ln p(x, \theta) \quad \forall x \in \mathcal{N}_p \quad \forall \theta \in \Theta$;

$\ln p(X, \theta)$ – *правдоподобие*, $\mathcal{U}(X, \theta) \coloneqq \dfrac{\partial}{\partial\theta} \ln p(X, \theta)$ – *вклад наблюдения* X;

3) для всех статистик $S(X)$, таких что $E_\theta S^2(X) < M < +\infty \quad \forall \theta \in \Theta$, можно менять местами интегрирование и дифференцирование в выражении $\dfrac{\partial}{\partial\theta} E_\theta S(X) \quad \forall \theta \in \Theta$;

4) $I_X(\theta) \coloneqq E_\theta \mathcal{U}^2(X, \theta)$ – *информация Фишера, содержащаяся в выборке* X; требуется выполнение неравенств $0 < I_X(\theta) < +\infty$.

Теорема 5.1 (Неравенство Рао-Крамера)**.** *Пусть выполнены условия регулярности (R) (определение (5.1)), $\hat{\tau}_n(X)$ – несмещённая оценка для $\tau(\theta)$ с равномерно по θ ограниченным 2-м моментом. Тогда $D_\theta \hat{\tau}_n(X) \geqslant \dfrac{(\tau'(\theta))^2}{I_X(\theta)} \quad \forall \theta \in \Theta$.*

Определение 5.2. Если для оценки $\hat{\tau}_n(X)$ в неравенстве Рао-Крамера (теорема (5.1)) достигается равенство, то такая оценка называется *эффективной.*

Определение 5.3. Пусть статистика $S(X)$ имеет плотность $g_\theta(s)$. Тогда

$$I_{S(X)}(\theta) \coloneqq E_\theta \left(\frac{\partial}{\partial\theta} \ln g_\theta(S(x)) \right)^2$$

– *информация Фишера, содержащаяся в статистике* $S(X)$.

Утверждение 5.1. *Пусть $X = (X_1, \ldots, X_n)$ – выборка, тогда*

$$I_X(\theta) = n I_{X_1}(\theta).$$

Замечание 5.1. Далее для краткости будем использовать обозначение $i(\theta) \coloneqq I_{X_1}(\theta)$.

Утверждение 5.2. *Пусть* $X = (X_1, \ldots, X_n)$ *– выборка,* $X_1 \sim \mathrm{Exp}(\theta)$*, тогда* $\sum\limits_{i=1}^{n} X_i \sim \mathrm{Gamma}(\theta, n)$*, то есть плотность вероятности суммы имеет вид*

$$\theta e^{-\theta x}\frac{(\theta x)^{n-1}}{\Gamma(n)} I(x > 0).$$

Утверждение 5.3. *Если* $X_1 \sim p_{X_1}(x)$*, то* $\dfrac{X_1}{n} \sim p_{X_1/n}(x) = |n| p_{X_1}(nx) \ \ \forall n \in \mathbb{R}\backslash\{0\}$.

Задача 5.1. Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim \mathrm{Exp}(\theta)$, то есть $p_\theta(x) = \theta e^{-\theta x} I(x > 0)$, где параметр $\theta > 0$. Проверить, что $\overline{X}$ – эффективная оценка для $\tau(\theta) = E_\theta X_1 = \dfrac{1}{\theta}$, и вычислить $I_{\overline{X}}(\theta)$.

Решение:

$$\begin{aligned}
&D_\theta \overline{X} = \frac{n \cdot D_\theta X_1}{n^2} = \frac{D_\theta X_1}{n} = \frac{1}{n\theta^2};\\
&\tau'(\theta) = -\frac{1}{\theta^2};\\
&\mathcal{U}(X_1, \theta) = \frac{\partial}{\partial\theta} \ln p_\theta(X_1) = \frac{\partial}{\partial\theta}(\ln\theta - \theta X_1) = \frac{1}{\theta} - X_1;\\
&i(\theta) = E_\theta\left(\frac{1}{\theta} - X_1\right)^2 = DX_1 = \frac{1}{\theta^2};\\
&\frac{(\tau'(\theta))^2}{ni(\theta)} = \frac{\theta^2}{n\theta^4} = \frac{1}{n\theta^2} = D_\theta\overline{X}.
\end{aligned}$$

Таким образом, в неравенстве Рао-Крамера (теорема (5.1)) достигается равенство, значит, $\overline{X}$ – эффективная оценка для $E_\theta X_1$.

Так как X_1 имеет плотность вероятности $p_\theta(x) = \theta e^{-\theta x} I(x > 0)$, то согласно утверждению (5.3) $\dfrac{X_1}{n}$ имеет плотность вероятности $n\theta e^{-n\theta x} I(nx > 0)$, где $n \in \mathbb{N}$. Тогда по утверждению (5.2) получаем, что $\overline{X} \sim \mathrm{Gamma}(n\theta, n)$, то есть плотность вероятности имеет вид $p_{\overline{X}}(x, \theta) = (n\theta)^n e^{-n\theta x}\dfrac{x^{n-1}}{(n-1)!}$.

$$\begin{aligned}
&\mathcal{U}(\overline{X}, \theta) = \frac{\partial}{\partial\theta} \ln p_{\overline{X}}(\overline{X}, \theta) = \frac{\partial}{\partial\theta} \ln\left((n\theta)^n e^{-n\theta\overline{X}}\frac{\overline{X}^{n-1}}{(n-1)!}\right) =\\
&= \frac{\partial}{\partial\theta}(n\ln\theta - n\theta\overline{X}) = \frac{n}{\theta} - n\overline{X};\\
&I_{\overline{X}}(\theta) = E_\theta\left(n^2\left(\frac{1}{\theta} - \overline{X}\right)^2\right) = n^2 \cdot D\overline{X} = n^2 \cdot \frac{1}{n\theta^2} = \frac{n}{\theta^2}.
\end{aligned}$$

□

Замечание 5.2. В задаче выше получилось, что $I_{\overline{X}}(\theta) = n \cdot i(\theta) = I_X(\theta)$. Такие статистики называются *достаточными*. Вообще говоря, информация Фишера, содержащаяся в какой-либо статистике, всегда не больше информации Фишера, содержащейся во всём наблюдении.

Критерий эффективности

Лемма 5.1 (Критерий эффективности). *Пусть выполнены условия регулярности (R) (определение (5.1)). Тогда $\hat{\tau}_n(X)$ – эффективная оценка для $\tau(\theta)$ тогда и только тогда, когда*

$$\hat{\tau}_n(X) - \tau(\theta) \overset{P_\theta\text{-п. н.}}{=} \frac{\tau'(\theta)}{I_X(\theta)}\mathcal{U}(X,\theta) \qquad \forall \theta \in \Theta.$$

Задача 5.2. Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim \mathcal{N}(a, \theta)$, где a – известное число, параметр $\theta > 0$. Какие функции $\tau(\theta)$ можно оценить эффективно? Найти для них эффективные оценки и $i(\theta)$.

Решение:

$$p_\theta(x) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{(x-a)^2}{2\theta}}; \qquad p(X,\theta) = \prod_{i=1}^{n} p_\theta(X_i); \qquad \ln p(X,\theta) = \sum_{i=1}^{n} \ln p_\theta(X_i);$$

$$\mathcal{U}(X,\theta) = \frac{\partial}{\partial\theta}\ln p(X,\theta) = \sum_{i=1}^{n}\frac{\partial}{\partial\theta}\ln p_\theta(X_i) = \sum_{i=1}^{n}\frac{\partial}{\partial\theta}\left(-\frac{\ln\theta}{2} - \frac{(X_i-a)^2}{2\theta}\right) =$$
$$= \sum_{i=1}^{n}\left(-\frac{1}{2\theta} + \frac{(X_i-a)^2}{2\theta^2}\right) = \frac{n}{2\theta^2}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-a)^2 - \theta\right). \tag{5.1}$$

Так как по критерию эффективности (лемма (5.1))

$$\mathcal{U}(X,\theta) \overset{P_\theta\text{-п. н.}}{=} \frac{I_X(\theta)}{\tau'(\theta)}\left(\hat{\tau}_n(X) - \tau(\theta)\right) \qquad \forall\theta\in\Theta, \tag{5.2}$$

то $\hat{\tau}_n(X) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i-a)^2$ – эффективная оценка для $\tau(\theta) = \theta$. Также эффективно можно оценить эту функцию после домножения её на константу и прибавления к ней константы.

Также из выражений (5.1) и (5.2) видно, что $\dfrac{n\cdot i(\theta)}{\tau'(\theta)} = \dfrac{n}{2\theta^2}$. Учитывая, что $\tau'(\theta) = 1$, получаем: $i(\theta) = \dfrac{1}{2\theta^2}$. □

Замечание 5.3. Вообще говоря, эффективной оценки может не существовать.

Задача 5.3. Пусть $X = (X_1, \ldots, X_n)$ – выборка с дискретным распределением: $P_\theta(X_1 = 0) = \theta$, $P_\theta(X_1 = 2) = 2\theta$, $P_\theta(X_1 = 1) = 1 - 3\theta$, где $\theta \in \left(0, \dfrac{1}{3}\right)$. Найти все эффективные оценки.

Решение:

$$
\begin{aligned}
&p(x,\theta) = P_\theta(X_1 = x), \qquad x = \{0,1,2\};\\
&p(x,\theta) = \theta^{I(x=0)} \cdot (2\theta)^{I(x=2)} \cdot (1-3\theta)^{I(x=1)};\\
&\frac{\partial}{\partial\theta}\ln p(x,\theta) = \frac{\partial}{\partial\theta}\left(I(x=0)\ln\theta + I(x=2)\ln(2\theta) + I(x=1)\ln(1-3\theta)\right) =\\
&= I(x\neq 1)\cdot\frac{1}{\theta} + I(x=1)\cdot\frac{-3}{1-3\theta} = I(x\neq 1)\cdot\left(\frac{1}{\theta}+\frac{3}{1-3\theta}\right) - \frac{3}{1-3\theta} =\\
&= \frac{3}{\theta(1-3\theta)}\left(\frac{I(x\neq 1)}{3} - \theta\right);\\
&\mathcal{U}(X,\theta) = \frac{3n}{\theta(1-3\theta)}\left(\frac{1}{3n}\sum_{i=1}^{n} I(X_i\neq 1) - \theta\right).
\end{aligned}
$$

Так как по критерию эффективности (лемма (5.1))

$$\mathcal{U}(X,\theta) \overset{P_\theta\text{-п. н.}}{=} \frac{I_X(\theta)}{\tau'(\theta)}\left(\hat{\tau}_n(X) - \tau(\theta)\right) \qquad \forall\theta\in\Theta,$$

то $\hat{\tau}_n(X) = \dfrac{1}{3n}\displaystyle\sum_{i=1}^{n} I(X_i\neq 1)$ – эффективная оценка для $\tau(\theta)=\theta$. Также эффективно можно оценить эту функцию после домножения её на константу и прибавления к ней константы. Значит, все эффективные оценки можно записать так:

$$\hat{\tau}_n(X) = \frac{a}{3n}\sum_{i=1}^{n} I(X_i\neq 1) + b \qquad \text{для} \qquad \tau(\theta) = a\theta + b.$$

□

Утверждение 5.4. *Если $\hat{\tau}_n(X)$ – эффективная оценка для $\tau(\theta)$, то*

$$p(X,\theta) \overset{P_\theta\text{-п. н.}}{=} h(x)e^{A(\theta)\hat{\tau}_n(X)+B(\theta)},$$

причём $\tau(\theta) = -\dfrac{B'(\theta)}{A'(\theta)}$.

Определение 5.4. Такой вид распределения, как в утверждении (5.4), называется *распределением экспоненциального типа.*

Проверка условий регулярности (R)

Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim \mathrm{Exp}(\theta)$, где $\theta > 0$.

Проверим условия регулярности (R) (определение (5.1)):

1) $\mathcal{N}_p = {R_+}^n$;

2) $\Theta = (0;\ +\infty)$ и $\exists \dfrac{\partial}{\partial\theta} \ln p(x, \theta)$ – это условие автоматически проверяется в ходе вычислений;

3) в частности, для статистики $S(X) \equiv 1$ получаем:

$$0 = \frac{\partial}{\partial\theta} \int\limits_{\mathcal{N}_p} S(x) p(x, \theta) \mu(dx) = \int\limits_{\mathcal{N}_p} \frac{\partial}{\partial\theta} p(x, \theta) \mu(dx) =$$

$$= \int\limits_{\mathcal{N}_p} \frac{\partial}{\partial\theta} \left(\ln p(x, \theta)\right) p(x, \theta) \mu(dx) = E_\theta \mathcal{U}(X, \theta);$$

тогда $I_X(\theta) = E_\theta \mathcal{U}^2(X, \theta) = D_\theta \mathcal{U}(X, \theta)$;

Утверждение 5.5. *Для выполнения условия (3) достаточно выполнения двух условий:*

1) $\sqrt{p(\theta)}$ непрерывно дифференцируем по θ;

2) $I_X(\theta)$ непрерывна по θ.

4) $0 < I_X(\theta) = \dfrac{n}{\theta^2} < +\infty$.

Таким образом, для распределения экспоненциального типа условия регулярности (R) выполнены.

Свойства эффективных оценок

По определению эффективная оценка является несмещённой.

Задача 5.4. Пусть $\hat{\theta}_n(X)$ – эффективная оценка для θ. Доказать, что оценка $\hat{\theta}_n(X)$ является $\sqrt{n}$-асимптотически гауссовской, и найти её асимптотическую дисперсию.

Решение:

По критерию эффективности (лемма (5.1)) получаем:

$$\hat{\theta}_n(X) - \theta = \frac{1}{n \cdot i(\theta)} \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln p(X_i, \theta);$$

$$\sqrt{n} \left(\hat{\theta}_n(X) - \theta\right) = \frac{\sqrt{n}}{n \cdot i(\theta)} \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln p(X_i, \theta) = \frac{\sqrt{n}}{i(\theta)} \left(\frac{1}{n} \sum_{i=1}^{n} \mathcal{U}(X_i, \theta)\right).$$

Так как $\{X_i\}$ – выборка, то $\{\mathcal{U}(X_i, \theta)\}$ – независимые одинаково распределённые величины. Так как выполнены условия регулярности (R), то

$$E_\theta \mathcal{U}(X_1, \theta) = 0, \qquad 0 < D_\theta \mathcal{U}(X_1, \theta) = E_\theta \mathcal{U}^2(X_1, \theta) = i(\theta) < +\infty.$$

Тогда по центральной предельной теореме (утверждение (1.5)) получаем:

$$\sqrt{n} \left(\hat{\theta}_n(X) - \theta\right) \xrightarrow{d} \frac{1}{i(\theta)} \mathcal{N}(0,\, i(\theta)).$$

По лемме Слуцкого (1.1) получаем:

$$\sqrt{n}\left(\hat{\theta}_n(X) - \theta\right) \xrightarrow{d} \mathcal{N}\left(0,\, \frac{1}{i(\theta)}\right) \qquad \forall\theta \in \Theta.$$

□

Определение 5.5. Пусть $\hat{\theta}_n(X)$ является $\sqrt{n}$-асимптотически гауссовской, то есть

$$\sqrt{n}(\hat{\theta}_n(X) - \theta) \xrightarrow{d} \mathcal{N}(0,\, \sigma^2(\theta)) \qquad \forall\theta \in \Theta,$$

причём $\sigma^2(\theta) = \dfrac{1}{i(\theta)}$ почти всюду при $\theta \in \Theta$. Тогда такая оценка $\hat{\theta}_n(X)$ называется *асимптотически эффективной.*

Задача 5.5. Пусть $X = (X_1,\, \ldots, X_n)$ – выборка, $X_1 \sim \mathrm{Exp}(\theta)$, где параметр $\theta > 0$. Показать, что θ можно оценить асимптотически эффективно.

Решение:

$\overline{X}$ – эффективная оценка для $\tau(\theta) = \dfrac{1}{\theta}$. По закону больших чисел (утверждение (1.3)) имеем: $\overline{X} \xrightarrow{P} EX_1 = \dfrac{1}{\theta}$. Тогда по теореме о непрерывности получаем: $\dfrac{1}{\overline{X}} \xrightarrow{P} \theta$.

По центральной предельной теореме (утверждение (1.5)) получаем:

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

Применим дельта-метод (утверждение (1.8)) для функции $h(x) = \dfrac{1}{x}$:

$$h'(x) = -\frac{1}{x^2}; \qquad h'\left(\frac{1}{\theta}\right) = -\theta^2;$$
$$\sqrt{n}\left(\frac{1}{\overline{X}} - \theta\right) = \sqrt{n}\left(h(\overline{X}) - h\left(\frac{1}{\theta}\right)\right) \xrightarrow{d} \mathcal{N}\left(0,\, \frac{1}{\theta^2}\cdot\left(\theta^2\right)^2\right) =$$
$$= \mathcal{N}(0, \theta^2) = \mathcal{N}\left(0,\, \frac{1}{i(\theta)}\right).$$

Таким образом, $\dfrac{1}{\overline{X}}$ является асимптотически эффективной оценкой для θ. □

Семинар 6

Многомерные оптимальные оценки

Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim P_\theta$, где параметр $\theta \in \Theta \subseteq \mathbb{R}^k$.

Определение 6.1. $\tau_n^*(X)$ – *оптимальная оценка для* $\tau(\theta)$, где $\tau\colon\ \mathbb{R}^k \to \mathbb{R}^d$, если:

1) $E_\theta \tau_n^*(X) = \tau(\theta) \quad \forall \theta \in \Theta$;

2) $D_\theta \tau_n^*(X) \leqslant D_\theta \tilde{\tau}_n(X) \quad \forall \theta \in \Theta$ для любых $\tilde{\tau}_n(X)$, являющихся несмещёнными оценками для $\tau(\theta)$.

Здесь $D_\theta \tau_n^* \coloneqq E_\theta(\tau_n^* - E_\theta \tau_n^*)(\tau_n^* - E_\theta \tau_n^*)^T$ – ковариационная матрица размера $d \times d$. Она является симметричной и неотрицательно определённой ($D_\theta \tau_n^* \geqslant 0$, смотри ниже).

Определение 6.2. Симметричная матрица D размера $d \times d$ называется *неотрицательно определённой* ($D \geqslant 0$) тогда и только тогда, когда $\lambda^T D \lambda \geqslant 0 \ \ \forall \lambda \in \mathbb{R}^d$.

Замечание 6.1. У симметричной матрицы все собственные числа вещественны. Неотрицательная определённость такой матрицы равносильна тому, что все её собственные числа неотрицательны.

Определение 6.3.

$$D_1 \leqslant D_2 \iff D_2 - D_1 \geqslant 0 \iff \lambda^T D_1 \lambda \leqslant \lambda^T D_2 \lambda \quad \forall \lambda \in \mathbb{R}^d.$$

Замечание 6.2. Так как

$$\begin{aligned}
&\lambda^T D_\theta \tau_n^* \lambda = E_\theta \left(\lambda^T (\tau_n^* - E_\theta \tau_n^*)(\tau_n^* - E_\theta \tau_n^*)^T \lambda\right) = \\
&= E_\theta \left(\lambda^T (\tau_n^* - E_\theta \tau_n^*) \left(\lambda^T (\tau_n^* - E_\theta \tau_n^*)\right)^T\right) = E_\theta \left(\lambda^T (\tau_n^* - E_\theta \tau_n^*) \lambda^T (\tau_n^* - E_\theta \tau_n^*)\right) = \\
&= E_\theta \left(\lambda^T (\tau_n^* - E_\theta \tau_n^*)\right)^2 = E_\theta \left(\lambda^T \tau_n^* - E_\theta \lambda^T \tau_n^*\right)^2 = D_\theta(\lambda^T \tau_n^*),
\end{aligned}$$

то оценка τ_n^* оптимальна тогда и только тогда, когда всевозможные линейные комбинации её компонент $\lambda^T \tau_n^*$ будут оптимально оценивать $\lambda^T \tau(\theta)$.

В частности, если взять вектор λ, у которого на i-м месте стоит 1, а на всех остальных 0, то получим, что каждая компонента оптимальной оценки τ_n^* должна оптимально оценивать соответствующую компоненту векторной функции $\tau(\theta)$.

Верно и обратное. Если мы можем оптимально оценить все компоненты $\tau(\theta)$, то объединение этих оптимальных оценок будет оптимальной оценкой $\tau(\theta)$. Покажем это ниже.

Задача 6.1. Пусть $\tau_1^*(X)$ и $\tau_2^*(X)$ – оптимальные оценки для $\tau_1(\theta)$ и $\tau_2(\theta)$ соответственно, где $\tau_1:\ \mathbb{R}^k \to \mathbb{R}$ и $\tau_2:\ \mathbb{R}^k \to \mathbb{R}$. Доказать, что $\lambda_1 \tau_1^* + \lambda_2 \tau_2^*$ оптимально оценивает $\lambda_1 \tau_1 + \lambda_2 \tau_2$, где $\lambda_1, \lambda_2 \in \mathbb{R}$.

Решение:

1) Докажем, что если $\varphi(X)$ – статистика, причём $E_\theta\varphi(X) = 0$, то $\operatorname{Cov}(\varphi(X),\, \tau_i^*(X)) = 0$, где $i = 1, 2$.

$$\begin{aligned}
&E_\theta(\tau_i^* + \lambda\varphi) = \tau_i \qquad \forall\lambda \in \mathbb{R};\\
&D_\theta\tau_i^* \leqslant D_\theta(\tau_i^* + \lambda\varphi) = D_\theta\tau_i^* + 2\lambda\operatorname{Cov}(\varphi, \tau_i^*) + \lambda^2 D_\theta\varphi;\\
&\lambda^2 D_\theta\varphi + 2\lambda\operatorname{Cov}(\varphi, \tau_i^*) \geqslant 0 \qquad \forall\lambda \in \mathbb{R}.
\end{aligned} \tag{6.1}$$

Если $D_\theta\varphi = 0$, то $\varphi \overset{P_\theta\text{-п. н.}}{=} 0$, так как $E_\theta\varphi = 0$. В этом случае $\operatorname{Cov}(\varphi,\, \tau_i^*) = 0$. Поэтому далее будем рассматривать случай $D_\theta\varphi > 0$. Тогда левая часть неравенства (6.1) является параболой с ветвями вверх, причём её корнями являются 0 и $\dfrac{-2\operatorname{Cov}(\varphi, \tau_i^*)}{D_\theta\varphi}$. Значит, неравенство (6.1) будет выполнено $\forall\lambda \in \mathbb{R}$ тогда и только тогда, когда корни совпадают, то есть $\operatorname{Cov}(\varphi, \tau_i^*) = 0$.

2) Пусть T – несмещённая оценка для $\tau = \lambda_1\tau_1 + \lambda_2\tau_2$. Пусть $\varphi = \tau^* - T$, где $\tau^* = \lambda_1\tau_1^* + \lambda_2\tau_2^*$. Так как $E_\theta\varphi = 0$, то в силу пункта (1) доказательства получаем:

$$\operatorname{Cov}(\varphi, \tau^*) = \lambda_1\operatorname{Cov}(\varphi, \tau_1^*) + \lambda_2\operatorname{Cov}(\varphi, \tau_2^*) = 0.$$

С другой стороны, так как $\varphi = \tau^* - T$, то

$$\operatorname{Cov}(\varphi, \tau^*) = D_\theta\tau^* - \operatorname{Cov}(T, \tau^*).$$

Значит, $D_\theta\tau^* = \operatorname{Cov}(T, \tau^*)$. По неравенству Коши-Буняковского получаем:

$$D_\theta\tau^* = \operatorname{Cov}(T, \tau^*) \leqslant \sqrt{D_\theta T \cdot D_\theta\tau^*}.$$

Значит, $D_\theta\tau^* \leqslant D_\theta T \quad \forall\theta \in \Theta$. Таким образом, τ^* – оптимальная оценка для τ. $\square$

Многомерное неравенство Рао-Крамера

Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim P_\theta$, где параметр $\theta \in \Theta \subseteq \mathbb{R}^k$. Пусть X_1 имеет плотность $p(x, \theta)$ относительно некоторой меры μ.

Определение 6.4. $\mathcal{U}(X, \theta) \coloneqq \left(\dfrac{\partial}{\partial\theta_1}\ln p(X, \theta),\ \ldots,\ \dfrac{\partial}{\partial\theta_k}\ln p(X, \theta)\right)^T$ – *вклад выборки.*

Определение 6.5. $I_X(\theta) \coloneqq E_\theta\mathcal{U}(X, \theta)\mathcal{U}^T(X, \theta)$ – *информация Фишера, содержащаяся в выборке X.*

Замечание 6.3. В условиях регулярности (R) получаем: $I_X(\theta) = D_\theta\mathcal{U}(X, \theta)$, так как $E_\theta\mathcal{U}(X, \theta) = 0$.

Определение 6.6. Пусть $\tau(\theta)\colon\ \mathbb{R}^k \to \mathbb{R}^d$, тогда $\tau'(\theta) = \begin{pmatrix}\nabla_\theta\tau_1(\theta)\\ \vdots\\ \nabla_\theta\tau_d(\theta)\end{pmatrix}$.

Условия регулярности (R) аналогичны одномерному случаю.

Теорема 6.1 (Многомерное неравенство Рао-Крамера). *Пусть $\hat{\tau}_n(X)$ – несмещённая оценка для $\tau(\theta)$ с равномерно по θ ограниченным 2-м моментом. Тогда в условиях регулярности (R) имеем:*

$$D_\theta \hat{\tau}_n(X) \geqslant \tau'(\theta) I_X^{-1}(\theta) \tau'^T(\theta) \qquad \forall \theta \in \Theta.$$

Определение 6.7. Если для оценки $\hat{\tau}_n(X)$ в многомерном неравенстве Рао-Крамера (теорема (6.1)) достигается равенство, то такая оценка называется *эффективной*.

Задача 6.2. Пусть $X = (X_1, \ldots, X_n)$ – выборка, где $n \geqslant 2$, $X_1 \sim \mathcal{N}(\theta_1, \theta_2)$, где параметры $\theta_1 \in \mathbb{R}$, $\theta_2 > 0$. Как известно, $\left(\overline{X}, S^2\right)^T$ является несмещённой оценкой для $\theta = (\theta_1, \theta_2)^T$, где

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2.$$

Является ли эта оценка эффективной?

Решение:

$$p(x, \theta) = \frac{1}{\sqrt{2\pi\theta_2}} e^{-\frac{(x-\theta_1)^2}{2\theta_2}}; \qquad \ln p(x, \theta) = -\frac{1}{2} \ln \theta_2 - \frac{(x-\theta_1)^2}{2\theta_2} - \frac{1}{2} \ln(2\pi);$$

$$\mathcal{U}(X_1, \theta) = \nabla_\theta \ln p(X_1, \theta) = \left(\frac{X_1 - \theta_1}{\theta_2}, \; -\frac{1}{2\theta_2} + \frac{(X_1 - \theta_1)^2}{2\theta_2^2} \right)^T;$$

$$I_{X_1}(\theta) = E_\theta \mathcal{U}(X_1, \theta) \mathcal{U}^T(X_1, \theta) := \begin{pmatrix} i_{11} & i_{12} \\ i_{21} & i_{22} \end{pmatrix} \ominus$$

$$i_{11} = \frac{1}{\theta_2^2} D_\theta (X_1 - \theta_1) = \frac{1}{\theta_2};$$

$$i_{22} = \frac{1}{4\theta_2^4} D_\theta (X_1 - \theta_1)^2 = \frac{1}{4\theta_2^4} \left(E_\theta (X_1 - \theta_1)^4 - \theta_2^2 \right) = \frac{1}{2\theta_2^2};$$

$$i_{12} = i_{21} = -\frac{1}{2\theta_2^2} E_\theta (X_1 - \theta_1) + \frac{1}{2\theta_2^3} E_\theta (X_1 - \theta_1)^3 = 0;$$

$$\ominus \begin{pmatrix} \frac{1}{\theta_2} & 0 \\ 0 & \frac{1}{2\theta_2^2} \end{pmatrix}; \qquad I_X(\theta) = n I_{X_1}(\theta); \qquad I_X^{-1}(\theta) = \begin{pmatrix} \frac{\theta_2}{n} & 0 \\ 0 & \frac{2\theta_2^2}{n} \end{pmatrix};$$

$$\tau(\theta) = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}; \qquad \tau'(\theta) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}; \qquad \tau'(\theta) I_X^{-1}(\theta) \tau'^T(\theta) = \begin{pmatrix} \frac{\theta_2}{n} & 0 \\ 0 & \frac{2\theta_2^2}{n} \end{pmatrix}.$$

Учитывая, что

$$\frac{(n-1)S^2}{\theta_2} \overset{d}{=} \chi^2(n-1) \overset{d}{=} \xi_1^2 + \ldots + \xi_{n-1}^2$$

(утверждение (4.1)), а также то, что $\overline{X} \sim \mathcal{N}\left(\theta_1, \dfrac{\theta_2}{n}\right)$ и величины $\overline{X}$ и S^2 независимы, получаем:

$$D_\theta S^2 = D_\theta \left(\frac{\theta_2}{n-1}\chi^2(n-1)\right) = \frac{\theta_2^2}{(n-1)^2}\cdot(n-1)D(\xi_1^2) = \frac{2\theta_2^2}{n-1};$$
$$D_\theta \begin{pmatrix} \overline{X} \\ S^2 \end{pmatrix} = \begin{pmatrix} \dfrac{\theta_2}{n} & 0 \\ 0 & \dfrac{2\theta_2^2}{n-1} \end{pmatrix}.$$

По неравенству Рао-Крамера (теорема (6.1)) получаем:

$$\begin{pmatrix} \dfrac{\theta_2}{n} & 0 \\ 0 & \dfrac{2\theta_2^2}{n-1} \end{pmatrix} \geqslant \begin{pmatrix} \dfrac{\theta_2}{n} & 0 \\ 0 & \dfrac{2\theta_2^2}{n} \end{pmatrix}.$$

Так как $\dfrac{2\theta_2^2}{n-1} > \dfrac{2\theta_2^2}{n}$ при $n \geqslant 2$, то в неравенстве Рао-Крамера реализуется случай строгого неравенства. Значит, $\left(\overline{X}, S^2\right)^T$ является неэффективной оценкой для $\theta = (\theta_1, \theta_2)^T$.

Для $\tau_1(\theta) = \theta_1$ по неравенству Рао-Крамера получаем: $D_\theta \overline{X} \geqslant \dfrac{\theta_2}{n}$. В этом случае реализуется равенство, значит, $\overline{X}$ является эффективной оценкой для θ_1.

Для $\tau_2(\theta) = \theta_2$ по неравенству Рао-Крамера получаем: $D_\theta S^2 \geqslant \dfrac{2\theta_2^2}{n}$. В этом случае реализуется строгое неравенство, значит, S^2 является неэффективной оценкой для θ_2. □

Замечание 6.4. Можно показать, что S^2 является оптимальной оценкой для θ_2 в среднеквадратичном смысле, а эффективных оценок для θ_2 нет.

Семинар 7

Оценки максимального правдоподобия (ОМП)

Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim P_\theta$, где параметр $\theta \in \Theta$. Пусть X_1 имеет плотность вероятности $p(x, \theta)$ относительно некоторой меры μ.

Определение 7.1. $L_n(X, \theta) \coloneqq \prod\limits_{i=1}^{n} p(X_i, \theta)$ – *правдоподобие выборки* X.

Определение 7.2. $\hat{\theta}_n(X) \coloneqq \arg\max\limits_{\theta \in \Theta} L_n(X, \theta)$ – *оценка максимального правдоподобия (ОМП).*

Замечание 7.1. ОМП всегда принадлежит множеству Θ. ОМП может не существовать или не быть единственной.

Определение 7.3. $l_n(X, \theta) \coloneqq \ln L_n(X, \theta)$ – *логарифмическое правдоподобие.*

Утверждение 7.1. *Если правдоподобие является достаточно гладкой функцией относительно параметра θ, то можно искать ОМП среди решений уравнения правдоподобия:* $\dfrac{\partial}{\partial \theta} l_n(X, \theta) = 0$.

Замечание 7.2. Так как $l_n(X, \theta) = \sum\limits_{i=1}^{n} \ln p(X_i, \theta)$, то уравнение правдоподобия можно записать в виде $\sum\limits_{i=1}^{n} \dfrac{\partial}{\partial \theta} \ln p(X_i, \theta)$.

Решение задач на оценки максимального правдоподобия

Задача 7.1. Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim \mathcal{N}(\theta, \sigma^2)$, где σ^2 известно, а θ – параметр, причём:

а) $\theta \in \mathbb{R}$;

б) $\theta \in [a, b]$, где $a < b$ – известные числа;

в) $\theta \in (a, +\infty)$, где a – известное число.

Найти ОМП для θ.

Решение:

$$L_n(X, \theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X_i - \theta)^2}{2\sigma^2}} = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2} \sum\limits_{i=1}^{n} (X_i - \theta)^2}.$$

Видно, что правдоподобие $L_n(X, \theta)$ максимально (в зависимости от θ) тогда и только тогда, когда сумма $\sum\limits_{i=1}^{n} (X_i - \theta)^2$ минимальна (в зависимости от θ).

$$\sum_{i=1}^{n}(X_i-\theta)^2 = n\left(\overline{X^2}-2\theta\overline{X}+\theta^2\right).$$

$\overline{X^2}-2\theta\overline{X}+\theta^2$ – парабола с ветвями вверх и вершиной $\theta=\overline{X}$.

а) $\hat{\theta}_n=\overline{X}$.

б) $\hat{\theta}_n=\begin{cases}\overline{X}, & \overline{X}\in[a,b]\\ a, & \overline{X}<a\\ b, & \overline{X}>b\end{cases}.$

в) $\hat{\theta}_n=\begin{cases}\overline{X}, & \overline{X}>a\\ \nexists, & \overline{X}\leqslant a\end{cases}.$ □

Замечание 7.3. Если $\theta\in\mathbb{R}$, то ОМП $\hat{\theta}_n=\overline{X}$ совпадает с эффективной оценкой.

Утверждение 7.2. *Всякая эффективная оценка $\hat{\theta}_n(X)$, если она существует, является ОМП.*

Доказательство:

В силу критерия эффективности (лемма (5.1))

$$\hat{\theta}_n(X)-\theta \overset{P_\theta\text{-п. н.}}{=} \frac{1}{I_X(\theta)}\frac{\partial}{\partial\theta}l_n(X,\theta); \qquad \frac{\partial}{\partial\theta}l_n(X,\theta) \overset{P_\theta\text{-п. н.}}{=} I_X(\theta)\left(\hat{\theta}_n(X)-\theta\right).$$

В силу условий регулярности (R) $0<I_X(\theta)<+\infty$. Значит, так как оценка $\hat{\theta}_n(X)$ эффективная, то $\dfrac{\partial}{\partial\theta}l_n(X,\theta)=0$, причём слева от точки $\theta=\hat{\theta}_n(X)$ эта производная положительна, а справа – отрицательна. Таким образом, $\theta=\hat{\theta}_n(X)$ – точка максимума функции $l_n(X,\theta)$, то есть $\hat{\theta}_n(X)$ – ОМП. ■

Задача 7.2. Пусть $X=(X_1,\ldots,X_n)$ – выборка из дискретного распределения, $X_1\sim\mathrm{Pois}(\theta)$, где параметр $\theta>0$. Найти ОМП для θ.

Решение:

$$\begin{aligned}
&p(x,\theta)=P(X_1=x)=\frac{\theta^x}{x!}e^{-\theta}, \quad \text{где} \quad x=0,1,2,\ldots;\\
&l_n(X,\theta)=\sum_{i=1}^{n}\ln p(X_i,\theta)=\sum_{i=1}^{n}(X_i\ln\theta-\theta-\ln X_i!);\\
&\frac{\partial}{\partial\theta}l_n(X,\theta)=n\left(\frac{\overline{X}}{\theta}-1\right)=0;\\
&\hat{\theta}_n=\overline{X}, \quad \text{если} \quad \overline{X}>0.
\end{aligned}$$

Найдём вероятность того, что ОМП не существует:

$$P(\overline{X}=0)=P^n(X_1=0)=e^{-n\theta}.$$

Эта вероятность положительна, но стремится к 0 при $n \to +\infty$. □

Задача 7.3. Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim p(x, \theta) = \dfrac{1}{2}e^{-|x-\theta|}$, где параметр $\theta \in \mathbb{R}$. Найти ОМП для θ.

Решение:

$$l_n(X, \theta) = -\sum_{i=1}^{n} |X_i - \theta| - n \ln 2.$$

Видно, что логарифмическое правдоподобие $l_n(X, \theta)$ максимально (в зависимости от θ) тогда и только тогда, когда сумма $\sum\limits_{i=1}^{n} |X_i - \theta|$ минимальна (в зависимости от θ). Так как эта сумма состоит из выпуклых вниз функций, то у неё существует точка минимума (возможно, не единственная, если график суммы имеет горизонтальный участок). Возьмём производную этой суммы в тех точках, в которых она существует:

$$\frac{\partial}{\partial\theta}\sum_{i=1}^{n} |X_i - \theta| = -\sum_{i=1}^{n} \operatorname{sgn}(X_i - \theta) = -\sum_{i=1}^{n} \operatorname{sgn}(X_{(i)} - \theta) \doteqdot 0, \tag{7.1}$$

где «$\doteqdot 0$» означает, что выражение слева равно 0 (это соответствует случаю, когда график суммы имеет горизонтальный участок) или меняет свой знак (это соответствует случаю, когда существует единственная точка минимума, но в ней производная не равна 0, так как она не существует).

При изменении θ от значения $< X_{(1)}$ до значения $> X_{(n)}$ левая часть выражения (7.1) меняется от значения $-n$ до значения n. Причём каждый раз, когда θ переходит через значение $X_{(i)}$, левая часть выражения (7.1) увеличивается на 2. Значит, если $n = 2k - 1$, то решением (7.1) является $X_{(k)}$, а если $n = 2k$, то решением (7.1) является $[X_{(k-1)}, X_{(k)}]$, где $k \in \mathbb{N}$. Тогда в качестве ОМП, например, можно взять медиану выборки:

$$\hat{\theta}_n = \begin{cases} X_{(k)}, & \text{если} \quad n = 2k - 1 \\ \dfrac{X_{(k-1)} + X_{(k)}}{2}, & \text{если} \quad n = 2k \end{cases}.$$

□

Задача 7.4. Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim \mathcal{U}(\theta_1, \theta_2)$, где $\theta_1 < \theta_2$. Найти ОМП для параметра $\theta = \begin{pmatrix}\theta_1 \\ \theta_2\end{pmatrix}$.

Решение:

$$p(x, \theta) = \frac{1}{\theta_2 - \theta_1} I(\theta_1 \leqslant x \leqslant \theta_2);$$

$$L_n(X, \theta) = \frac{1}{(\theta_2 - \theta_1)^n} I(X_{(1)} \geqslant \theta_1,\ X_{(n)} \leqslant \theta_2). \tag{7.2}$$

Максимум функции правдоподобия $L_n(X, \theta)$ на множестве $\theta_1 < \theta_2$ достигается, когда в выражении (7.2) индикатор равен 1 и знаменатель дроби наименьший. Значит, $\hat{\theta}_1 = X_{(1)}$, $\hat{\theta}_2 = X_{(n)}$ – ОМП (если существует).

ОМП не существует только в том случае, когда $X_{(1)} = X_{(n)}$, но вероятность этого равна 0 (так как распределение равномерное). Таким образом, $\hat{\theta}_1 = X_{(1)}$, $\hat{\theta}_2 = X_{(n)}$ – ОМП с вероятностью 1, если $X_{(1)} < X_{(n)}$. $\square$

Замечание 7.4. Найденная оценка ОМП не является $\sqrt{n}$-асимптотически нормальной. Это связано с тем, что не выполнены условия регулярности (R), так как носитель распределения зависит от параметра θ.

Задача 7.5. Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim \mathcal{N}(\theta, 1)$, где параметр $\theta \geqslant 1$. Исследовать на $\sqrt{n}$-асимптотическую нормальность ОМП для θ.

Решение:

Согласно решению задачи (7.1) получаем ОМП для θ:

$$\hat{\theta}_n = \begin{cases} \overline{X}, & \text{если} \quad \overline{X} \geqslant 1 \\ 1, & \text{если} \quad \overline{X} < 1 \end{cases} = \overline{X} I(\overline{X} \geqslant 1) + I(\overline{X} < 1) = \left(\overline{X} - 1\right) I(\overline{X} \geqslant 1) + 1.$$

При $\theta > 1$ условия регулярности (R) выполнены, поэтому оценка будет $\sqrt{n}$-асимптотически нормальной. Далее рассмотрим случай $\theta = 1$:

$$\sqrt{n}(\hat{\theta}_n - \theta) = \sqrt{n}\left(\overline{X} - 1\right) I(\overline{X} \geqslant 1) = \sqrt{n}\left(\overline{X} - 1\right) I\left(\sqrt{n}\left(\overline{X} - 1\right) \geqslant 0\right) \ominus$$

Так как $X_1 \sim \mathcal{N}(\theta, 1)$, то $\overline{X} \sim \mathcal{N}\left(\theta, \dfrac{1}{n}\right)$, тогда $\sqrt{n}\left(\overline{X} - \theta\right) \sim \mathcal{N}(0, 1)$. Пусть $Y_n \coloneqq \sqrt{n}\left(\overline{X} - 1\right)$. Тогда $Y_n \sim \mathcal{N}(0, 1)$.

$$\ominus\, Y_n I(Y_n \geqslant 0).$$

Таким образом, при $\theta = 1$ получаем:

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} Y I(Y \geqslant 0), \quad \text{где} \quad Y \sim \mathcal{N}(0, 1).$$

Значит, при $\theta = 1$ оценка $\hat{\theta}_n$ не является $\sqrt{n}$-асимптотически нормальной для θ. $\square$

Условия регулярности для ОМП

Задача 7.6. Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim \mathrm{Exp}(\theta)$, где параметр $\theta > 0$. $p_\theta(x) = \theta e^{-\theta x} I(x > 0)$ – плотность. Проверить выполнение условий регулярности для ОМП $\hat{\theta} = \dfrac{1}{\overline{X}}$.

Решение:

($R0$): семейство распределений $\{P_\theta,\ \theta \in \Theta\}$, которому принадлежит наблюдение X_i, доминируемое с плотностью наблюдения $p_\theta(x)$ относительно некоторой меры λ, причём $p_\theta(x)$ различима по θ.

($R1$): $X = (X_1, \ldots, X_n)$ – выборка из P_θ при $n \to +\infty$. $\mathcal{N}_p = \{x > 0\}$.

($R2$): $\Theta \subseteq \mathbb{R}$ – открытый интервал. $\Theta = (0;\ +\infty)$.

($R3$): $\mathcal{N}_p \coloneqq \{p_\theta(x) > 0\}$ не зависит от θ.

($R4$): $p_\theta(x)$ непрерывно дифференцируема по θ $\ \forall x \in \mathcal{N}_p$.

В этих условиях уравнение максимального правдоподобия имеет состоятельное решение с вероятностью, стремящейся к 1.

($R5$): $p_\theta(x)$ трижды непрерывно дифференцируема по θ $\ \forall x \in \mathcal{N}_p$.

($R6$): равенство $\int\limits_{\mathcal{N}_p} p_\theta(x)\,dx$ можно дважды дифференцировать по θ, причём производную можно заносить под знак интеграла.

$$\int\limits_{\mathcal{N}_p} \frac{\partial}{\partial\theta} p_\theta(x)\,dx = \int\limits_0^{+\infty} \left(e^{-\theta x} - \theta x e^{-\theta x}\right) dx = \frac{1}{\theta} - E_\theta X_1 = 0.$$

Аналогично можно показать, что вторую производную можно заносить под знак интеграла.

($R7$): $0 < i(\theta) < +\infty$ $\ \forall \theta \in \Theta$.

$$\frac{\partial}{\partial\theta} \ln p_\theta(x) = \frac{1}{\theta} - x; \qquad i(\theta) = E_\theta \left(\frac{1}{\theta} - X_1\right)^2 = DX_1 = \frac{1}{\theta^2}.$$

($R8$):

$$\forall \theta_0 \in \Theta \quad \exists \delta > 0 \text{ и } M(x)\colon\ \forall \theta \in [\theta_0 - \delta,\ \theta_0 + \delta] \qquad \left|\frac{\partial^3}{\partial\theta^3} \ln p_\theta(x)\right| \leqslant M(x)$$

и $E_\theta M(X_1) < +\infty$.

Так как $\dfrac{\partial^3}{\partial\theta^3} \ln p_\theta(x) = \dfrac{2}{\theta^3}$, то условие выполнено.

В этих условиях уравнение максимального правдоподобия имеет $\sqrt{n}$-асимптотически нормальное и асимптотически эффективное решение с вероятностью, стремящейся к 1. □

Инвариантность ОМП относительно замены параметра

Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim \mathrm{Exp}(\theta)$, где параметр $\theta > 0$. Тогда $\overline{X}$ – эффективная оценка для $E_\theta X_1 = \dfrac{1}{\theta}$. Так как эффективные оценки являются оценками максимального правдоподобия, то $\overline{X}$ – ОМП для $E_\theta X_1 = \dfrac{1}{\theta}$, причём она асимптотически эффективна. ОМП – это $\hat{\theta} = \arg\max\limits_{\theta\in\Theta} L(X,\theta)$, где $L(X,\theta) = \prod\limits_{i=1}^{n} p_\theta(X_i)$.

Рассмотрим взаимно однозначное преобразование $\tau = \tau(\theta)\colon\ \Theta \to T$. Тогда $\theta = \theta(\tau) = \tau^{-1}(\tau)$ и
$$\hat{\tau} = \arg\max_{\tau\in T} L(X,\, \theta(\tau)) = \tau(\hat{\theta}).$$
Причём если оценка $\hat{\theta}$ асимптотически эффективна и преобразование $\tau(\theta)$ дифференцируемое и ненулевое, то из дельта-метода (утверждение (1.8)) следует, что оценка $\hat{\tau}$ тоже асимптотически эффективна.

Задача 7.7. Пусть $X_1,\,\ldots,X_n$ – независимые одинаково распределённые случайные величины, $X_1 \sim \mathrm{Exp}(\theta)$, где параметр $\theta > 0$. Наблюдение: $X = (X_{(1)}, X_{(2)})$. Найти ОМП для θ.

Решение:

$X_{(1)}, X_{(2)}$ – зависимые случайные величины. Найдём их совместную плотность при $0 < x_1 < x_2$ (иначе она равна 0):

$$\begin{aligned}
&p_{X_{(1)},\,X_{(2)}}(x_1,x_2) = \lim_{|O_{x_1}|,\,|O_{x_2}|\to 0} \frac{P(X_{(1)}\in O_{x_1},\, X_{(2)}\in O_{x_2})}{|O_{x_1}|\cdot|O_{x_2}|} = \\
&= \lim_{|O_{x_1}|,\,|O_{x_2}|\to 0} \frac{P(X_1\in O_{x_1},\, X_2\in O_{x_2},\, X_3,\,\ldots,X_n > x_2)}{|O_{x_1}|\cdot|O_{x_2}|} = \\
&= \lim_{|O_{x_1}|,\,|O_{x_2}|\to 0} \left(n(n-1)\cdot \frac{P(X_1\in O_{x_1})}{|O_{x_1}|}\cdot\frac{P(X_2\in O_{x_2})}{|O_{x_2}|}\cdot P^{n-2}(X_3 > x_2)\right) = \\
&= n(n-1)p_{X_1}(x_1)p_{X_1}(x_2)\left(1 - F_{X_1}(x_2)\right)^{n-2}.
\end{aligned}$$

Учитывая, что $p_{X_1}(x) = \theta e^{-\theta x}$ и $F_{X_1}(x) = 1 - e^{-\theta x}$ при $x > 0$ (иначе равны 0), получаем при $0 < x_1 < x_2$:

$$\begin{aligned}
&p_{X_{(1)},\,X_{(2)}}(x_1,x_2) = n(n-1)\theta^2 e^{-\theta(x_1+x_2)}\left(e^{-\theta x_2}\right)^{n-2};\\
&L_\theta(X_{(1)},X_{(2)}) = n(n-1)\theta^2 e^{-\theta\left(X_{(1)}+(n-1)X_{(2)}\right)};\\
&\frac{\partial}{\partial\theta}\ln L_\theta(X_{(1)},X_{(2)}) = \frac{2}{\theta} - \left(X_{(1)}+(n-1)X_{(2)}\right) = 0;\\
&\hat{\theta} = \frac{2}{X_{(1)}+(n-1)X_{(2)}}.
\end{aligned}$$

□

Семинар 8

Условное математическое ожидание в дискретном случае

Сначала рассмотрим пример. Эксперимент заключается в подбрасывании кубика 1 раз. Пусть ξ – выпавшее число очков, $\eta = I(\xi$ – нечётное$)$.

$$\begin{aligned}
&E\xi = 1\cdot\frac{1}{6} + \ldots + 6\cdot\frac{1}{6} = 3{,}5;\\
&E(\xi\,|\,\eta) = \left(2\cdot\frac{1}{3} + 4\cdot\frac{1}{3} + 6\cdot\frac{1}{3}\right)\cdot I(\eta = 0) + \left(1\cdot\frac{1}{3} + 3\cdot\frac{1}{3} + 5\cdot\frac{1}{3}\right)\cdot I(\eta = 1) =\\
&= 4I(\eta = 0) + 3I(\eta = 1) = 4 - \eta.
\end{aligned}$$

Пусть $(\Omega, \mathcal{F}, P)$ – дискретное вероятностное пространство, ξ, η – дискретные случайные величины.

$E(\xi\,|\,\eta) = \sum\limits_y E(\xi\,|\,\eta = y)\cdot I(\eta = y)$ – *условное математическое ожидание в дискретном случае*, где $E(\xi\,|\,\eta = y) = \sum\limits_x xP(\xi = x\,|\,\eta = y) = \dfrac{E(\xi\cdot I(\eta = y))}{P(\eta = y)} =:$ $=: \varphi(y)$. Тогда $E(\xi\,|\,\eta) = \sum\limits_y \varphi(y)I(\eta = y) = \varphi(\eta)$.

Позже мы получим общие формулы, которые возьмём за определение. Эти формулы для дискретного случая являются частным случаем общих формул.

Задача 8.1. Пусть ξ, η – независимые случайные величины с дискретным распределением, $\xi \sim \mathrm{Pois}(\lambda)$, $\eta \sim \mathrm{Pois}(\mu)$. Найти $E(\xi\,|\,\xi + \eta)$.

Решение:

Учитывая, что ξ и η – независимые случайные величины, получаем:

$$P(\xi = x\,|\,\xi + \eta = y) = \frac{P(\xi = x,\, \eta = y - x)}{P(\xi + \eta = y)} = \frac{P(\xi = x)P(\eta = y - x)}{P(\xi + \eta = y)} \ominus$$

Так как $\xi \sim \mathrm{Pois}(\lambda)$ и $\eta \sim \mathrm{Pois}(\mu)$, то $\xi + \eta \sim \mathrm{Pois}(\lambda + \mu)$.

$$\ominus\ \frac{\lambda^x e^{-\lambda}}{x!}\cdot\frac{\mu^{y-x}e^{-\mu}}{(y-x)!}\cdot\frac{y!e^{\lambda+\mu}}{(\lambda+\mu)^y} = C_y^x\left(\frac{\lambda}{\lambda+\mu}\right)^x\left(\frac{\mu}{\lambda+\mu}\right)^{y-x},$$

где $y = 0, 1, 2, \ldots$ и $x = 0, 1, 2, \ldots, y$.

Заметим, что $\xi\,|\,\xi + \eta = y \sim \mathrm{Bin}\left(y,\ \dfrac{\lambda}{\lambda+\mu}\right)$, тогда

$$E(\xi\,|\,\xi + \eta = y) = \sum_{x=0}^{y} xP(\xi = x\,|\,\xi + \eta = y) = y\cdot\frac{\lambda}{\lambda+\mu};$$

$$E(\xi \,|\, \xi + \eta) = (\xi + \eta) \cdot \frac{\lambda}{\lambda + \mu}.$$

□

Условное математическое ожидание в общем случае

Пусть η – одномерная случайная величина, тогда $\sigma(\eta) = \{\eta^{-1}(B) \ \ \forall B \in \mathcal{B}(\mathbb{R})\}$ – сигма-алгебра, порождённая η.

Приведём пример. Пусть η – дискретная случайная величина, принимающая не более чем счётное количество значений $\{y_k\}$, где $k \geqslant 1$. Тогда $\sigma(\eta) = \sigma\{D_k := \eta^{-1}(y_k)\}$.

Утверждение 8.1. *Случайная величина ξ измерима относительно сигма-алгебры $\sigma(\eta)$ тогда и только тогда, когда $\xi = \varphi(\eta)$ для некоторой борелевской функции φ.*

Пусть ξ, η – случайные величины на вероятностном пространстве $(\Omega, \mathcal{F}, P)$, причём $E|\xi| < +\infty$. Пусть $\hat{\xi} := E(\xi \,|\, \eta)$ – условное математическое ожидание. Выпишем свойства рассмотренного ранее условного математического ожидания для дискретной величины:

1) $\hat{\xi} = \varphi(\eta)$ – случайная величина, тогда по утверждению (8.1) $\hat{\xi}$ измерима относительно $\sigma(\eta)$;

2) $\forall A \in \sigma(\eta) \quad E\hat{\xi} I_A = E\xi I_A$.

Таким образом, из выписанных ранее формул для условного математического ожидания следуют эти свойства. Но можно показать, что верно и обратное, то есть если выполнены эти свойства, то выполняются выписанные ранее формулы для условного математического ожидания. Таким образом, эти свойства являются определяющими.

Определение 8.1. $\hat{\xi} := E(\xi \,|\, \eta)$ – *условное математическое ожидание* – это такая случайная величина, что:

1) $\hat{\xi}$ измерима относительно $\sigma(\eta)$;

2) $\forall A \in \sigma(\eta) \quad E\hat{\xi} I_A = E\xi I_A$.

Замечание 8.1. Условное математическое ожидание определяется относительно не случайной величины η, а относительно порождённой ей сигма-алгебры $\sigma(\eta)$, поэтому $\hat{\xi}$ можно записать так: $\hat{\xi} = E(\xi \,|\, \sigma(\eta))$.

Замечание 8.2. Можно показать (из теоремы Радона-Никодима), что если $E|\xi| < +\infty$, то $\hat{\xi}$ существует и почти наверное единственна.

Задача 8.2. Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim \mathcal{U}[0, 1]$. Найти $E(X_1 \,|\, X_{(n)})$.

Решение:

$E(X_1 \,|\, X_{(n)}) = \varphi(X_{(n)})$, где φ – борелевская функция, причем

$$P_{X_{(n)}}(x) = nx^{n-1} I(0 \leqslant x \leqslant 1).$$

Пусть $A = \{X_{(n)} \leqslant z\} = \{Z_1 \leqslant z, \ldots, X_n \leqslant z\}$, где $z \in [0, 1]$.

$$E\varphi(X_{(n)})I_A = \int\limits_0^z \varphi(y) n y^{n-1}\, dy;$$

$$EX_1 I_A = \int\limits_0^z \ldots \int\limits_0^z x_1\, dx_1 \ldots dx_n = \frac{z^2}{2} \cdot z^{n-1} = \frac{z^{n+1}}{2}.$$

Так как $E\varphi(X_{(n)})I_A = EX_1 I_A$, то, продифференцировав это равенство по z, получим следующее равенство, выполняющееся почти всюду в $z = [0, 1]$:

$$\varphi(z) n z^{n-1} = \frac{n+1}{2} z^n; \qquad \varphi(z) = \frac{n+1}{2n} z.$$

Так как условное математическое ожидание единственно, то

$$E(X_1 \,|\, X_{(n)}) = \frac{n+1}{2n} X_{(n)}.$$

$\square$

Свойства условного математического ожидания

Будем предполагать, что математические ожидания, участвующие в свойствах, конечны. Кроме этого, так как математическое ожидание выполняется с точностью до множества меры 0, то равенства в свойствах выполняются почти наверное.

Утверждение 8.2. *Свойства условного математического ожидания:*

1) $E(\varphi(\eta) \,|\, \eta) = \varphi(\eta) \quad \forall$ борелевской функции φ; в частности, $E(a \,|\, \eta) = a$;

2) $E(a\xi + b\zeta \,|\, \eta) = aE(\xi \,|\, \eta) + bE(\zeta \,|\, \eta)$;

3) $EE(\xi \,|\, \eta) = E\xi$ – формула полной вероятности (формула полного условного математического ожидания);

4) $E(\xi \cdot \varphi(\eta) \,|\, \eta) = \varphi(\eta) E(\xi \,|\, \eta) \quad \forall$ борелевской функции φ;

5) если ξ и η независимы, то $E(\xi \,|\, \eta) = E\xi$;

6) если ξ и η независимы, то $Eg(\xi, \eta) = EE(g(\xi, \eta) \,|\, \eta) = E\left(Eg(\xi, y)\right)_{y=\eta}$.

Задача 8.3. Пусть ξ, η – независимые случайные величины, $\xi \sim \mathrm{Exp}(\lambda)$, $\eta \sim \mathrm{Exp}(\mu)$. Найти $P(\xi < \eta)$.

Решение:

Используя свойство (6) из утверждения (8.2), получим:

$$P(\xi < \eta) = EI(\xi < \eta) = E\left(EI(\xi < y)\right)_{y=\eta} = EF_\xi(\eta) = \int\limits_{\mathbb{R}} F_\xi(y) p_\eta(y)\, dy =$$

$$= \int\limits_0^{+\infty} (1 - e^{-\lambda y})\mu e^{-\mu y}\, dy = \mu \left(-\frac{e^{-\mu y}}{\mu} + \frac{e^{-(\lambda+\mu)y}}{\lambda+\mu} \right)\Bigg|_0^{+\infty} = \mu \left(\frac{1}{\mu} - \frac{1}{\lambda+\mu} \right) = \frac{\lambda}{\lambda+\mu}.$$

□

Задача 8.4. Пусть $\xi \sim \mathcal{U}[-1,\, 1]$. Найти $E(\xi \,|\, \xi^2)$.

Решение:

1-й способ.

Так как две случайные величины, одна из которых является взаимно однозначным преобразованием другой, порождают одинаковые сигма-алгебры, а условное математическое ожидание определено относительно сигма-алгебры, то получаем:

$$E(\xi \,|\, \xi^2) = E(\xi \,|\, |\xi|) = E(|\xi| \operatorname{sgn}(\xi) \,|\, |\xi|) = |\xi| \cdot E(\operatorname{sgn}(\xi) \,|\, |\xi|) \ominus$$

(использовали свойство (4) из утверждения (8.2)).

В силу симметричности распределения относительно 0 получаем:

$$P(\operatorname{sgn}(\xi) = 1,\, |\xi| \leqslant z) = P(0 < \xi \leqslant z) = \frac{1}{2} P(|\xi| \leqslant z) = P(\operatorname{sgn}(\xi) = 1) \cdot P(|\xi| \leqslant z);$$

$$P(\operatorname{sgn}(\xi) = -1,\, |\xi| \leqslant z) = P(-z \leqslant \xi < 0) = \frac{1}{2} P(|\xi| \leqslant z) = P(\operatorname{sgn}(\xi) = -1) \cdot P(|\xi| \leqslant z);$$

$$F_{\operatorname{sgn}(\xi),\, |\xi|}(x, z) = F_{\operatorname{sgn}(\xi)}(x) \cdot F_{|\xi|}(z).$$

Таким образом, случайные величины $\operatorname{sgn}(\xi)$ и $|\xi|$ независимы, тогда согласно свойству (5) из утверждения (8.2) получаем:

$$\ominus\, |\xi| \cdot E(\operatorname{sgn}(\xi)) = |\xi| \cdot \left(1 \cdot \frac{1}{2} + 0 \cdot 0 - 1 \cdot \frac{1}{2} \right) = 0.$$

□

Утверждение 8.3. *Пусть величины $\xi, \tilde{\xi}, \eta$ заданы на одном вероятностном пространстве, причём $\begin{pmatrix} \xi \\ \eta \end{pmatrix} = \begin{pmatrix} \tilde{\xi} \\ \eta \end{pmatrix}$. Тогда $E(\xi \,|\, \eta) \overset{п.н.}{=} E(\tilde{\xi} \,|\, \eta)$.*

Приведём ещё один способ решения задачи (8.4).

Решение:

2-й способ.

Так как ξ распределена симметрично относительно 0, то $\xi \overset{d}{=} -\xi$. Тогда $h(\xi) = -h(\xi)$ $\forall$ борелевской функции h. Пусть $h(x) = \begin{pmatrix} x \\ x^2 \end{pmatrix}$. Тогда $\begin{pmatrix} \xi \\ \xi^2 \end{pmatrix} \overset{d}{=} \begin{pmatrix} -\xi \\ \xi^2 \end{pmatrix}$.

Значит, $E(\xi \,|\, \xi^2) \overset{п.н.}{=} E(-\xi \,|\, \xi^2)$. Воспользуемся свойством (1) из утверждения (8.2), получим: $E(\xi \,|\, \xi^2) \overset{п.н.}{=} -E(\xi \,|\, \xi^2)$. Значит, $E(\xi \,|\, \xi^2) \overset{п.н.}{=} 0$. □

Задача 8.5. Пусть $X = (X_1, \ldots, X_n)$ – выборка, $E|X_1| < +\infty$. Найти $E\left(X_1 \,|\, \overline{X}\right)$.

Решение:

Случайные векторы $\begin{pmatrix} X_i \\ \overline{X} \end{pmatrix}$ одинаково распределены $\forall i = \overline{1, n}$. Значит все условные математические ожидания вида $E\left(X_i \,|\, \overline{X}\right)$ равны почти наверное.

$$\sum_{i=1}^{n} E\left(X_i \,|\, \overline{X}\right) \overset{\text{п. н.}}{=} \sum_{i=1}^{n} E\left(X_1 \,|\, \overline{X}\right) \overset{\text{п. н.}}{=} nE(X_1 \,|\, \overline{X});$$
$$\sum_{i=1}^{n} E\left(X_i \,|\, \overline{X}\right) \overset{\text{п. н.}}{=} E\left(\sum_{i=1}^{n} X_i \,|\, \overline{X}\right) \overset{\text{п. н.}}{=} E\left(n\overline{X} \,|\, \overline{X}\right) \overset{\text{п. н.}}{=} n\overline{X};$$
$$E(X_1 \,|\, \overline{X}) \overset{\text{п. н.}}{=} \overline{X}.$$

□

Многомерное гауссовское распределение

Определение 8.2. Пусть $\xi \in \mathbb{R}^n$. Будем говорить, что ξ имеет *многомерное гауссовское распределение с параметрами a и Q* (обозначение: $\xi \sim \mathcal{N}_n(a, Q)$), где $a \in \mathbb{R}^n$, $Q \geqslant 0$ – матрица размера $n \times n$, если характеристическая функция ξ имеет вид

$$\varphi_\xi(t) = Ee^{it^T\xi} = e^{it^T a - \frac{t^T Q t}{2}}, \quad \text{где} \quad t \in \mathbb{R}^n.$$

Замечание 8.3. Если $Q > 0$, то плотность вероятности для случайного вектора ξ существует и её можно выписать в явном виде:

$$p_\xi(x) = \frac{1}{\left(\sqrt{2\pi}\right)^n \sqrt{\det Q}} e^{-\frac{1}{2}(x-a)^T Q^{-1}(x-a)}, \quad \text{где} \quad x \in \mathbb{R}^n.$$

Утверждение 8.4. *Пусть $\xi \sim \mathcal{N}_n(a, Q)$, A – матрица такого размера, чтобы её можно было умножить на ξ, тогда*

$$A\xi \sim \mathcal{N}(Aa,\, AQA^T).$$

Доказательство:

$$\varphi_{A\xi}(t) = Ee^{it^T A\xi} = \varphi_\xi(A^T t) = e^{it^T Aa - \frac{t^T AQA^T t}{2}}.$$

■

Сформулируем следствие. В качестве A возьмём вектор-строчку, у которой на i-м месте 1, а остальные элементы 0. Пусть $\xi = (\xi_1, \ldots, \xi_n)^T$, $a = (a_1, \ldots, a_n)^T$, $Q = (q_{ij})$. Тогда из утверждения (8.4) получим: $\xi_i \sim \mathcal{N}(a_i, q_{ii})$. Таким образом, все компоненты (и любой подвектор) гауссовского вектора тоже гауссовские. Обратное, вообще говоря, неверно.

Утверждение 8.5. *Пусть* $\xi \sim \mathcal{N}_n(a, Q)$, $\xi = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$. *Случайные величины (или векторы)* ξ_1 *и* ξ_2 *независимы тогда и только тогда, когда* $\operatorname{Cov}(\xi_1, \xi_2) = 0$.

Задача 8.6. Пусть $\begin{pmatrix} \xi \\ \eta \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 3 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}\right)$.

Найти: а) $E(\xi + \eta \,|\, \xi - \eta)$; б) $E(e^{\xi+\eta} \,|\, \xi - \eta)$.

Решение:

а) Пусть $X := \xi + \eta$, $Y := \xi - \eta$.

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \xi \\ \eta \end{pmatrix} \sim$$
$$EX = E\xi + E\eta = 3 + 3 = 6; \qquad EY = E\xi - E\eta = 3 - 3 = 0;$$
$$DX = D\xi + D\eta + 2\operatorname{Cov}(\xi, \eta) = 1 + 2 + 2 = 5;$$
$$DY = D\xi + D\eta - 2\operatorname{Cov}(\xi, \eta) = 1 + 2 - 2 = 1;$$
$$\operatorname{Cov}(X, Y) = D\xi + \operatorname{Cov}(\xi, \eta) - \operatorname{Cov}(\xi, \eta) - D\eta = 1 - 2 = -1;$$
$$\sim \mathcal{N}_2\left(\begin{pmatrix} 6 \\ 0 \end{pmatrix}, \begin{pmatrix} 5 & -1 \\ -1 & 1 \end{pmatrix}\right).$$

Пусть $Z := X - aY$, где a – некоторая константа. Тогда $X = aY + Z$. Подберём константу a так, чтобы случайные величины Y и Z были независимы. Так как $\begin{pmatrix} Y \\ Z \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & -a \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}_2$, то Y и Z независимы тогда и только тогда, когда $\operatorname{Cov}(Y, Z) = 0$.

$$\operatorname{Cov}(Y, Z) = \operatorname{Cov}(X, Y) - a \cdot DY = -1 - a \cdot 1 = 0; \qquad a = -1.$$

Таким образом, $X = Z - Y$. Теперь, учитывая, что Y и Z независимы, получаем:

$$E(\xi + \eta \,|\, \xi - \eta) = E(X \,|\, Y) = E(Z - Y \,|\, Y) = E(Z \,|\, Y) - E(Y \,|\, Y) = EZ - Y =$$
$$EZ = E(X + Y) = EX + EY = 6 + 0 = 6;$$
$$= 6 - Y = 6 + \eta - \xi.$$

б) Так как Y и Z независимы, то Y и e^Z тоже независимы. Тогда получаем:

$$E(e^{\xi+\eta} \,|\, \xi - \eta) = E(e^X \,|\, Y) = E(e^{Z-Y} \,|\, Y) = E(e^Z \cdot e^{-Y} \,|\, Y) = e^{-Y} \cdot Ee^Z =;$$
$$Z = 2\xi \sim \mathcal{N}(6, 4); \qquad \varphi_Z(-i) = Ee^Z = e^{6 + \frac{1}{2} \cdot 4} = e^8;$$
$$= e^{8-Y} = e^{8+\eta-\xi}.$$

□

Семинар 9

Условное распределение

Определение 9.1. Пусть ξ, η – случайные величины, заданные на вероятностном пространстве $(\Omega, \mathcal{F}, P)$, причём $E|\xi| < +\infty$. Тогда

$$\forall C \in \mathcal{F} \qquad P(C \,|\, \eta) \coloneqq E(I_C \,|\, \eta)$$

– *условная вероятность события C относительно η.*

Так как условное математическое ожидание является случайной величиной, определённой с точностью до множества меры 0, то условная вероятность является функцией, зависящей от множества C и элементарного исхода ω, определённой с точностью до множества меры 0: $P(C \,|\, \eta) = g(C, \omega)$. Хотелось бы, чтоб при фиксированном ω условная вероятность обладала свойствами вероятности: вероятность пустого множества равна 0, вероятность всего пространства элементарных исходов равна 1, аддитивность. Однако с аддитивностью возникают сложности.

Пусть $C_1, C_2 \in \mathcal{F}$ – непересекающиеся события, тогда

$$P(C_1 + C_2 \,|\, \eta) \overset{\text{п. н.}}{=} P(C_1 \,|\, \eta) + P(C_2 \,|\, \eta).$$

Пусть это равенство не выполняется на множестве нулевой меры $\mathcal{N}_{C_1C_2}$. Рассмотрим множество $\mathcal{N} \coloneqq \bigcup \mathcal{N}_{C_1C_2}$ – объединение множеств нулевой вероятности по всем непересекающимся событиям C_1, C_2. Таких множеств под знаком объединения может быть более чем счётное количество. Тогда множество $\mathcal{N}$ может иметь положительную вероятность. В этом случае для всех элементарных исходов из множества $\mathcal{N}$ функция $g(\cdot, \omega)$ не является мерой, так как для неё не выполнено свойство аддитивности.

Определение 9.2. Если $P(C \,|\, \eta)$ является мерой для почти всех элементарных исходов, то будем называть её *регулярной условной вероятностью события C относительно η.*

Определение 9.3. *Условное распределение ξ относительно η* – это регулярная условная вероятность $P_\xi(B \,|\, \eta) \coloneqq P(\xi \in B \,|\, \eta) \quad \forall B \in \mathcal{B}(\mathbb{R})$.

Условное распределение ξ относительно η является борелевской функцией от случайной величины η: $g_B(\eta) \coloneqq P_\xi(B \,|\, \eta)$.

Определение 9.4. Функцию $g_B(y) = P_\xi(B \,|\, \eta = y)$ будем называть *условным распределением ξ относительно события $\{\eta = y\}$.*

Теорема 9.1. *Пусть случайный вектор (ξ, η) имеет совместную плотность $f_{\xi,\eta}(x, y)$ относительно $\lambda \times \mu$. Тогда*

$$\forall B \in \mathcal{B}(\mathbb{R}) \qquad P_\xi(B \,|\, \eta = y) = \int\limits_B f_{\xi\,|\,\eta}(x \,|\, y)\lambda(dx),$$

где $f_{\xi\,|\,\eta}(x\,|\,y) = \begin{cases} \dfrac{f_{\xi,\eta}(x,y)}{f_\eta(y)}, & \text{если} \quad f_\eta(y) > 0 \\ 0, \quad \text{если} \quad f_\eta(y) = 0 \end{cases}.$

Определение 9.5. Функция $f_{\xi\,|\,\eta}(x\,|\,y)$ из теоремы (9.1) называется *условной плотностью* ξ *относительно события* $\{\eta = y\}$.

В качестве следствия получаем формулу для расчёта условного математического ожидания $h(\xi)$ относительно $\eta = y$:

$$E(h(\xi)\,|\,\eta = y) = \int\limits_{\mathbb{R}} h(x) f_{\xi\,|\,\eta}(x\,|\,y) \lambda(dx).$$

Решение задач на условное распределение

Задача 9.1. Пусть $X, Y \sim \mathcal{U}(0,1)$ – независимые одинаково распределённые случайные величины. Найти $E(X\,|\,XY)$.

Решение:

Пусть $Z \coloneqq XY$ и $g\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} X \\ Z \end{pmatrix}$ – взаимно однозначное преобразование. Тогда $g^{-1}\begin{pmatrix} X \\ Z \end{pmatrix} = \begin{pmatrix} X \\ Z/X \end{pmatrix}$.

Запишем совместную плотность:

$$f_{X,Y}(x,y) = f_X(x) f_Y(y) = I(0 < x < 1,\ 0 < y < 1).$$

Используем формулу

$$f_{g(X,Y)}(u,v) = f_{X,Y}(x,y) \cdot |J_{g^{-1}}(u,v)|,$$

где $\begin{pmatrix} x \\ y \end{pmatrix} = g^{-1}\begin{pmatrix} u \\ v \end{pmatrix}$.

$$J_{g^{-1}} = \det \begin{pmatrix} 1 & 0 \\ -\dfrac{z}{x^2} & \dfrac{1}{x} \end{pmatrix} = \frac{1}{x};$$

$$f_{X,Z}(x,z) = I\left(0 < x < 1,\ 0 < \frac{z}{x} < 1\right) \cdot \frac{1}{|x|} = \frac{1}{x} I(0 < z < x < 1);$$

$$\forall z \in (0,1) \qquad p_Z(z) = \int\limits_{\mathbb{R}} p_{X,Z}(x,z)\,dx = \int\limits_z^1 \frac{dx}{x} = \ln x \Big|_z^1 = -\ln z;$$

$$p_{X\,|\,Z}(x\,|\,z) = \frac{p_{X,Z}(x,z)}{p_Z(z)} = -\frac{1}{x \ln z} I(0 < z < x < 1);$$

$$
\begin{aligned}
&E(X \mid Z=z) = \int\limits_{\mathbb{R}} x p_{X\mid Z}(x \mid z)\,dx = -\int\limits_{z}^{1} \frac{x}{x \ln z}\,dx = \frac{z-1}{\ln z};\\
&E(X \mid XY) = \frac{XY-1}{\ln(XY)}.
\end{aligned}
$$

$\square$

Задача 9.2. Пусть $\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}_2\left(\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}\right)$. Найти $f_{X\mid Y}(x \mid y)$.

Решение:

$$
\begin{aligned}
&Q = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix}; \qquad \det Q = \sigma_1^2\sigma_2^2(1-\rho^2);\\
&Q^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix};\\
&\begin{pmatrix} x-a_1 & y-a_2 \end{pmatrix} Q^{-1} \begin{pmatrix} x-a_1 \\ y-a_2 \end{pmatrix} =\\
&= \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \left(\sigma_2^2(x-a_1)^2 + \sigma_1^2(y-a_2)^2 - 2\rho\sigma_1\sigma_2(x-a_1)(y-a_2)\right);\\
&f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{\sigma_2^2(x-a_1)^2+\sigma_1^2(y-a_2)^2-2\rho\sigma_1\sigma_2(x-a_1)(y-a_2)}{2\sigma_1^2\sigma_2^2(1-\rho^2)}};\\
&Y \sim \mathcal{N}(a_2, \sigma_2^2); \qquad f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(y-a_2)^2}{2\sigma_2^2}};\\
&f_{X\mid Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} =\\
&= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}} e^{-\frac{1}{2\sigma_1^2(1-\rho^2)}\left((x-a_1)^2 + \frac{\sigma_1^2}{\sigma_2^2}\rho^2(y-a_2)^2 - 2\rho\frac{\sigma_1}{\sigma_2}(x-a_1)(y-a_2)\right)} =\\
&= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}} e^{-\frac{1}{2\sigma_1^2(1-\rho^2)}\left((x-a_1) - \frac{\sigma_1}{\sigma_2}\rho(y-a_2)\right)^2}.
\end{aligned}
$$

$\square$

Замечание 9.1. $f_{X\mid Y}(x \mid y)$ – плотность $\mathcal{N}\left(a_1 + \frac{\sigma_1}{\sigma_2}\rho(y-a_2),\ \sigma_1^2(1-\rho^2)\right)$. Тогда получаем:

$$
E(X \mid Y=y) = a_1 + \frac{\sigma_1}{\sigma_2}\rho(y-a_2); \qquad D(X \mid Y=y) = \sigma_1^2(1-\rho^2).
$$

Семинар 10

Байесовские оценки

Пусть $X \sim P_\theta$, где параметр $\theta \in \Theta$.

Определение 10.1. $R_{\hat{\theta}(X)}(\theta) = R(\theta, \hat{\theta}(X)) \coloneqq E_\theta(\hat{\theta}(X) - \theta)^2$ – *среднеквадратическая функция риска.*

Определение 10.2. $r_Q \coloneqq \min\limits_{\hat{\theta}(X)} \int\limits_\Theta R_{\hat{\theta}(X)}(t) Q(dt)$ – *байесовский риск*, где Q – вероятностное распределение на множестве Θ.

Распределение Q должно быть известно и называется *априорным* (известным до опыта) распределением.

Определение 10.3. $\theta^*_Q \coloneqq \arg\min\limits_{\hat{\theta}(X)} \int\limits_\Theta R_{\hat{\theta}(X)}(t) Q(dt)$ – *байесовская оценка.*

Пусть $\Theta \subseteq \mathbb{R}^1$, $q(t)$ – плотность априорного распределения Q относительно некоторой меры λ, $p_\theta(x)$ – плотность распределения P_θ относительно некоторой меры μ (условная плотность $X|\theta$). Тогда $(Q, X) \sim q(t)p_t(X)$.

Теорема 10.1. *Байесовская оценка со среднеквадратической функцией риска вычисляется по формуле*

$$\theta^*_Q(X) = E(\theta \,|\, X) = \int\limits_\Theta t \cdot q(t \,|\, X) \lambda(dt),$$

где $q(t \,|\, X) = \dfrac{q(t)p_t(X)}{p(X)}$ – *апостериорная (полученная из опыта) условная плотность, где* $p(X) = \int\limits_\Theta q(t)p_t(X)\lambda(dt)$.

Решение задач на байесовские оценки

Задача 10.1. Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim \mathrm{Bin}(1, \theta)$, где параметр $\theta \in (0, 1)$. $P(X_1 = x) = \theta^x (1 - \theta)^{1-x}$, где $x = 0$ или 1. Тогда $p_\theta(X) = \theta^{\sum\limits_{i=1}^n X_i} (1 - \theta)^{n - \sum\limits_{i=1}^n X_i}$. Пусть $\theta \sim \mathrm{Beta}(\alpha, \beta)$ – априорное распределение, где $a > 0$, $\beta > 0$. $q(t) = \dfrac{t^{\alpha-1}(1-t)^{\beta-1}}{B(\alpha, \beta)} I(t \in (0, 1))$. Вычислить байесовскую оценку θ^*_Q.

Решение:

$$\begin{aligned}
&q(t)p_t(X) = \frac{t^{n\overline{X}+\alpha-1}(1-t)^{n-n\overline{X}+\beta-1}}{B(\alpha,\beta)} I(t\in(0,1));\\
&\theta|X \sim \mathrm{Beta}\left(n\overline{X}+\alpha,\, n-n\overline{X}+\beta\right);\\
&\theta^*_Q = E(\theta|X) = \frac{n\overline{X}+\alpha}{n+\alpha+\beta}.
\end{aligned}$$

□

Замечание 10.1. Сравним найденную байесовскую оценку θ^*_Q с эффективной оценкой $\overline{X}$. Для этого запишем байесовскую оценку в следующем виде:

$$\theta^*_Q = \frac{n}{n+\alpha+\beta}\overline{X} + \frac{\alpha+\beta}{n+\alpha+\beta}\cdot\frac{\alpha}{\alpha+\beta}.$$

Так как $E\theta = \dfrac{\alpha}{\alpha+\beta}$ для распределения $Q \sim \mathrm{Beta}(\alpha,\beta)$, то байесовская оценка является взвешенным средним эффективной оценки $\overline{X}$ и априорной оценки $E\theta = \dfrac{\alpha}{\alpha+\beta}$.

Если $\alpha \to \infty$, $\beta \to \infty$ при условии, что $\dfrac{\alpha}{\beta}$ постоянно и n фиксировано, то $\theta^*_Q \xrightarrow{P} \dfrac{\alpha}{\alpha+\beta}$.

Если α и β фиксированы, а $n\to\infty$, то $\theta^*_Q \xrightarrow{P} \theta \quad \forall\theta\in(0,1)$.

$\theta^*_Q = \overline{X}$, если $\alpha=\beta=0$, но распределение $\mathrm{Beta}(0,0)$ не является вероятностным.

Если $\theta \sim \mathcal{U}(0,1) = \mathrm{Beta}(1,1)$ – априорное распределение, то $\theta^*_{\mathcal{U}} = \dfrac{\overline{X}+\frac{1}{n}}{1+\frac{2}{n}}$.

Задача 10.2. Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim \mathcal{N}(0,\theta)$, где параметр $\theta > 0$. Используя априорное сопряжённое распределение, найти байесовскую оценку θ^*_Q.

Решение:

$$p_\theta(X) = \frac{1}{\left(\sqrt{2\pi}\right)^n}\cdot\frac{1}{\theta^{\frac{n}{2}}} e^{-\frac{1}{2\theta}\sum\limits_{i=1}^n X_i^2}.$$

Пусть $\tau = \dfrac{1}{\theta} \sim \mathrm{Gamma}(\lambda,k)$, тогда

$$p_\tau(X) \propto \tau^{\frac{n}{2}} e^{-\frac{\tau}{2}\cdot n\overline{X^2}},$$

где $\propto$ – равенство с точностью до константы (которая не важна, так как в будущем сократится).

$$
\begin{aligned}
&q(t) \propto t^{k-1}e^{-\lambda t}I(t>0);\\
&q(t)p_t(X) \propto t^{\frac{n}{2}+k-1}e^{-t\left(\frac{n}{2}\overline{X^2}+\lambda\right)}I(t>0);\\
&\tau|X \sim \mathrm{Gamma}\left(\frac{n}{2}\overline{X^2}+\lambda,\ \frac{n}{2}+k\right);\\
&\theta^*_Q = E(\theta|X) = E\left(\frac{1}{\tau}\middle| X\right) \circleddash\\
&E\left(\frac{1}{\tau}\right) = \int\limits_0^{+\infty}\frac{\lambda^k}{t}\cdot\frac{t^{k-1}e^{-\lambda t}}{\Gamma(k)}\,dt = \frac{\lambda}{k-1}\int\limits_0^{+\infty}\frac{t^{k-2}e^{-\lambda t}\lambda^{k-1}}{\Gamma(k-1)}\,dt = \frac{\lambda}{k-1},
\end{aligned}
$$

так как $\dfrac{t^{k-2}e^{-\lambda t}\lambda^{k-1}}{\Gamma(k-1)}$ – плотность распределения $\Gamma(\lambda,\, k-1)$.

$$
\circleddash \frac{\frac{n}{2}\overline{X^2}+\lambda}{\frac{n}{2}+k-1}.
$$

□

Замечание 10.2. Так как

$$
\theta^*_Q = \frac{\frac{n}{2}}{\frac{n}{2}+k-1}\overline{X^2} + \frac{k-1}{\frac{n}{2}+k-1}\cdot\frac{\lambda}{k-1},
$$

то байесовская оценка представляет собой взвешенную среднее эффективной оценки $\overline{X^2}$ для θ и априорной оценки.

Эта оценка состоятельна, так как $\theta^*_Q \underset{n\to+\infty}{\to} \theta \quad \forall\theta>0$.

Задача 10.3. Пусть $X = (X_1,\, \ldots, X_n)$ – выборка, $X_1 \sim \mathcal{N}(\theta, 1)$, где параметр $\theta \in \in \Theta = \{\pm a\}$. Найти байесовскую оценку θ^*_Q для $Q \sim \pm a$ с вероятностями $\frac{1}{2}$. Будет ли эта оценка состоятельной?

Решение:

$$
\begin{aligned}
&q(t) = \frac{1}{2}I(t=\pm a);\\
&p_\theta(X) \propto e^{-\frac{1}{2}\sum\limits_{i=1}^n (X_i-\theta)^2} = e^{-\frac{n\overline{X^2}}{2}}\cdot e^{n\theta\overline{X}}\cdot e^{-\frac{n\theta^2}{2}}
\end{aligned}
$$

где $\propto$ – равенство с точностью до константы (которая не важна, так как в будущем сократится).

$$q(t)p_t(X) \propto I(t=\pm a)\cdot e^{nt\overline{X}}\cdot e^{-\frac{nt^2}{2}} \propto I(t=\pm a)\cdot e^{nt\overline{X}};$$

$$q(t|X) = \frac{e^{nt\overline{X}} I(t=\pm a)}{e^{-na\overline{X}}+e^{na\overline{X}}};$$

$$\theta_Q^* = E(\theta|X) = \frac{a\left(e^{na\overline{X}}-e^{-na\overline{X}}\right)}{e^{na\overline{X}}+e^{-na\overline{X}}} = a\,\mathrm{th}\left(na\overline{X}\right).$$

Рассмотрим случай $\theta = a$. Тогда $\overline{X} \overset{\text{п. н.}}{\to} \theta$, значит, $a\overline{X} \overset{\text{п. н.}}{\to} \theta^2$, значит, $\mathrm{th}(na\overline{X}) \overset{\text{п. н.}}{\to} 1$, значит, $\theta_Q^* \overset{\text{п. н.}}{\to} \theta$. Таким образом, при $\theta = a$ оценка состоятельна. Аналогично можно показать состоятельности оценки и при $\theta = -a$. □

Минимаксные оценки

Определение 10.4. $\theta^*(X) = \min\limits_{\hat\theta(X)}\sup\limits_{\theta\in\Theta} R(\theta, \hat\theta(X))$ – *минимаксная оценка.*

Определение 10.5. Q – *наименее благоприятное априорное распределение*, если $r_Q \geqslant r_{Q'}\quad \forall$ априорного распределения Q' на Θ.

Лемма 10.1. *Если $\exists$ байесовская оценка θ_Q^*: $\sup\limits_{\theta\in\Theta} R_{\theta_Q^*}(\theta) = r_Q$, то θ_Q^* – минимаксная оценка и Q – наименее благоприятное априорное распределение.*

В качестве следствия получаем: если θ_Q^* – байесовская оценка с постоянным риском (то есть не зависит от θ), то θ_Q^* – минимаксная оценка.

Решение задач на минимаксные оценки

Задача 10.4. Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim \mathrm{Bin}(1,\theta)$, где параметр $\theta \in (0,1)$. Найти минимаксную оценку для θ.

Решение:

При решении задачи (10.1) мы получили для сопряжённого априорного распределения $\theta \sim \mathrm{Beta}(\alpha,\beta)$, где $\alpha,\beta > 0$, байесовскую оценку $\theta_Q^* = \dfrac{n\overline{X}+\alpha}{n+\alpha+\beta} = a\overline{X}+b$, где $a = \dfrac{n}{n+\alpha+\beta}$ и $b = \dfrac{\alpha}{n+\alpha+\beta}$, причём $a, b \in (0,1)$.

Вычислим функцию риска для байесовской оценки:

$$R(\theta,\theta_Q^*) = E(\theta_Q^*-\theta)^2 = D_\theta\theta_Q^* + (E\theta_Q^*-\theta)^2 = a^2 D_\theta\overline{X} + ((a-1)\theta+b)^2 = \\ = \frac{a^2}{n}\theta(1-\theta) + (a-1)^2\theta^2 + 2b(a-1)\theta + b^2. \tag{10.1}$$

Чтобы функция $R(\theta, \theta_Q^*)$ не зависела от θ, в выражении (10.1) коэффициенты при θ^2 и при θ должны быть равны 0:

$$\begin{cases} -\dfrac{a^2}{n} + (a-1)^2 = 0 \quad (1) \\ \dfrac{a^2}{n} + 2b(a-1) = 0 \quad (2) \end{cases};$$

$$(1) \quad 1 - a = \frac{a}{\sqrt{n}}; \qquad a = \frac{1}{1 + \dfrac{1}{\sqrt{n}}} = \frac{\sqrt{n}}{\sqrt{n}+1};$$

$$(1) + (2): \quad (a-1)(a-1+2b) = 0; \qquad b = \frac{1-a}{2} = \frac{a}{2\sqrt{n}} = \frac{1}{2\left(\sqrt{n}+1\right)}.$$

Запишем байесовскую оценку с постоянным риском:

$$\theta_Q^* = a\overline{X} + \frac{a}{2\sqrt{n}} = \frac{\left(\overline{X} + \dfrac{1}{2\sqrt{n}}\right)\sqrt{n}}{\sqrt{n}+1} = \frac{n\overline{X} + \dfrac{\sqrt{n}}{2}}{n + \sqrt{n}}.$$

Такая байесовская оценка соответствует априорному распределению $Q \sim \sim \mathrm{Beta}\left(\dfrac{\sqrt{n}}{2}, \dfrac{\sqrt{n}}{2}\right)$. Тогда по лемме (10.1) получаем, что θ_Q^* – минимаксная оценка.

$\square$

Замечание 10.3. Сравним найденную минимаксную оценку с эффективной оценкой $\overline{X}$. Для этого запишем риск минимаксной оценки и оценки $\overline{X}$:

$$R_{\theta_Q^*}(\theta) = b^2 = \frac{1}{4\left(\sqrt{n}+1\right)^2}; \qquad R_{\overline{X}}(\theta) = D_\theta \overline{X} = \frac{\theta(1-\theta)}{n}.$$

Графиком $R_{\overline{X}}(\theta)$ является парабола с ветвями вниз и вершиной в точке $\theta = = \dfrac{1}{2}$, $R = \dfrac{1}{4n}$, проходящая через точки $(0,0)$ и $(0,1)$. Значений $R_{\theta_Q^*}(\theta)$ постоянно и меньше значения вершины параболы $R_{\overline{X}}(\theta)$. Тогда в некотором интервале $I_n = = \left(\dfrac{1}{2} - \delta_n;\ \dfrac{1}{2} + \delta_n\right)$ вокруг вершины параболы получаем, что $R_{\theta_Q^*}(\theta) < R_{\overline{X}}(\theta)$, а за его пределами $R_{\theta_Q^*}(\theta) \geqslant R_{\overline{X}}(\theta)$.

Найдём δ_n из условия $R_{\theta_Q^*}(\theta) = R_{\overline{X}}(\theta)$ при $\theta = \dfrac{1}{2} + \delta_n$:

$$\frac{1}{n}\left(\frac{1}{2} + \delta_n\right)\left(\frac{1}{2} - \delta_n\right) = \frac{1}{4\left(\sqrt{n}+1\right)^2};$$

$$\frac{1}{4} - \delta_n^2 = \frac{n}{4\left(\sqrt{n}+1\right)^2};$$

МЕХАНИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ МГУ ИМЕНИ М.В. ЛОМОНОСОВА

teach-in
ЛЕКЦИИ УЧЕНЫХ МГУ

$$\delta_n = \frac{1}{2}\sqrt{1 - \left(\frac{\sqrt{n}}{\sqrt{n}+1}\right)^2}.$$

Как видно, для малых значений n получается, что δ_n близко к 1, поэтому интервал I_n в этом случае будет покрывать большую часть интервала $(0,1)$. А при $n \to +\infty$ интервал I_n вырождается в малую окрестность точки $\left\{\frac{1}{2}\right\}$, причём

$$\frac{\sup\limits_{\theta} R_{\overline{X}}}{\sup\limits_{\theta} R_{\theta^*_Q}} = \frac{4\left(\sqrt{n}+1\right)^2}{4n} = \left(1 + \frac{1}{\sqrt{n}}\right)^2 \underset{n\to+\infty}{\to} 1.$$

Таким образом, минимаксная оценка может быть пригодна только при малых значениях n.

Семинар 11

Достаточные статистики и оптимальные оценки

Пусть $X = (X_1, \ldots, X_n)$ – выборка. Распределение наблюдения X принадлежит параметрическому классу $\{P_\theta,\ \theta \in \Theta\}$. Запишем для оценки $\hat{\theta}(X)$ среднеквадратичную функцию риска: $R_{\hat{\theta}(X)}(\theta) \coloneqq E_\theta(\hat{\theta}(X) - \theta)^2$. Мы хотим, чтобы у наилучшей оценки функция $R_{\hat{\theta}(X)}(\theta)$ была равномерно наименьшей по всем значениям $\theta \in \Theta$. Если рассматривать всевозможные оценки параметра θ, то наилучшей оценки в написанном выше смысле не найдётся. Поэтому сузим класс оценок, в котором будем искать наилучшую, до класса несмещённых оценок. В этом случае среднеквадратичная функция риска превращается в дисперсию.

Определение 11.1. $\tau^*(X)$ – *оптимальная оценка для* $\tau(\theta)$, если:

1) $E_\theta \tau^*(X) = \tau(\theta) \quad \forall \theta \in \Theta$;

2) $D_\theta \tau^*(X) \leqslant D_\theta \hat{\tau}(X) \quad \forall \theta \in \Theta$ для любых $\hat{\tau}(X)$, являющихся несмещёнными оценками для $\tau(\theta)$.

Определение 11.2. $S(X)$ – *достаточная статистика для* $\{P_\theta,\ \theta \in \Theta\}$ *(для* θ*)*, если $\exists$ вариант условного распределения $P_\theta(X \in B \,|\, S(X) = s)$, не зависящий от параметра θ при любых значениях s.

Теорема 11.1 (Критерий факторизации)**.** *Пусть наблюдение X имеет плотность $p_\theta(X)$ относительно некоторой меры μ. Тогда статистика $S(X)$ достаточна для θ тогда и только тогда, когда $p_\theta(X) = h(X) \cdot g(\theta,\ S(X))$ для почти всех значений X по мере μ.*

Теорема 11.2 (Теорема Колмогорова-Блэкуэлла-Рао)**.** *Пусть $\hat{\tau}(X)$ – несмещённая оценка для $\tau(\theta)$, $S(X)$ – достаточная статистика для θ, $\tau^*(X) \coloneqq E_\theta(\hat{\tau}(X) \,|\, S(X))$. Тогда:*

1) $\tau^*(X)$ – несмещённая оценка для θ;

2) $D_\theta \tau^*(X) \leqslant D_\theta \hat{\tau}(X) \quad \forall \theta \in \Theta$;

3) $D_\theta \tau^*(X) = D_\theta \hat{\tau}(X) \iff \hat{\tau}(X)$ – измеримая функция от $S(X)$.

Сформулируем следствие из теоремы выше.

Утверждение 11.1. *Если $\tau(\theta)$ допускает оптимальное оценивание, то оптимальная оценка является функцией от достаточной статистики.*

Определение 11.3. Достаточная статистика $S(X)$ называется *полной для* $\{P_\theta,\ \theta \in \Theta\}$ *(для* θ*)*, если уравнение $E_\theta \varphi(S(X)) = 0 \ \ \forall \theta \in \Theta$ имеет единственное решение $\varphi(S(X)) \overset{P_\theta\text{-п. н.}}{=} 0 \ \ \forall \theta \in \Theta$.

Если существуют две несмещённых оценки для $\tau(\theta)$, то есть $E_\theta \varphi_1(S(X)) = E_\theta \varphi_2(S(X)) = \tau(\theta) \ \ \forall \theta \in \Theta$, то рассмотрим функцию $\varphi = \varphi_1 - \varphi_2$. Если статистика полная, то $\varphi \overset{P_\theta\text{-п. н.}}{=} 0 \ \ \forall \theta \in \Theta$.

Утверждение 11.2. *Оптимальная оценка является решением уравнения несмещённости $E_\theta \varphi(S(X)) = \tau(\theta) \ \ \forall \theta \in \Theta$.*

Решение задач на достаточные статистики и оптимальные оценки

Задача 11.1. Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim \mathrm{Bin}(1, \theta)$, где параметр $\theta \in (0, 1)$. Найти оптимальную оценку для $\tau(\theta) = D_\theta X_1 = \theta(1-\theta)$.

Решение:

1-й шаг. Найдём достаточную статистику по критерию факторизации (теорема (11.1)):

$$L(X, \theta) = \prod_{i=1}^{n} \theta^{X_i}(1-\theta)^{1-X_i} = \theta^{n\overline{X}}(1-\theta)^{n-n\overline{X}} = h(X) \cdot g(\theta, n\overline{X}),$$

где $h(X) = 1,\ \ g(\theta,\, n\overline{X}) = \theta^{n\overline{X}}(1-\theta)^{n-n\overline{X}}$. Таким образом $\overline{X}$ (или $S(X) \coloneqq \sum\limits_{i=1}^{n} X_i$) – достаточная статистика.

2-й шаг. Проверим полноту $S(X)$. Так как $S(X)$ является суммой n независимых одинаково распределённых биномиальных величин, то $S(X) \sim \mathrm{Bin}(n, \theta)$.

$$\begin{aligned}
&E_\theta \varphi(S(X)) = 0 \qquad \forall \theta \in (0,1);\\
&\sum_{k=0}^{n} \varphi(k) C_n^k \theta^k (1-\theta)^{n-k} = 0 \qquad \forall \theta \in (0,1);\\
&\sum_{k=0}^{n} \varphi(k) C_n^k \left(\frac{\theta}{1-\theta}\right)^k = 0 \qquad \forall \theta \in (0,1);\\
&\text{Замена переменной:} \quad \frac{\theta}{1-\theta} \eqqcolon z \in (0, +\infty);\\
&\sum_{k=0}^{n} \varphi(k) C_n^k z^k = 0 \qquad \forall z > 0.
\end{aligned}$$

Если хотя бы какой-то коэффициент перед z^k в сумме $\sum\limits_{k=0}^{n} \varphi(k) C_n^k z^k$ не равен 0, то эта сумма является ненулевым многочленом степени не больше n. Тогда эта сумма имеет не более n нулей. Но она равна 0 при всех значениях $z > 0$. Значит, так как $C_n^k \neq 0\ \ \forall k = \overline{0, n}$, то $\varphi(k) = 0\ \ \forall k = \overline{0, n}$. Тогда $\varphi(S(X)) = 0$.

3-й шаг. Найдём измеримую функцию $\varphi(S(X))$ от полной достаточной статистики $S(X)$, несмещённо оценивающую $\tau(\theta)$. Можно решать уравнение несмещённости:

$$E_\theta \varphi(S(X)) = \tau(\theta) \qquad \forall \theta \in \Theta.$$

Но мы вместо этого используем разумный подбор. Так как любое взаимно однозначное преобразование достаточную полную статистику переводит в достаточную полную статистику, то вместо $S(X)$ можно взять $\overline{X}$.

$$E_\theta\varphi\left(\overline{X}\right) = \tau(\theta) = D_\theta X_1 = \theta - \theta^2;$$

$$E_\theta\overline{X} = \theta; \qquad E_\theta\overline{X}^2 = D_\theta\overline{X} + E_\theta^2\overline{X} = \frac{D_\theta X_1}{n} + \theta^2 = \frac{\theta - \theta^2}{n} + \theta^2;$$

$$E_\theta\left(\overline{X} - \overline{X}^2\right) = \left(\theta - \theta^2\right) - \frac{(\theta - \theta^2)}{n} = \left(1 - \frac{1}{n}\right)(\theta - \theta^2) = \frac{n-1}{n}\cdot\tau(\theta);$$

$$E_\theta\left(\frac{n}{n-1}\left(\overline{X} - \overline{X}^2\right)\right) = \tau(\theta) \qquad \forall\theta\in\Theta = (0,1).$$

Таким образом, $\hat{\tau}(X) = \dfrac{n}{n-1}\overline{X}\left(1 - \overline{X}\right)$ – оптимальная оценка для $\tau(\theta) = \theta(1-\theta)$. □

Задача 11.2. Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim \mathrm{Bin}(1,\theta)$, где параметр $\theta \in (0,1)$. Какие функции $\tau(\theta)$ допускают оптимальное оценивании? Найти соответствующие оптимальные оценки $\hat{\tau}(\theta)$ для $\tau(\theta)$.

Решение:

Так как оптимальные оценки являются несмещёнными, то найдём сначала несмещённые оценки. Пусть $\varphi(X)$ – измеримая функция от выборки.

$$E_\theta\varphi(X) = \sum_{x\in\{0,1\}^n}\varphi(x)\cdot\theta^{\sum\limits_{i=1}^n x_i}(1-\theta)^{n-\sum\limits_{i=1}^n x_i}.$$

Так как $0 \leqslant \sum\limits_{i=1}^n x_i \leqslant n$, то выражение выше является многочленом по θ степени не выше n. Если мы сможем предъявить оптимальную оценку для $\tau(\theta) = \theta^m$, где $1 \leqslant m \leqslant n$ (если $m = 0$, то τ – константа, которая оценивается тривиально), то многочлен $E_\theta\varphi(X)$ тоже будет оптимальной оценкой как линейная комбинация оптимальных оценок.

Из шагов 1 и 2 решения предыдущей задачи (11.1) получаем, что $S(X) = \sum\limits_{i=1}^n X_i$ является полной достаточной статистикой.

3-й шаг. Решим уравнение несмещённости:

$$E_\theta\varphi(S(X)) = \theta^m \qquad \forall\theta\in(0,1);$$

$$\sum_{k=0}^n \varphi(k)C_n^k\theta^k(1-\theta)^{n-k} = \theta^m \qquad \forall\theta\in(0,1);$$

$$\sum_{k=0}^n \varphi(k)C_n^k\left(\frac{\theta}{1-\theta}\right)^k = \frac{\theta^m}{(1-\theta)^n} \qquad \forall\theta\in(0,1);$$

Замена переменной: $\dfrac{\theta}{1-\theta} =: z \in (0, +\infty); \qquad z + 1 = \dfrac{1}{1-\theta};$

$$\sum_{k=0}^{n} \varphi(k) C_n^k z^k = z^m (z+1)^{n-m} \qquad \forall z > 0;$$

$$\sum_{k=0}^{n} \varphi(k) C_n^k z^k = z^m \sum_{i=0}^{n-m} C_{n-m}^i z^i \qquad \forall z > 0;$$

Замена индекса суммирования: $k = m + i$;

$$\sum_{k=0}^{n} \varphi(k) C_n^k z^k = \sum_{k=m}^{n} C_{n-m}^{k-m} z^k \qquad \forall z > 0.$$

Так как два многочлена равны при всех значениях $z > 0$, то равны их коэффициенты:

$$\varphi(k) = 0 \quad \text{при} \quad k < m;$$

$$\varphi(k) = \frac{C_{n-m}^{k-m}}{C_n^k} = \frac{(n-m)!k!(n-k)!}{(k-m)!(n-k)!n!} = \frac{k!/(k-m)!}{n!/(n-m)!} = \frac{(k)_m}{(n)_m} \quad \text{при} \quad m \leqslant k \leqslant n,$$

где мы ввели обозначение $(b)_a := \dfrac{b!}{(b-a)!} = b(b-1) \cdot \ldots \cdot (b-a+1)$.

Таким образом, $\hat{\tau}(\theta) = \dfrac{\left(\sum\limits_{i=1}^{n} X_i\right)_m}{(n)_m}$ – оптимальная оценка для $\tau(\theta) = \theta^m$, где $1 \leqslant m \leqslant n$. □

Достаточные статистики и оптимальные оценки для распределений экспоненциального типа

Рассмотрим *распределение экспоненциального типа*, то есть плотность наблюдений имеет следующий вид:

$$p_\theta(X) = h(X) \cdot e^{\sum\limits_{i=1}^{k} u_i(X) \cdot a_i(\theta) + v(\theta)}, \tag{11.1}$$

где $h(X) \geqslant 0$ и параметр $\theta \in \Theta \subset \mathbb{R}^k$.

По критерию факторизации (теорема (11.1)) получаем, что функция $u(X) = (u_1(X), \ldots, u_k(X))^T$ является достаточной статистикой. Также введём функцию $a(\theta) = (a_1(\theta), \ldots, a_k(\theta))$.

Теорема 11.3. *Пусть $\{p_\theta(X),\ \theta \in \Theta \subseteq \mathbb{R}^k\}$ – семейство распределений экспоненциального типа (11.1). Тогда если $a(\theta)$ зачёркивает k-мерный шар, то $u(X)$ является полной достаточной статистикой для θ.*

Задача 11.3. Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim \mathcal{N}(\theta, 1)$, где параметр $\theta \in \in \mathbb{R}$. Найти оптимальную оценку для $\tau(\theta) = P_\theta(X_1 \leqslant x_0)$ – функции распределения в некоторой точке x_0.

Решение:

1-й и 2-й шаги. Пусть $\Phi(X)$ – функция распределения стандартной нормальной величины $\mathcal{N}(0,1)$. Тогда $\tau(\theta) = \Phi(x_0 - \theta)$.

$$p(X,\theta) = \frac{1}{\left(\sqrt{2\pi}\right)^n} e^{-\frac{1}{2}\sum\limits_{i=1}^{n}(X_i-\theta)^2} = \frac{1}{\left(\sqrt{2\pi}\right)^n} e^{-\frac{n}{2}\overline{X^2}+n\overline{X}\theta-\frac{1}{2}n\theta^2} = h(X)\cdot e^{u(X)a(\theta)+v(\theta)},$$

где $h(X) = \dfrac{1}{\left(\sqrt{2\pi}\right)^n} e^{-\frac{n}{2}\overline{X^2}}$, $u(X) = n\overline{X}$, $a(\theta) = \theta$, $v(\theta) = -\dfrac{1}{2}n\theta^2$. При этом $a(\Theta) = \mathbb{R}$ содержит одномерный шар (интервал). Тогда по теореме (11.3) получаем, что $\overline{X}$ (или $n\overline{X}$) является полной достаточной статистикой.

3-й шаг. Найдём измеримую функцию $\varphi(\overline{X})$, несмещённо оценивающую $\tau(\theta)$, то есть $E_\theta\varphi(\overline{X}) = \tau(\theta)$. Пусть $\tilde{\tau}(X)$ – несмещённая оценка для $\tau(\theta)$. Тогда по теореме Колмогорова-Блэкуэлла-Рао (11.2) получаем, что $\varphi(\overline{X}) = E_\theta\left(\tilde{\tau}(X)\,|\,\overline{X}\right)$. Так как $\tau(\theta) = E_\theta I(X_1 \leqslant x_0)$, то $\tilde{\tau}(X) = I(X_1 \leqslant x_0)$. Таким образом,

$$\tau^*(X) = E(I(X_1 \leqslant x_0)\,|\,\overline{X})$$

– оптимальная оценка для $\tau(\theta)$.

Подберём константу c так, чтобы случайные величины $X_1 + c\overline{X}$ и $\overline{X}$ были независимы. Вектор

$$\begin{pmatrix} X_1 + c\overline{X} \\ \overline{X} \end{pmatrix} = \begin{pmatrix} 1 + \frac{c}{n} & \frac{c}{n} & \ldots & \frac{c}{n} \\ \frac{1}{n} & \frac{1}{n} & \ldots & \frac{1}{n} \end{pmatrix} \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

является гауссовским как линейная гауссовского вектора X. Тогда независимость $X_1 + c\overline{X}$ и $\overline{X}$ равносильна некоррелированности.

$$0 = \mathrm{Cov}(X_1 + c\overline{X}, \overline{X}) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{Cov}(X_1, X_i) + c\cdot D\overline{X} = \frac{1}{n}\cdot DX_1 + c\cdot\frac{1}{n} = \frac{1+}{n}.$$

Таким образом, $c = -1$, то есть случайные величины $X_1 - \overline{X}$ и $\overline{X}$ независимы. Тогда получаем:

$$E(I(X_1 \leqslant x_0)\,|\,\overline{X} = x) = E(I(X_1 - \overline{X} \leqslant x_0 - x)\,|\,\overline{X} = x) = P(X_1 - \overline{X} \leqslant x_0 - x) =$$

$$X_1 - \overline{X} \sum \mathcal{N}\left(0,\, 1 + \frac{1}{n} - 2\cdot\frac{1}{n}\right) = \mathcal{N}\left(0,\, \frac{n-1}{n}\right);$$

$$= P\left(\left(X_1 - \overline{X}\right)\sqrt{\frac{n}{n-1}} \leqslant (x_0 - x)\sqrt{\frac{n}{n-1}}\right) = \Phi\left((x_0 - x)\sqrt{\frac{n}{n-1}}\right).$$

$$\tau^*(X) = E(I(X_1 \leqslant x_0) \,|\, \overline{X}) = \Phi\left(\sqrt{\frac{n}{n-1}}\,(x_0 - \overline{X})\right).$$

□

Достаточные статистики и оптимальные оценки для нерегулярных распределений

Задача 11.4. Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim \mathrm{Exp}(1)+\theta$, то есть плотность одного наблюдения из выборки имеет вид $p_{X_1}(x,\theta) = e^{-x+\theta} I(x > \theta)$, где параметр $\theta \in \mathbb{R}$. Найти оптимальную оценку для θ.

Решение:

1-й шаг. Найдём достаточную статистику по критерию факторизации (теорема (11.1)).

$$p(X,\theta) = e^{-n\overline{X}+n\theta} I(X_{(1)} > \theta) = h(X) \cdot g(\theta,\, S(X)),$$

где $h(X) = e^{-n\overline{X}}$, $g(\theta,\, S(X)) = e^{n\theta} I(X_{(1)} > \theta)$. Таким образом, $S(X) = X_{(1)}$ – достаточная статистика.

2-й шаг. Проверим $S(X)$ на полноту.

$$\begin{aligned}
&p_{X_{(1)}}(x,\theta) = n p_{X_1}(x) \left(1 - F_{X_1}(x)\right)^{n-1} = n e^{-x+\theta} \left(e^{-x+\theta}\right)^{n-1} I(x > \theta) = \\
&= n e^{-nx+n\theta} I(x > \theta); \\
&E_\theta \varphi(S(X)) = 0 \qquad \forall \theta \in \mathbb{R}; \\
&\int\limits_{\theta}^{+\infty} \varphi(x) \cdot n e^{-nx+n\theta}\, dx = 0 \quad \Big| : e^{n\theta}; \\
&\int\limits_{\theta}^{+\infty} \varphi(x) \cdot n e^{-nx}\, dx = 0 \quad \Big| \ \frac{d}{dx} \quad \Big| : \left(-n e^{-n\theta}\right); \\
&\varphi(\theta) = 0 \quad \text{для почти всех} \quad \theta \in \mathbb{R}; \\
&P_\theta(\varphi(X_{(1)}) \neq 0) = \int\limits_{x:\ \varphi(x) \neq 0} p_{X_{(1)}}(x,\theta)\, dx = 0; \\
&\varphi(X_{(1)}) \underset{P_\theta\text{-п. н.}}{=} 0 \qquad \forall \theta \in \mathbb{R}.
\end{aligned}$$

3-й шаг. Найдём измеримую функцию $\varphi(X_{(1)})$, несмещённо оценивающую θ, то есть $E_\theta \varphi(X_{(1)}) = \theta \quad \forall \theta \in \mathbb{R}$.

$$E_\theta \varphi(X_{(1)}) = \int\limits_{\theta}^{+\infty} \varphi(x) \cdot n e^{-nx+n\theta}\, dx;$$

$$\int\limits_{\theta}^{+\infty} \varphi(x) \cdot ne^{-nx}\, dx = \theta e^{-n\theta} \quad \Big| \frac{d}{dx};$$

$$-\varphi(\theta) ne^{-n\theta} = e^{-n\theta} - n\theta e^{-n\theta};$$

$$\varphi(\theta) = \theta - \frac{1}{n} \quad \text{для почти всех} \quad \theta \in \mathbb{R}.$$

Таким образом, $X_{(1)} - \dfrac{1}{n}$ – оптимальная оценка для θ.

3-й шаг (альтернативный способ). Вместе решения уравнения несмещённости можно воспользоваться разумным перебором.

Так как $p_{X_{(1)}}(x, \theta) = ne^{-n(x-\theta)} I(x > \theta)$, то $X_{(1)} \sim \mathrm{Exp}(n) + \theta$. Тогда получаем:

$$EX_{(1)} = E_\theta(\mathrm{Exp}(n)) + \theta = \frac{1}{n} + \theta.$$

Значит, $X_{(1)} - \dfrac{1}{n}$ – оптимальная оценка для θ. □

Вообще говоря, если полная достаточная статистика $\hat{\theta}$ является оптимальной оценкой для θ и требуется найти оптимальную оценку для $\tau(\theta)$, то бывает полезно вычислить $E_\theta \tau(\hat{\theta})$ и попытаться скорректировать полученный результат так, чтобы получить оптимальную оценку для $\tau(\theta)$.

Семинар 12

Доверительные интервалы

Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X \sim p_\theta$, где параметр $\theta \in \Theta$.

Определение 12.1. $I(X) = (T_1(X), T_2(X)) \subset \Theta$ – *доверительный интервал уровня доверия* $1-\alpha$ *для* θ, где $\alpha \in (0,1)$, $T_1(X)$ и $T_2(X)$ – некоторые статистики, если

$$P_\theta(\theta \in I(X)) \geqslant 1-\alpha \qquad \forall \theta \in \Theta.$$

Задача 12.1. Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim \mathrm{Exp}(1)+\theta$, то есть плотность одного наблюдения из выборки имеет вид $p_{X_1}(x,\theta) = e^{-x+\theta} I(x > \theta)$, где параметр $\theta \in \mathbb{R}$. $X_{(1)} \sim \mathrm{Exp}(n) + \theta$ – полная достаточная статистика, $X_{(1)} - \dfrac{1}{n}$ – оптимальная оценка для θ. Построить доверительный интервал для θ уровня $1-\alpha$.

Решение:

Так как $X_{(1)} \sim \mathrm{Exp}(n) + \theta$, то $n(X_{(1)} - \theta) \sim \mathrm{Exp}(1)$. Обозначим $F(x) = (1 - e^{-x}) I(x > 0)$ функцию распределения стандартной экспоненциальной величины, тогда получаем:

$$P_\theta(g_1 < n(X_{(1)} - \theta) < g_2) = F(g_2) - F(g_1) = 1 - \alpha;$$
$$P_\theta\left(X_{(1)} - \frac{g_2}{n} < \theta < X_{(1)} - \frac{g_1}{n}\right) = 1-\alpha.$$

Получили доверительный интервал $X_{(1)} - \dfrac{g_2}{n} < \theta < X_{(1)} - \dfrac{g_1}{n}$ уровня доверия $1-\alpha$, который имеет длину $l_n(g_1, g_2) = \dfrac{g_2 - g_1}{n}$. Чтобы точнее определить параметр θ, минимизируем длину $l_n(g_1, g_2)$.

Пусть $F(g_1) = \alpha_1$, где $\alpha_1 \in [0, \alpha]$, тогда $F(g_2) = 1 - \alpha + \alpha_1$. Получаем:

$$F(g_1) = \alpha_1, \qquad 1 - e^{-g_1} = \alpha_1, \qquad g_1 = -\ln(1-\alpha_1);$$
$$F(g_2) = 1 - \alpha + \alpha_1, \qquad 1 - e^{-g_1} = 1 - \alpha + \alpha_1, \qquad g_1 = -\ln(\alpha - \alpha_1);$$
$$l_n = \frac{1}{n}\left(\ln(1-\alpha_1) - \ln(\alpha - \alpha_1)\right) = \frac{1}{n}\ln\frac{1-\alpha_1}{\alpha-\alpha_1} = \frac{1}{n}\ln\left(1 + \frac{1-\alpha}{\alpha-\alpha_1}\right).$$

Надо минимизировать l_n при $\alpha_1 \in [0, \alpha]$. Наименьшее значение достигается при $\alpha_1 = 0$ и оно равно $l_n = -\dfrac{\ln\alpha}{n}$. При $\alpha_1 = 0$ получаем: $g_1 = 0$, $g_2 = -\ln\alpha$. Таким образом, имеем:

$$P_\theta\left(X_{(1)} - \frac{\ln\alpha}{n} < \theta < X_{(1)}\right) = 1-\alpha,$$

то есть $X_{(1)} - \dfrac{\ln\alpha}{n} < \theta < X_{(1)}$ – доверительный интервал уровня доверия $1-\alpha$ минимальной длины. □

Замечание 12.1. Если мы хотим получить значение вероятности как можно больше, тогда надо брать значение α как можно меньше, но в таком случае длина доверительного интервала стремится к $+\infty$ (при фиксированном значении n), что неинформативно. Обычно берут $\alpha = 0{,}05$ или $0{,}01$.

Замечание 12.2. Если α фиксировано, а $n \to +\infty$, то длина доверительного интервала стремится к 0 со скоростью $O\left(\frac{1}{n}\right)$.

Метод центральной статистики

Возьмём оптимальную оценку и сделаем некоторое преобразование так, чтобы распределение преобразованной случайной величины $G(X,\theta)$ не зависело от неизвестного параметра θ. Потребуем, чтобы функция распределения этой случайной величины F_G была непрерывна. Это нужно для того, чтобы можно было строить точные доверительные интервалы, то есть для любого значения α можно было бы подобрать соответствующие значения g_1 и g_2.

Также потребуем, чтобы функция $G(X,\theta)$ была непрерывна и монотонна по θ.

Определение 12.2. Случайная величина $G(X,\theta)$ называется *центральной статистикой* (или *центральной величиной*).

Запишем доверительный интервал уровня доверия $1-\alpha$:

$$P_\theta(g_1 < G(X,\theta) < g_2) = F_G(g_2) - F_G(g_1) = 1 - \alpha.$$

Так распределение $G(X,\theta)$ не зависит от параметра θ, то мы можем подобрать значения g_1 и g_2. Так как $G(X,\theta)$ непрерывна и монотонна по θ, то получаем:

$$P_\theta(g_1 < G(X,\theta) < g_2) = P_\theta(T_1(X) < \theta < T_2(X)),$$

где $T_1(X)$ – значение параметра θ, при котором $G(X,\theta) = g_1$, а $T_2(X)$ – значение параметра θ, при котором $G(X,\theta) = g_2$.

Значения g_1 и g_2 можно подобрать из условия

$$\begin{cases} l_n = T_2(X,g_2) - T_1(X,g_1) \to \min\limits_{g_1,\,g_2} \\ F_G(g_2) - F_G(g_1) = 1-\alpha \end{cases}.$$

Но такая задача поиска минимума может не решаться аналитически. В таком случае выбирают значения g_1 и g_2 из условия симметрии: $F_G(g_2) = 1 - \frac{\alpha}{2}$, $F_G(g_1) = \frac{\alpha}{2}$. Тогда g_1 и g_2 являются квантилями уровней $\frac{\alpha}{2}$ и $1-\frac{\alpha}{2}$ соответственно:

$$\begin{cases} g_1 = F_G^{-1}\left(\frac{\alpha}{2}\right) \\ g_2 = F_G^{-1}\left(1-\frac{\alpha}{2}\right) \end{cases}.$$

При таком выборе g_1 и g_2 доверительный интервал называется *центральным*.

Доверительные интервалы для параметров гауссовской выборки

Задача 12.2. Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim \mathcal{N}(a, \sigma^2)$. Построить доверительный интервал уровня $1-\alpha$ для a, считая, что значение σ^2 известно.

Решение:

$\overline{X}$ – оптимальная оценка a. Так как $\overline{X} \sim \mathcal{N}\left(a, \dfrac{\sigma^2}{n}\right)$, то $\dfrac{\sqrt{n}\left(\overline{X}-a\right)}{\sigma} \sim \mathcal{N}(0,1)$.

Используя обозначение $\Phi(x)$ в качестве функции распределения стандартной нормальной величины, получаем для доверительного интервала:

$$P_\theta = \left(g_1 < \frac{\sqrt{n}\left(\overline{X}-a\right)}{\sigma} < g_2\right) = \Phi(g_2) - \Phi(g_1) = 1-\alpha; \quad P_\theta\left(\overline{X} - g_2\frac{\sigma}{\sqrt{n}} < a < \overline{X} - g_1\frac{\sigma}{\sqrt{n}}\right).$$

Тогда длина доверительного интервала $l_n = \dfrac{\sigma}{\sqrt{n}}(g_2 - g_1)$. Найдём следующий условный минимум:

$$\begin{cases} l_n = \dfrac{\sigma}{\sqrt{n}}(g_2-g_1) \to \min\limits_{g_1,\, g_2} \\ \Phi(g_2) - \Phi(g_1) = 1-\alpha \end{cases}.$$

Воспользуемся функцией Лагранжа для минимизации значения $g_2 - g_1$. Учитывая, что $\Phi'(x) = \varphi(x)$ – плотность распределения стандартной гауссовской величины, получаем:

$$L = g_2 - g_1 + \lambda(\Phi(g_2) - \Phi(g_1) - 1 + \alpha);$$

$$\begin{cases} L'_{g_1} = -1 - \lambda\varphi(g_1) = 0 \\ L'_{g_2} = -1 + \lambda\varphi(g_2) = 0 \\ L'_{\lambda} = \Phi(g_2) - \Phi(g_1) = 1-\alpha \end{cases}.$$

Отсюда получаем, что $\varphi(g_1) = \varphi(g_2)$. Так как $1-\alpha > 0$, то $g_1 \neq g_2$. Тогда в силу симметричности $\varphi(x)$ относительно $x = 0$ получаем, что $-g_1 = g_2$. Значит, так как $\Phi(g_2) - \Phi(g_1) = 1-\alpha$, то

$$g_1 = \Phi^{-1}\left(\frac{\alpha}{2}\right) \equiv Z_{\alpha/2};$$

$$g_2 = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \equiv Z_{1-\alpha/2}.$$

Так как $Z_{1-\alpha/2} = -Z_{\alpha/2} > 0$, получаем:

$$P_\theta\left(\theta \in \left(\overline{X} - \frac{\sigma}{\sqrt{n}}Z_{1-\alpha/2},\, \overline{X} + \frac{\sigma}{\sqrt{n}}Z_{1-\alpha/2}\right)\right) = 1-\alpha.$$

□

Замечание 12.3. Для $\alpha = 0{,}05$ имеем $Z_{1-\alpha/2} \approx 1{,}96$, а для $\alpha = 0{,}01$ имеем $Z_{1-\alpha/2} \approx$

$\approx 2{,}58.$

Теорема 12.1. *Пусть* $X = (X_1, \ldots, X_n)$ *– выборка из* $\mathcal{N}(a, \sigma^2)$*, где* $a \in \mathbb{R}$ *и* $\sigma > 0$ *– неизвестные параметры. Тогда* $\overline{X}$ *и* $S^2 = \dfrac{1}{n-1}\sum\limits_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ *– оптимальные оценки для* a *и* σ^2 *соответственно, причём* $\overline{X} \sim \mathcal{N}\left(a, \dfrac{\sigma^2}{n}\right)$ *не зависит от* $S^2 \sim \dfrac{\sigma^2}{n-1}\chi^2(n-1)$.

Напоминание: $\chi^2(k) \overset{d}{=} \xi_1^2 + \ldots + \xi_k^2$, где $\{\xi_i\}$ – независимые одинаково распределённые стандартные гауссовские величины.

Задача 12.3. Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim \mathcal{N}(a, \sigma^2)$. Построить доверительный интервал уровня $1 - \alpha$ для a, считая, что значение σ^2 неизвестно.

Решение:

Так как $\dfrac{\sqrt{n}\left(\overline{X} - a\right)}{\sigma} \sim \mathcal{N}(0,1)$, то

$$\frac{\sqrt{n}(\overline{X} - a)}{S} = \frac{\sqrt{n}(\overline{X} - a)/\sigma}{\sqrt{S^2/\sigma^2}}, \tag{12.1}$$

причём $\sqrt{n}(\overline{X}-a)/\sigma \sim \mathcal{N}(0,1)$ и $S^2/\sigma^2 \sim \dfrac{1}{n-1}\chi^2(n-1)$. Так как $\overline{X}$ не зависит от S^2, то и любая борелевская функция от $\overline{X}$ не зависит от любой борелевской функции от S^2. Значит, числитель и знаменатель в выражении (12.1) независимы. Поэтому распределение функции $\dfrac{\sqrt{n}(\overline{X} - a)}{S}$ не зависит от параметров и полностью известно, а значит, подходит в качестве центральной статистики. Это распределение называется распределение Стьюдента с числом степеней свободы $n - 1$ (записывается $t(n-1)$).

Определение 12.3. $t(k) \overset{d}{=} \dfrac{\xi_0}{\sqrt{\dfrac{1}{k}(x_1^2 + \ldots + \xi_k^2)}}$ *– распределение Стьюдента с числом степеней свободы* k, где $\{\xi_i\}$ – независимые одинаково распределённые стандартные гауссовские величины.

Распределение Стьюдента симметрично относительно 0, поэтому получаем:

$$P_\theta\left(-t_{n-1,\ 1-\alpha/2} \leqslant \frac{\sqrt{n}(\overline{X} - a)}{S} \leqslant t_{n-1,\ 1-\alpha/2}\right) = 1 - \alpha,$$

где $t_{k,\alpha}$ – α-квантиль распределения $t(k)$.

$$P_\theta\left(a \in \left(\overline{X} - \frac{S}{\sqrt{n}}t_{n-1,\ 1-\alpha/2},\ \overline{X} + \frac{S}{\sqrt{n}}t_{n-1,\ 1-\alpha/2}\right)\right) = 1 - \alpha.$$

$\square$

Задача 12.4. Доказать, что $t(k)$ – симметричное относительно 0 распределение, то есть $t(k) \stackrel{d}{=} -t(k)$.

Решение:

$\{\xi_i\}$ – независимые одинаково распределённые стандартные гауссовские величины, поэтому вектор $\xi \equiv \begin{pmatrix} \xi_0 \\ \vdots \\ \xi_k \end{pmatrix}$ имеет симметричное относительно 0 распределение, так как $\varphi_\xi(\lambda) = e^{-\dfrac{|\lambda|^2}{2}}$, где $\lambda \in \mathbb{R}^{k+1}$. Так как $\xi \stackrel{d}{=} -\xi$, то $h(\xi) \stackrel{d}{=} h(-\xi)$, где h – борелевская функция. Возьмём $h(\xi_0, \ldots, \xi_k) = \dfrac{\xi_0}{\sqrt{\frac{1}{k}(\xi_1^2 + \ldots + \xi_k^2)}}$. Тогда $t(k) \stackrel{d}{=} -t(k)$.

$\square$

Так как при $k \to +\infty$ имеем: $\sqrt{\frac{1}{k}\chi^2(k)} = \sqrt{\frac{1}{k}(\xi_1^2 + \ldots + \xi_k^2)} \xrightarrow{P} \sqrt{E\xi_1^2} = 1$ и $\xi_0 \xrightarrow{d} \mathcal{N}(0,1)$, то по лемме Слуцкого (1.1) получаем: $t(k) \stackrel{d}{=} \dfrac{\xi_0}{\sqrt{\frac{1}{k}\chi^2(k)}} \xrightarrow{d} \mathcal{N}(0,1)$.

Тогда при больших значениях k для квантилей получаем: $t_{\alpha,k} \approx Z_\alpha$.

Если $k = 1$, то $t(k) \stackrel{d}{=} \dfrac{\xi_0}{|\xi_1|} \sim \mathrm{Cauchy}(0,1)$.

Задача 12.5. Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim \mathcal{N}(a, \sigma^2)$, где $a \in \mathbb{R}$ и $\sigma > 0$ – неизвестные параметры. Построить доверительный интервал уровня $1-\alpha$ для σ^2.

Решение:

$S^2 \coloneqq \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ – оптимальная оценка для σ^2, причём $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$. Эта случайная величина будем центральной статистикой, так как она монотонна по σ^2 и её распределение не зависит от неизвестных параметров.

Так как невозможно аналитически указать квантили, чтобы длина доверительного интервала была наименьшей, то построим центральный доверительный интервал:

$$P_{a,\sigma^2}\left(\chi_{\alpha/2,\,n-1} < \frac{(n-1)S^2}{\sigma^2} < \chi_{1-\alpha/2,\,n-1}\right) = 1-\alpha;$$
$$P_{a,\sigma^2}\left(\frac{n-1}{\chi_{1-\alpha/2,\,n-1}}S^2 < \sigma^2 < \frac{n-1}{\chi_{\alpha/2,\,n-1}}S^2\right) = 1-\alpha.$$

$\square$

Замечание 12.4. Можно построить доверительный интервал и для σ:

$$P_{a,\sigma^2}\left(\sqrt{\frac{n-1}{\chi_{1-\alpha/2,\,n-1}}S^2} < \sqrt{\sigma^2} < \sqrt{\frac{n-1}{\chi_{\alpha/2,\,n-1}}S^2}\right) = 1-\alpha.$$

Оптимальный доверительный интервал

Вернёмся к обсуждению задачи (12.1). У нас была выборка $X = (X_1, \ldots, X_n)$, $X_1 \sim \mathrm{Exp}(1) + \theta$, где $\theta \in \mathbb{R}$. Мы получили доверительный интервал $\left(X_{(1)} - \dfrac{\ln\alpha}{n},\, X_{(1)}\right)$. Его длина имеет порядок $O\left(\dfrac{1}{n}\right)$. Мы выбрали этот интервал из условия минимальной длины, но мы использовали только статистику $X_{(1)}$. А могли бы использовать и другие статистики. Например, $n\left(\overline{X} - \theta\right) = \sum\limits_{i=1}^{n}(X_i - \theta) \sim \Gamma(1, n)$. Чтобы понять, какой из двух доверительных интервалов лучше, воспользуемся центральной предельной теоремой:

$$\sqrt{n}(\overline{X} - \theta - 1) \xrightarrow[n\to+\infty]{d} \mathcal{N}(0,1).$$

Тогда длина второго доверительного интервала будет иметь порядок $O\left(\dfrac{1}{\sqrt{n}}\right)$. Значит, для больших значений n первый доверительный интервал лучше, так как будет иметь меньшую длину.

Но вообще говоря, длина интервала не является удачной характеристикой для того, чтобы определить оптимальный доверительный интервал.

Определение 12.4. Доверительный интервал $I(X)$ уровня $1-\alpha$ для θ называется *несмещённым*, если

$$P_\theta(\theta' \in I(X)) \leqslant 1-\alpha \qquad \forall \theta, \theta'\colon\ \theta \neq \theta'.$$

Определение 12.5. Несмещённый доверительный интервал $I(X)$ уровня $1-\alpha$ для θ называется *наиболее точным*, то есть оптимальным, если он минимизирует

$$P_\theta(\theta' \in I(X)) \qquad \forall \theta, \theta'\colon\ \theta \neq \theta'$$

среди всех несмещённых интервалов уровня $1-\alpha$.

Семинар 13

Асимптотические доверительные интервалы

Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X \sim P_\theta$, где параметр $\theta \in \Theta \subset \mathbb{R}^1$.

Определение 13.1. $I_n(X) = (T_{1,n}(X), T_{2,n}(X)) \subset \Theta$ называется *асимптотическим доверительным интервалом для θ уровня* $1-\alpha$, где $\alpha \in (0,1)$, если

$$P_\theta(\theta \in I_n(X)) \underset{n\to+\infty}{\to} 1-\alpha \qquad \forall \theta \in \Theta.$$

Для построения асимптотических доверительных интервалов будем использовать $\sqrt{n}$-асимптотически гауссовские оценки:

$$\sqrt{n}\left(\hat{\theta}_n(X)-\theta\right) \xrightarrow{d} \mathcal{N}\left(0,\, \sigma^2(\theta)\right) \qquad \forall \theta \in \Theta, \quad \text{где} \quad \sigma^2(\theta) > 0.$$

Тогда получаем:

$$\frac{\sqrt{n}\left(\hat{\theta}_n(X)-\theta\right)}{\sigma(\theta)} \xrightarrow{d} \mathcal{N}(0,1) \qquad \forall \theta \in \Theta;$$

$$P_\theta\left(Z_{\alpha/2} < \frac{\sqrt{n}\left(\hat{\theta}_n(X)-\theta\right)}{\sigma(\theta)} < -Z_{\alpha/2}\right) \to 1-\alpha.$$

Если выражение $Z_{\alpha/2} < \dfrac{\sqrt{n}\left(\hat{\theta}_n(X)-\theta\right)}{\sigma(\theta)} < -Z_{\alpha/2}$ удастся записать в виде неравенства для θ, то мы получим асимптотический доверительный интервал.

Задача 13.1. Пусть $X = (X_1, \ldots, X_n)$ – выборка, где $X_1 \sim \mathrm{Pois}(\theta)$, где параметр $\theta > 0$. Построить асимптотический доверительный интервал для θ уровня $1-\alpha$.

Решение:

$$\frac{\sqrt{n}\left(\overline{X}-\theta\right)}{\sqrt{\theta}} \underset{n\to+\infty}{\xrightarrow{d}} \mathcal{N}(0,1) \qquad \forall \theta > 0;$$

$$P_\theta\left(Z_{\alpha/2} < \frac{\sqrt{n}\left(\overline{X}-\theta\right)}{\sqrt{\theta}} < -Z_{\alpha/2}\right) \to 1-\alpha;$$

$$\begin{cases} \theta - \dfrac{Z_{\alpha/2}}{\sqrt{n}}\sqrt{\theta} - \overline{X} > 0 \\ \theta + \dfrac{Z_{\alpha/2}}{\sqrt{n}}\sqrt{\theta} - \overline{X} < 0 \end{cases}.$$

Для обоих квадратных трёхчленов относительно $\sqrt{\theta}$ получаем: $\dfrac{D}{4} = \overline{X} + \dfrac{Z^2_{\alpha/2}}{4n}$.

$$\begin{cases} \sqrt{\theta} > \sqrt{\dfrac{D}{4}} + \dfrac{Z_{\alpha/2}}{2\sqrt{n}} \\ \sqrt{\theta} < \sqrt{\dfrac{D}{4}} - \dfrac{Z_{\alpha/2}}{2\sqrt{n}} \end{cases};$$

$$\theta \in \left(\left(\sqrt{\overline{X} + \frac{Z_{\alpha/2}}{4n}} + \frac{Z_{\alpha/2}}{2\sqrt{n}} \right)^2; \ \left(\sqrt{\overline{X} + \frac{Z_{\alpha/2}}{4n}} - \frac{Z_{\alpha/2}}{2\sqrt{n}} \right)^2 \right).$$

$\square$

Использование состоятельной оценки для дисперсии

Так как вычисления получаются громоздкими и асимптотическая дисперсия не всегда может быть такой простой, чтобы можно было выразить интервал для θ в явном виде, то необходим ещё способ получения доверительного интервала. Заменим дисперсию известной состоятельной оценкой.

Пусть $\sigma(\theta)$ – непрерывная функция по θ. Так как $\hat{\theta}_n(X)$ – $\sqrt{n}$-асимптотически гауссовская оценка для θ с асимптотической дисперсией $\sigma^2(\theta)$, то $\hat{\theta}_n \underset{n\to+\infty}{\xrightarrow{P}} \theta \quad \forall\theta \in \Theta$. Тогда по теореме (1.1) о наследовании сходимости получаем: $\sigma\left(\hat{\theta}_n\right) \xrightarrow{P} \sigma(\theta)$.

$$\frac{\sqrt{n}\left(\hat{\theta}_n(X) - \theta\right)}{\sigma(\hat{\theta}_n)} = \frac{\sqrt{n}\left(\hat{\theta}_n(X) - \theta\right)/\sigma(\theta)}{\sigma(\hat{\theta}_n)/\sigma(\theta)} \xrightarrow{d} \mathcal{N}(0,1).$$

Отсюда получаем асимптотический доверительный интервал для θ уровня $1-\alpha$:

$$I_n(X) = \left(\hat{\theta}_n(X) + \frac{\sigma(\hat{\theta}_n)}{\sqrt{n}} Z_{\alpha/2};\ \hat{\theta}_n(X) - \frac{\sigma(\hat{\theta}_n)}{\sqrt{n}} Z_{\alpha/2} \right).$$

Задача 13.2. Пусть $X = (X_1, \ldots, X_n)$ – выборка, где $X_1 \sim \mathrm{Pois}(\theta)$, где параметр $\theta > 0$. Построить асимптотический доверительный интервал для θ уровня $1-\alpha$ на основе состоятельной оценки для σ.

Решение:

$\overline{X}$ – эффективная и асимптотически эффективная оценка для θ (то есть ОМП), тогда $\sigma(\theta) = \sqrt{\theta}$ и $\sigma(\overline{X}) = \sqrt{\overline{X}}$.

$$I_n(X) = \left(\overline{X} + \frac{\sqrt{\overline{X}}}{\sqrt{n}} Z_{\alpha/2};\ \overline{X} - \frac{\sqrt{\overline{X}}}{\sqrt{n}} Z_{\alpha/2} \right).$$

$\square$

Использование преобразования, стабилизующего дисперсию

Так как для построения асимптотических доверительных интервалов мы используем $\sqrt{n}$-асимптотически гауссовские оценки, то

$$\sqrt{n}\left(\hat{\theta}_n(X)-\theta\right) \xrightarrow{d} \mathcal{N}\left(0,\, \sigma^2(\theta)\right) \qquad \forall \theta \in \Theta, \quad \text{где} \quad \sigma^2(\theta)>0.$$

Хотим использовать преобразование $\tau(\theta)$, стабилизирующее дисперсию. Пусть $\tau(\theta)$ дифференцируемо по θ и монотонно. Тогда по дельта-методу (1.8) получаем:

$$\sqrt{n}\left(\tau(\hat{\theta}_n)-\tau(\theta)\right) \xrightarrow{d} \mathcal{N}\left(0,\, (\tau'(\theta)\sigma(\theta))^2\right).$$

Выберем $\tau(\theta)$ так, чтобы выполнялось условие $\tau'(\theta)\sigma(\theta)\equiv 1 \quad \forall \theta \in \Theta$. Тогда получаем:

$$\tau(\theta)=\int \frac{d\theta}{\sigma(\theta)}.$$

Построим доверительный интервал для $\tau(\theta)$ уровня $1-\alpha$:

$$P_\theta\left(\frac{Z_{\alpha/2}}{\sqrt{n}}<\tau(\theta)<-\frac{Z_{\alpha/2}}{\sqrt{n}}\right) \to 1-\alpha.$$

Так как $\tau(\theta)$ – монотонная функция, то этот интервал можно обратить и выразить из него интервал для θ. Таким образом мы получим асимптотический доверительный интервал для θ.

Задача 13.3. Пусть $X=(X_1,\,\ldots,X_n)$ – выборка, где $X_1\sim \mathrm{Pois}(\theta)$, где параметр $\theta>0$. Построить асимптотический доверительный интервал для θ уровня $1-\alpha$ на основе преобразования, стабилизирующего дисперсию.

Решение:

Так как $\overline{X}$ – ОМП, то

$$\sqrt{n}\left(\overline{X}-\theta\right) \xrightarrow{d} \mathcal{N}(0,\theta),$$

где $\theta=\sigma^2(\theta)>0$.

$$\tau(\theta)=\int\frac{d\theta}{\sigma(\theta)}=\int\frac{d\theta}{\sqrt{\theta}}=2\sqrt{\theta};$$

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

$$2\sqrt{\theta}\in\left(2\sqrt{\overline{X}}+\frac{Z_{\alpha/2}}{\sqrt{n}};\ 2\sqrt{\overline{X}}-\frac{Z_{\alpha/2}}{\sqrt{n}}\right);$$

$$\theta\in\left(\left(\sqrt{\overline{X}}+\frac{Z_{\alpha/2}}{2\sqrt{n}}\right)^2;\ \left(\sqrt{\overline{X}}-\frac{Z_{\alpha/2}}{2\sqrt{n}}\right)^2\right).$$

□

Задача 13.4. Пусть $\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \ \ldots, \ \begin{pmatrix} X_n \\ Y_n \end{pmatrix}$ – независимые одинаково распределённые двумерные гауссовские величины с коэффициентом корреляции $\rho = \dfrac{\mathrm{Cov}(X_1, Y_1)}{\sqrt{DX_1 \cdot DY_1}}$. Построить асимптотический доверительный интервал для ρ уровня $1-\alpha$ на основе преобразования, стабилизирующего дисперсию.

Решение:

Используем выборочные коэффициенты корреляции:

$$r_n = \frac{\sum\limits_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum\limits_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \cdot \sum\limits_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}}.$$

Для построения асимптотического доверительного интервала нужно найти асимптотическую дисперсию для выборочной корреляции. Применяя многомерный дельта-метод и многомерную центральную предельную теорему, можно получить:

$$\sqrt{n}(r_n - \rho) \xrightarrow{d} \mathcal{N}\left(0, \left(1-\rho^2\right)^2\right).$$

Таким образом, $\sigma^2(\rho) = \left(1-\rho^2\right)^2$ – асимптотическая дисперсия.

$$\tau(\rho) = \int \frac{d\rho}{\sigma(\rho)} = \int \frac{d\theta}{1-\rho^2} = \frac{1}{2}\int\left(\frac{1}{1-\rho} + \frac{1}{1+\rho}\right)d\rho = \frac{1}{2}\ln\frac{1+\rho}{1-\rho} = \operatorname{arth}\rho;$$

$$\operatorname{arth}\rho \in \left(\operatorname{arth} r_n + \frac{Z_{\alpha/2}}{\sqrt{n}};\ \operatorname{arth} r_n - \frac{Z_{\alpha/2}}{\sqrt{n}}\right);$$

$$\rho \in \left(\operatorname{th}\left(\operatorname{arth} r_n + \frac{Z_{\alpha/2}}{\sqrt{n}}\right);\ \operatorname{th}\left(\operatorname{arth} r_n - \frac{Z_{\alpha/2}}{\sqrt{n}}\right)\right).$$

□

Семинар 14

Линейная регрессия

Приведём пример. Пусть есть два грузика разных масс, которые мы хотим измерить. $X_1 = m_1 + \varepsilon_1$, $X_2 = m_2 + \varepsilon_2$ – результаты измерений масс m_1 и m_2 первого и второго грузика с погрешностями ε_1 и ε_2 соответственно. $X_3 = m_1 + m_2 + \varepsilon_3$ – результат измерения суммарной массы $m_1 + m_2$ двух грузиков с погрешностью ε_3. Пусть $\{\varepsilon_i\}$ – независимые одинаково распределённые случайные величины $\mathcal{N}(0, \sigma^2)$. Требуется оптимально оценить (m_1, m_2).

Запишем условия в матричном виде:

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} m_1 \\ m_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}.$$

Пусть $X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$ – вектор наблюдений, $Z = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}$, $\theta = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}$ – вектор неизвестных параметров, $\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{pmatrix}$ – вектор случайных ошибок. Тогда матричное уравнение можно записать в виде $X = Z\theta + \varepsilon$.

Пусть в общем случае $\theta \in \mathbb{R}^k$ – неизвестный параметр модели, Z – неслучайная известная матрица размера $n \times k$, причём $\operatorname{rank} Z = k < n$, $E\varepsilon = 0$, $D\varepsilon = \sigma^2 I_n$. Такая модель называется *моделью линейной регрессии.*

Определение 14.1. *Оценкой наименьших квадратов (ОНК)* в модели линейной регрессии называется

$$\hat{\theta}_n := \arg\min_{\theta \in \mathbb{R}^k} \|X - Z\theta\|^2.$$

Рассмотрим, что геометрически представляет из себя ОНК. Обозначим $l := EX = Z\theta$ и $Z_1, \ldots, Z_k$ – столбцы матрицы Z. Так как $\operatorname{rank} Z = k$, то столбцы $Z_1, \ldots, Z_k$ линейно независимы. Тогда $l = \theta_1 Z_1 + \ldots + \theta_k Z_k$, значит, математическое ожидание наблюдения лежит в линейном подпространстве, которое порождено системой линейно независимых векторов $Z_1, \ldots, Z_k$: $\quad l \in L\langle Z_1, \ldots, Z_k \rangle$ – известное линейное подпространство в $\mathbb{R}^n$, причём $\dim L = k < n$. Следовательно, $\hat{l} = Z\hat{\theta}_n = \operatorname{proj}_L X$, то есть $(X - Z\hat{\theta}_n)^T Z = 0$. Отсюда получаем явный вид ОНК: $\hat{\theta}_n = \left(Z^T Z\right)^{-1} Z^T X$. Но на практике обычно удобнее решать задачу $\hat{\theta}_n := \arg\min\limits_{\theta \in \mathbb{R}^k} \|X - Z\theta\|^2$.

Теорема 14.1 (Теорема Гаусса-Маркова)**.** *В модели линейной регрессии ОНК $\hat{\theta}_n = BLUE(\theta)$ – наилучшая несмещённая оценка в классе линейных несмещённых оценок для θ.*

Найдём ОНК для пример с двумя грузика, описанного выше.

Задача 14.1. $\begin{cases} X_1 = m_1 + \varepsilon_1 \\ X_2 = m_2 + \varepsilon_2 \\ X_3 = m_1 + m_2 + \varepsilon_3 \end{cases}$, где $\{\varepsilon_i\}$ – независимые одинаково распределённые случайные величины, $E\varepsilon_i = 0$, $D\varepsilon_i = \sigma^2$. Найти ОНК для $\theta = (m_1, m_2)$.

Решение:

$$\|X - Z\theta\|^2 = (X_1 - m_1)^2 + (X_2 - m_2)^2 + (X_3 - m_1 - m_2)^2 \to \min_{m_1, m_2};$$

$$\begin{cases} \dfrac{\partial}{\partial m_1}\|X - Z\theta\|^2 = -2(X_1 - m_1) - 2(X_3 - m_1 - m_2) = 0 \\ \dfrac{\partial}{\partial m_2}\|X - Z\theta\|^2 = -2(X_2 - m_2) - 2(X_3 - m_1 - m_2) = 0 \end{cases};$$

$$\begin{cases} 2m_1 + m_2 = X_1 + X_3 \\ m_1 + 2m_2 = X_2 + X_3 \end{cases};$$

$$\begin{cases} m_1 = \dfrac{1}{3}(2X_1 - X_2 + X_3) \\ m_2 = \dfrac{1}{3}(2X_2 - X_1 + X_3) \end{cases}.$$

Найденное решение является минимумом, как следует из квадратичного вида функции $\|X - Z\theta\|^2$, значит, $\begin{cases} \hat{m}_1 = \dfrac{1}{3}(2X_1 - X_2 + X_3) \\ \hat{m}_2 = \dfrac{1}{3}(2X_2 - X_1 + X_3) \end{cases}$ – ОНК для θ. □

Утверждение 14.1. *$\hat{\theta}_n = \left(Z^T Z\right)^{-1} Z^T X$ и $\hat{\sigma}_n^2 = \dfrac{1}{n-k}\|X - Z\hat{\theta}\|^2$ – несмещённые оценки для θ и σ^2 соответственно.*

Гауссовская линейная регрессия

Будем предполагать, что случайные погрешности гауссовские, тогда $X \sim \sim \mathcal{N}_n(Z\theta, \sigma^2 I_n)$, где $\operatorname{rank} Z = k < n$, $\theta \in \mathbb{R}^k$ – неизвестный параметр.

Теорема 14.2. *В гауссовской линейной регрессии:*

1) *ОНК $(\hat{\theta}_n, \hat{\sigma}_n^2)$ – оптимальные оценки для (θ, σ^2);*
2) $\hat{\theta}_n \sim \mathcal{N}_k\left(\theta, \sigma^2\left(Z^T Z\right)^{-1}\right)$, $\quad \hat{\sigma}^2 \sim \dfrac{\sigma^2}{n-k}\chi^2(n-k)$;
3) *$\hat{\theta}_n$ и $\hat{\sigma}_n^2$ – независимые случайные величины.*

Задача 14.2. Рассмотрим простую линейную регрессию: $X_i = a + bz_i + \varepsilon_i$, где $i = = \overline{1, n}$, $\theta = (a, b)$ – неизвестный параметр, $\{z_i\}$ – известные числа, $\{\varepsilon_i\}$ – независимые одинаково распределённые случайные величины $\mathcal{N}(0, \sigma^2)$.

а) Найти оптимальные оценки для a, b, σ^2 методом наименьших квадратов.
б) Исследовать эти оценки на состоятельность.
в) Построить доверительные интервалы уровня $1-\alpha$ для b и $a+bz$, где $z \in \mathbb{R}$.

Решение:

Сделаем замену переменных, упрощающую решение:

$$X_i = a + b\overline{z} + b(z_i - \overline{z}) + \varepsilon_i = c + b(z_i - \overline{z}) + \varepsilon_i,$$

где $c = a + b\overline{z}$. Тогда вместо параметра (a, b) будем рассматривать параметр (c, b), где $a = c - b\overline{z}$. В силу линейности преобразования можно искать оптимальную оценку для нового параметра.

а)
$$\sum_{i=1}^{n}(X_i - c - b(z_i - \overline{z}))^2 \to \min_{b,c};$$

$$\begin{cases} -2\sum\limits_{i=1}^{n}(X_i - c - b(z_i - \overline{z})) = 0 \\ -2\sum\limits_{i=1}^{n} z_i(X_i - c - b(z_i - \overline{z})) = 0 \end{cases};$$

$$\begin{cases} c = \overline{X} \\ b = \dfrac{\sum\limits_{i=1}^{n} z_i\left(X_i - \overline{X}\right)}{\sum\limits_{i=1}^{n} z_i(z_i - \overline{z})} \end{cases}.$$

Так как $X_i - \overline{X}$ и $z_i - \overline{z}$ – центрированные величины, то получаем оптимальные ОНК:

$$\hat{c}_n = \overline{X}, \qquad \hat{b}_n = \frac{\sum\limits_{i=1}^{n}(z_i - \overline{z})(X_i - \overline{X})}{\sum\limits_{i=1}^{n}(z_i - \overline{z})^2};$$

$$\hat{a}_n = \overline{X} - \hat{b}_n\overline{z};$$

$$\hat{\sigma}_n^2 = \frac{1}{n-2}\sum_{i=1}^{n}\left(X_i - \overline{X} - \hat{b}_n(z_i - \overline{z})\right)^2.$$

б) Нам известно, что $\begin{pmatrix} \hat{c}_n \\ \hat{b}_n \end{pmatrix} \sim \mathcal{N}_2\left(\begin{pmatrix} c \\ b \end{pmatrix},\ \sigma^2\left(Z^TZ\right)^{-1}\right)$ и $\hat{\sigma}_n^2 \sim \dfrac{\sigma^2}{n-2}\chi^2(n-2)$, причём эти величины независимые.

На семинаре 1 мы доказали, что если $E\xi_n \to c$ и $D\xi_n \to 0$, то $\xi_n \xrightarrow{P} c$.

$$D\left(\hat{\sigma}_n^2\right) = \frac{\sigma^4}{n-2} \cdot (n-2) \cdot D\left(\mathcal{N}^2(0,1)\right) = \frac{2\sigma^4}{n-2} \underset{n\to+\infty}{\to} 0.$$

Таким образом, оценка $\hat{\sigma}^2$ состоятельная.

$$Z = \begin{pmatrix} 1 & z_1 - \overline{z} \\ \vdots & \vdots \\ 1 & z_n - \overline{z} \end{pmatrix}; \qquad Z^T Z = \begin{pmatrix} n & 0 \\ 0 & \sum\limits_{i=1}^{n} (z_i - \overline{z})^2 \end{pmatrix};$$

$$D \begin{pmatrix} \hat{c}_n \\ \hat{b}_n \end{pmatrix} = \begin{pmatrix} \dfrac{\sigma^2}{n} & 0 \\ 0 & \dfrac{\sigma^2}{\sum\limits_{i=1}^{n} (z_i - \overline{z})^2} \end{pmatrix}. \tag{14.1}$$

Утверждение 14.2. *Пусть* $\xi_n \sim \mathcal{N}(c, \sigma_n^2)$. *Тогда* $\xi_n \xrightarrow{P} c \iff \sigma_n \to 0.$

Доказательство:

На семинаре 1 мы доказывали, что если дисперсия стремится к 0, то оценка состоятельна. То есть доказано $\Longleftarrow$.

Докажем $\Longrightarrow$.

$$\forall \varepsilon > 0 \qquad P(|\xi_n - c| \leqslant \varepsilon) = P(-\varepsilon \leqslant \xi_n - c \leqslant \varepsilon) = P\left(-\frac{\varepsilon}{\sigma_n} \leqslant \frac{\xi_n - c}{\sigma_n} \leqslant \frac{\varepsilon}{\sigma_n}\right) =$$

$$= \Phi\left(\frac{\varepsilon}{\sigma_n}\right) - \Phi\left(-\frac{\varepsilon}{\sigma_n}\right) \underset{n \to +\infty}{\to} 1.$$

Так как $\Phi \in [0, 1]$, то $\Phi\left(\dfrac{\varepsilon}{\sigma_n}\right) \to 1$ и $\Phi\left(-\dfrac{\varepsilon}{\sigma_n}\right) \to 0$ при $n \to +\infty$. Тогда $\sigma_n \to 0$ при $n \to +\infty$. ■

Из выражения (14.1) и утверждения (14.2) получаем:

$$\hat{c}_n = \overline{X}, \qquad D(\hat{c}_n) = \frac{\sigma^2}{n} \to 0 \quad \Longrightarrow \quad \hat{c}_n \text{ – состоятельная;}$$

$$\hat{b}_n \xrightarrow{P} b \quad \iff \quad D(\hat{b}_n) = \frac{\sigma^2}{\sum\limits_{i=1}^{n} (z_i - \overline{z})^2} \to 0 \quad \iff \quad \sum_{i=1}^{n} (z_i - \overline{z}) \to +\infty.$$

Если $\hat{b}_n$ состоятельна, то и $\hat{a}_n = \overline{X} - \hat{b}_n \overline{z}$ состоятельна.

в) Сначала построим доверительный интервал уровня $1 - \alpha$ для b.

Пусть $nS_z^2 \coloneqq \sum\limits_{i=1}^{n} (z_i - \overline{z})^2$, тогда $\hat{b}_n \sim \mathcal{N}\left(b, \dfrac{\sigma^2}{nS_z^2}\right)$.

$$\frac{\left(\hat{b}_n - b\right) \sqrt{n} S_z}{\hat{\sigma}_n} = \frac{\left(\hat{b}_n - b\right) \sqrt{n} S_z / \sigma}{\hat{\sigma}_n / \sigma} = \frac{\mathcal{N}(0, 1)}{\sqrt{\dfrac{1}{n-2} \chi^2(n-2)}},$$

причём числитель и знаменатель независимы, так как $\hat{b}_n$ и $\hat{\sigma}_n^2$ независимы. Тогда получаем распределение Стьюдента: $\dfrac{\mathcal{N}(0,1)}{\sqrt{\dfrac{1}{n-2}\chi^2(n-2)}} \sim t(n-2)$.

$$P_{a,\,b,\,\sigma^2}\left(t_{\alpha/2,\,n-2} < \left(\hat{b}_n - b\right)\frac{\sqrt{n}S_z}{\hat{\sigma}_n} < -t_{\alpha/2,\,n-2}\right) \to 1-\alpha;$$

$$P_{a,\,b,\,\sigma^2}\left(b \in \left(\hat{b}_n + t_{\alpha/2,\,n-2}\frac{\hat{\sigma}_n}{\sqrt{n}S_z};\ \hat{b}_n - t_{\alpha/2,\,n-2}\frac{\hat{\sigma}_n}{\sqrt{n}S_z}\right)\right) \to 1-\alpha.$$

Теперь построим доверительный интервал уровня $1-\alpha$ для $a+bz$, где $z \in \mathbb{R}$.

Вспоминаем, что $\begin{pmatrix}\hat{c}_n\\ \hat{b}_n\end{pmatrix} \sim \mathcal{N}_2\left(\begin{pmatrix}c\\ b\end{pmatrix},\ \begin{pmatrix}\dfrac{\sigma^2}{n} & 0\\ 0 & \dfrac{\sigma^2}{nS_z^2}\end{pmatrix}\right)$, причём $\hat{c}_n$ и $\hat{b}_n$ независимы.

$$\hat{a}_n + \hat{b}_n z = \hat{c}_n + \hat{b}_n(z-\overline{z});$$

$$D\left(\hat{a}_n + \hat{b}_n z\right) = D\hat{c}_n + (z-\overline{z})^2 D\hat{b}_n = \frac{\sigma^2}{n^2}\left(1+\frac{(z-\overline{z})^2}{S_z^2}\right);$$

$$\hat{a}_n + \hat{b}_n z \sim \mathcal{N}\left(a+bz,\ \frac{\sigma^2}{n^2}\left(1+\frac{(z-\overline{z})^2}{S_z^2}\right)\right);$$

$$\frac{\sqrt{n}\left(\hat{a}_n + \hat{b}_n z - (a+bz)\right)}{\hat{\sigma}_n\sqrt{1+\dfrac{(z-\overline{z})^2}{S_z^2}}} \sim t(n-2).$$

Тогда, как и в предыдущем пункте получаем доверительный интервал для $a+bz$ уровня $1-\alpha$:

$$\left(a+bz+t_{\alpha/2,\,n-2}\frac{\hat{\sigma}_n}{\sqrt{n}}\sqrt{1+\frac{(z-\overline{z})^2}{S_z^2}};\ a+bz-t_{\alpha/2,\,n-2}\frac{\hat{\sigma}_n}{\sqrt{n}}\sqrt{1+\frac{(z-\overline{z})^2}{S_z^2}}\right).$$

$\square$

Задача 14.3. Пусть $X_i = a + \varepsilon_i$, где $i = \overline{1,n}$, a – неизвестные параметр, $\{\varepsilon_i\}$ – независимые случайные величины $\mathcal{N}\left(0,\ \dfrac{\sigma^2}{w_i^2}\right)$, где σ^2 – неизвестный параметр, $\{w_i\}$ – известные положительные веса. Найти оптимальные оценки для параметров a, σ^2 и их распределения.

Решение:

Так как дисперсии ошибок различны, то эта задача не является задачей линейной регрессии по умолчанию. Сведём её к задаче линейной регрессии. Для этого умножим выражения для наблюдений на известные веса:

$$w_i X_i = w_i a + \delta_i,$$

где $\delta_i = w_i \varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ – независимые одинаково распределённые случайные величины. Получили модель линейной регрессии для наблюдений $w_i X_i$. Тогда оптимальные оценки дадут ОНК:

$$\sum_{i=1}^{n} w_i^2 (X_i - a)^2 \to \min_a;$$

$$\frac{d}{da} \sum_{i=1}^{n} w_i^2 (X_i - a)^2 = -2 \sum_{i=1}^{n} w_i^2 (X_i - a) = 0;$$

$$\hat{a}_n = \frac{\sum\limits_{i=1}^{n} w_i^2 X_i}{\sum\limits_{i=1}^{n} w_i^2} \sim \mathcal{N}\left(a,\ \frac{\sigma^2}{\sum\limits_{i=1}^{n} w_i^2}\right);$$

$$\hat{\sigma}_n^2 \sim \frac{\sigma^2}{n-1} \chi^2(n-1)$$

□

Семинар 15

Проверка гипотез

Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim p(x, \theta)$, где параметр $\theta \in \Theta$. Пусть $H_0\colon\ \theta \in \Theta_0 \subset \Theta$ – гипотеза. Тогда $H_1\colon\ \theta \in \Theta_1 \subseteq \Theta \setminus \Theta_0$ – альтернативная гипотеза.

Пусть $\mathcal{X}$ – множество всех возможных значений наблюдения X. Пусть $\mathcal{X} = \mathcal{X}_{1\alpha} \bigsqcup \mathcal{X}_{0\alpha}$. Тогда будем действовать по следующему правилу: если $X \in \mathcal{X}_{1\alpha}$, то принимаем гипотезу H_1 (отвергаем H_0), а если $X \in \mathcal{X}_{0\alpha}$, то принимаем гипотезу H_0 (отвергаем H_1). Множество $\mathcal{X}_{1\alpha}$ называется *критическим множеством* (*критерием*), а множество $\mathcal{X}_{0\alpha}$ называется *областью принятия гипотезы* H_0.

H_0	Верна	Неверна
Принимаем	+	Ошибка 2-го рода
Отвергаем	Ошибка 1-го рода	+

$\alpha \coloneqq P(H_1|H_0)$ – вероятность ошибки 1-го рода.

$\beta \coloneqq P(H_0|H_1)$ – вероятность ошибки 2-го рода.

Определение 15.1. *Функцией мощности критерия* $\mathcal{X}_{1\alpha}$ называется

$$W(\theta) = W(\theta, \mathcal{X}_{1\alpha}) = P_\theta(X \in \mathcal{X}_{1\alpha}).$$

Таким образом, $\alpha = W(\theta)$ при $\theta \in \Theta_0$, $\beta = 1 - W(\theta)$ при $\theta \in \Theta_1$.

Определение 15.2. Если $W(\theta, \mathcal{X}_{1\alpha}) \leqslant \alpha \quad \forall \theta \in \Theta_0$, то α – *уровень значимости критерия* $\mathcal{X}_{1\alpha}$.

Определение 15.3. *Размер критерия* $\mathcal{X}_{1\alpha}$ – наименьший уровень значимости этого критерия.

Пусть $\{\mathcal{X}_{1\alpha}\}$ – совокупность критериев уровня значимости α.

Определение 15.4. Оптимальный критерий $\mathcal{X}^*_{1\alpha}$ (*равномерно наиболее мощный (РНМ) критерий, или тест*) – такой критерий, для которого

$$W(\theta, \mathcal{X}^*_{1\alpha}) \geqslant W(\theta, \mathcal{X}_{1\alpha}) \qquad \forall \theta \in \Theta_1 \quad \forall \mathcal{X}_{1\alpha}.$$

Определение 15.5. Если $H_0:\ \ \theta = \theta_0$ – одно гипотетическое распределение, то H_0 – *простая гипотеза.* Аналогично, альтернатива H_1 называется *простой*, если состоит ровно из одного распределения.

Определение 15.6. Критерий $\mathcal{X}_{1\alpha}$ называется *несмещённым*, если

$$W(\theta_0, \mathcal{X}_{1\alpha}) \leqslant W(\theta_1, \mathcal{X}_{1\alpha}) \qquad \forall \theta \in \Theta_0 \quad \forall \theta_1 \in \Theta_1.$$

Лемма 15.1 (Лемма Неймана-Пирсона)**.** *Пусть проверяется простая гипотеза $H_0: \ \theta = \theta_0$ против простой альтернативы $H_1: \ \theta = \theta_1$, где $\theta_1 \neq \theta_0$. Запишем правдоподобие наблюдения X: $L(X, \theta) = \prod\limits_{i=1}^{n} p(X_i, \theta)$. Пусть*

$$S_\lambda \coloneqq \{X: \ L(X, \theta_1) - \lambda L(X, \theta_0) \geqslant 0\}$$

– критическое множество. Он называется критерий отношения правдоподобия. Значение λ определяется из условия $P_{\theta_0}(X \in S_\lambda) = \alpha$. Пусть это уравнение имеет решение по λ. Тогда S_λ – наиболее мощный и несмещённый тест уровня α.

Задача 15.1. Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim \mathcal{N}(\theta, \sigma^2)$, где $\theta \in \mathbb{R}$ – неизвестный параметр, а $\sigma^2 > 0$ известно. Построить оптимальный тест уровня α для проверки простой гипотезы $H_0\colon \ \theta = \theta_0$ против простой альтернативы $H_1\colon \ \theta = \theta_1$, где $\theta_1 > \theta_0$ – известные числа.

Решение:

$$L(X, \theta) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \theta)^2\right);$$

$$\frac{L(X, \theta_1)}{L(X, \theta_0)} = \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left((X_i - \theta_1)^2 - (X_i - \theta_0)^2\right)\right) \geqslant \lambda;$$

$$(\theta_1 - \theta_0)\sum_{i=1}^{n}\left(X_i - \frac{\theta_0 + \theta_1}{2}\right) \geqslant \lambda \qquad (\lambda \text{ – новая константа});$$

$$\sum_{i=1}^{n} X_i(\theta_1 - \theta_0) \geqslant \lambda \qquad (\lambda \text{ – новая константа});$$

$$\overline{X} \geqslant \lambda \qquad (\lambda \text{ – новая константа}).$$

Таким образом, $\{\overline{X} \geqslant \lambda\}$ – вид наиболее мощного критерия. Эта константа λ называется критическим значением, а $\overline{X}$ называется тестовой статистикой. Найдём критическое значение λ из соотношения $\alpha = P_{\theta_0}\left(\overline{X} \geqslant \lambda\right)$. Так как $\overline{X} \sim \mathcal{N}\left(0, \dfrac{\sigma^2}{n}\right)$, то получаем:

$$\alpha = P_{\theta_0}\left(\overline{X} \geqslant \lambda\right) = P_{\theta_0}\left(\frac{\sqrt{n}\left(\overline{X} - \theta_0\right)}{\sigma} \geqslant \frac{\sqrt{n}(\lambda - \theta_0)}{\sigma}\right) = 1 - \Phi\left(\frac{\sqrt{n}(\lambda - \theta_0)}{\sigma}\right);$$

$$\frac{\sqrt{n}(\lambda - \theta_0)}{\sigma} = Z_{1-\alpha} = -Z_\alpha; \qquad \lambda = \theta_0 - \frac{\sigma}{\sqrt{n}} Z_\alpha;$$

$$\mathcal{X}_{1\alpha}^* = \left\{X: \ \overline{X} \geqslant \theta_0 - \frac{\sigma}{\sqrt{n}} Z_\alpha\right\}.$$

□

Замечание 15.1. Критическая граница $\theta_0 - \dfrac{\sigma}{\sqrt{n}} Z_\alpha$ не зависит от θ_1, значит, $\mathcal{X}^*_{1\alpha}$ – РНМ тест уровня α для проверки гипотезы $H_0: \ \theta = \theta_0$ против альтернативы $H_1^+: \ \theta > \theta_0$ (такая альтернатива называется *правосторонней*).

Состоятельный критерий

Определение 15.7. Критерий $\mathcal{X}_{1\alpha}$ называется *состоятельным*, если

$$W(\theta, \mathcal{X}_{1\alpha}) \underset{n\to+\infty}{\to} 1 \qquad \forall \theta \in \Theta_1.$$

Задача 15.2. Для $\mathcal{X}^*_{1\alpha}$ из задачи (15.1) вычислить $W(\theta, \mathcal{X}^*_{1\alpha})$, где $\theta > \theta_0$, и проверить $\mathcal{X}^*_{1\alpha}$ на состоятельность.

Решение:

$$\mathcal{X}^*_{1\alpha} = \left\{ X: \ \overline{X} \geqslant \theta_0 - \frac{\sigma}{\sqrt{n}} Z_\alpha \right\};$$

$$W(\theta, \mathcal{X}^*_{1\alpha}) = P_\theta \left(\overline{X} \geqslant \theta_0 - \frac{\sigma}{\sqrt{n}} Z_\alpha \right) = P_\theta \left(\frac{\sqrt{n}\left(\overline{X} - \theta\right)}{\sigma} \geqslant \frac{\sqrt{n}(\theta_0 - \theta)}{\sigma} - Z_\alpha \right) =$$

$$= 1 - \Phi\left(\frac{\sqrt{n}(\theta_0 - \theta)}{\sigma} - Z_\alpha \right) \underset{n\to+\infty}{\to} 1.$$

□

Задача 15.3. Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim \mathcal{N}(\theta, \sigma^2)$, где $\theta \in \mathbb{R}$ – неизвестный параметр, а $\sigma^2 > 0$ известно. Найти наименьшее количество наблюдений n, необходимое для выполнения неравенства $1 - W(\theta_1, \mathcal{X}^*_{1\alpha}) \leqslant \beta$.

Решение:

$$1 - W(\theta_1, \mathcal{X}^*_{1\alpha}) = \Phi\left(\frac{\sqrt{n}(\theta_0 - \theta_1)}{\sigma} - Z_\alpha \right) = \beta;$$

$$\frac{\sqrt{n}(\theta_0 - \theta_1)}{\sigma} - Z_\alpha = Z_\beta;$$

$$n = \left\lceil \left(\frac{\sigma(Z_\alpha + Z_\beta)}{\theta_0 - \theta_1} \right)^2 \right\rceil.$$

□

Замечание 15.2. В случае проверки гипотезы $H_0: \ \theta = \theta_0$ относительно левосторонней альтернативы $H_1: \ \theta < \theta_0$ получим:

$$\mathcal{X}^*_{1\alpha} = \{X: \ \overline{X} \leqslant \lambda\} = \left\{ X: \ \frac{\sqrt{n}\left(\overline{X} - \theta_0\right)}{\sigma} \leqslant Z_\alpha \right\}.$$

Так как критическая граница не зависит от θ_1, то получили РНМ тест уровня α. Этот тест несмещённый и состоятельный.

Замечание 15.3. В случае проверки гипотезы $H_0: \ \theta = \theta_0$ относительно двусторонней альтернативы $H_1: \ \theta \neq \theta_0$ получим, что $\nexists$ РНМ теста.

Задача 15.4. Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim \mathrm{Pois}(\theta)$, где параметр $\theta > 0$. Построить оптимальный тест уровня α для простой гипотезы $H_0: \ \theta = \theta_0$ против левосторонней альтернативы $H_1^-: \ \theta < \theta_0$.

Решение:

$$L(X, \theta) = \prod_{i=1}^{n} \frac{\theta^{X_i}}{X_i!} e^{-\theta} = \frac{\theta^{\sum\limits_{i=1}^{n} X_i}}{\prod\limits_{i=1}^{n} X_i!} e^{-n\theta};$$

$$\text{При} \ \ \theta_1 < \theta_0 \qquad \frac{L(X, \theta_1)}{L(X, \theta_0)} \geqslant \lambda;$$

$$\left(\frac{\theta_1}{\theta_0}\right)^{\sum\limits_{i=1}^{n} X_i} e^{-n(\theta_1 - \theta_0)} \geqslant \lambda;$$

$$\sum_{i=1}^{n} X_i \ln \frac{\theta_1}{\theta_0} \geqslant \lambda \qquad (\lambda \text{ – новая константа})$$

$$\sum_{i=1}^{n} X_i \leqslant \lambda \qquad (\lambda \text{ – новая константа}).$$

Таким образом, $\left\{\sum\limits_{i=1}^{n} X_i \leqslant \lambda\right\}$ – наиболее мощное критическое множество. Найдём для него критическую границу λ из условия $P_{\theta_0}\left(\sum\limits_{i=1}^{n} X_i \leqslant \lambda\right) = \alpha$. Так как $\sum\limits_{i=1}^{n} X_i \sim \mathrm{Pois}(n\theta)$, то при $\lambda \in \mathbb{Z}$ получаем:

$$\sum_{k=0}^{\lambda} \frac{(n\theta_0)^k}{k!} e^{-n\theta_0} \leqslant \alpha < \sum_{k=0}^{\lambda+1} \frac{(n\theta_0)^k}{k!} e^{-n\theta_0}.$$

Пусть $\alpha' = \sum\limits_{k=0}^{\lambda} \frac{(n\theta_0)^k}{k!} e^{-n\theta_0}$. Заметим, что α' не зависит от θ_1. Тогда критическое множество $\mathcal{X}^*_{1\alpha'} = \{X: \ \sum\limits_{i=1}^{n} X_i \leqslant \lambda\}$ – РНМ тест уровня α' для проверки гипотезы $H_0: \ \theta = \theta_0$ против левосторонней альтернативы $H_1^-: \ \theta < \theta_0$.

Вычислим критическую границу λ приближённо по центральной предельной теореме:

$$\frac{\sqrt{n}\left(\overline{X}-\theta\right)}{\sqrt{\theta}} \xrightarrow{d} \mathcal{N}(0,1) \qquad \forall\theta > 0;$$

$$\alpha = P_{\theta_0}\left(\sum_{i=1}^{n} X_i \leqslant \lambda\right) = P_{\theta_0}\left(\frac{\sqrt{n}\left(\overline{X}-\theta_0\right)}{\sqrt{\theta_0}} \leqslant \frac{\sqrt{n}\left(\dfrac{\lambda}{n}-\theta_0\right)}{\sqrt{\theta_0}}\right);$$

$$\frac{\sqrt{n}}{\sqrt{\theta_0}}\left(\frac{\lambda}{n}-\theta_0\right) \approx Z_\alpha; \qquad\qquad \lambda \approx n\theta_0 + \sqrt{n\theta_0}Z_\alpha.$$

Докажем состоятельность критерия:

$$\forall\theta < \theta_0 \qquad P_\theta\left(\overline{X} \leqslant \frac{\lambda}{n}\right) = P_\theta\left(\frac{\sqrt{n}\left(\overline{X}-\theta\right)}{\sqrt{\theta}} \leqslant \frac{\sqrt{n}\left(\dfrac{\lambda}{n}-\theta\right)}{\sqrt{\theta}}\right) \approx$$

$$\approx \Phi\left(\frac{\sqrt{n}(\theta_0-\theta)+\sqrt{\theta_0}Z_\alpha}{\sqrt{\theta}}\right) \underset{n\to+\infty}{\to} 1.$$

$\square$

Семинар 16

Связь доверительного оценивания и проверки гипотез

Пусть X – наблюдение, $X \sim P_\theta$, где параметр $\theta \in \Theta$.

1) Пусть $\forall \theta_0 \in \Theta \quad \mathcal{X}_{1\alpha}(\theta_0)$ – критическое множество уровня α для проверки гипотезы $H_0\colon\ \theta = \theta_0$ против альтернативы $H_1\colon\ \theta \neq \theta_0$. Тогда

$$I_{1-\alpha}(X) = \{\theta \in \Theta\colon\ X \notin \mathcal{X}_{1\alpha}(\theta)\}$$

– доверительное множество для θ уровня $1-\alpha$.

2) Пусть $I_{1-\alpha}(X)$ – доверительное множество для параметра θ уровня $1-\alpha$. Тогда

$$\mathcal{X}_{1\alpha}(\theta_0) = \{X\colon\ \theta_0 \notin I_{1-\alpha}(X)\}$$

– критерий уровня α для проверки гипотезы $H_0\colon\ \theta = \theta_0$.

Если критерий несмещённый, то и доверительный интервал будет несмещённым, и наоборот. Также оптимальному критерию в классе несмещённых соответствует оптимальный доверительный интервал в классе несмещённых.

Решение задач на связь доверительного оценивания и проверки гипотез

Задача 16.1. Пусть $X = (X_1, \ldots, X_n)$ – выборка, $X_1 \sim \mathcal{N}(\theta, \sigma^2)$, где параметр $\theta \in \mathbb{R}^1$ неизвестный, а значение σ^2 известно. Построить критерий уровня α для проверки гипотезы $H_0\colon\ \theta = \theta_0$ против двусторонней альтернативы $H_1\colon\ \theta \neq \theta_0$. Вычислить мощность критерия. Исследовать его на состоятельность и несмещённость.

Решение:

$\overline{X}$ – оптимальная оценка для θ, $\dfrac{\sqrt{n}\left(\overline{X}-\theta\right)}{\sigma} \sim \mathcal{N}(0,1)$ – центральная статистика. Доверительный интервал:

$$I_{1-\alpha}(X) = \left\{\theta\colon\ \left|\frac{\sqrt{n}\left(\overline{X}-\theta\right)}{\sigma}\right| < -Z_{\alpha/2}\right\},$$

где $\Phi(Z_\alpha) = \alpha$, то есть Z_α – α-квантиль $\mathcal{N}(0,1)$.

Построим критическое множество уровня α для проверки гипотезы $H_0\colon\ \theta = \theta_0$:

$$\mathcal{X}_{1\alpha}(\theta_0) = \left\{X\colon\ \left|\frac{\sqrt{n}\left(\overline{X}-\theta_0\right)}{\sigma}\right| \geqslant -Z_{\alpha/2}\right\}.$$

Вычислим мощность критерия:

$$W(\theta) = P_\theta\left(X \in \mathcal{X}_{1\alpha}(\theta_0)\right) = P_\theta\left(\left|\frac{\sqrt{n}\left(\overline{X}-\theta_0\right)}{\sigma}\right| \geqslant -Z_{\alpha/2}\right) =$$

$$= 1 - P_\theta\left(Z_{\alpha/2} < \frac{\sqrt{n}\left(\overline{X}-\theta_0\right)}{\sigma} < -Z_{\alpha/2}\right) =$$

$$= 1 - P_\theta\left(\frac{\sqrt{n}(\theta_0-\theta)}{\sigma} + Z_{\alpha/2} < \frac{\sqrt{n}\left(\overline{X}-\theta\right)}{\sigma} < \frac{\sqrt{n}(\theta_0-\theta)}{\sigma} - Z_{\alpha/2}\right) =$$

$$= 1 - \left(\Phi\left(\frac{\sqrt{n}(\theta_0-\theta)}{\sigma} - Z_{\alpha/2}\right) - \Phi\left(\frac{\sqrt{n}(\theta_0-\theta)}{\sigma} + Z_{\alpha/2}\right)\right) \ominus$$

Так как $1-\Phi(x) = \Phi(-x)$, получаем:

$$\ominus\ \Phi\left(Z_{\alpha/2} + \frac{\sqrt{n}(\theta-\theta_0)}{\sigma}\right) + \Phi\left(Z_{\alpha/2} - \frac{\sqrt{n}(\theta-\theta_0)}{\sigma}\right).$$

Чтобы понять как ведёт себя функция мощности, вычислим её производную:

$$W'(\theta) = \frac{\sqrt{n}}{\sigma}\left(\varphi\left(Z_{\alpha/2} + \frac{\sqrt{n}(\theta-\theta_0)}{\sigma}\right) - \varphi\left(Z_{\alpha/2} - \frac{\sqrt{n}(\theta-\theta_0)}{\sigma}\right)\right).$$

Отсюда видно, что $\theta = \theta_0$ – точка минимума. Таким образом, найденный критерий является состоятельным и несмещённым. $\square$

Замечание 16.1.

$$\mathcal{X}_{1\alpha} = \left\{X\colon\ \left|\frac{\sqrt{n}\left(\overline{X}-\theta_0\right)}{\sigma}\right| \geqslant -Z_{\alpha/2}\right\} =$$

$$= \left\{\frac{\sqrt{n}\left(\overline{X}-\theta_0\right)}{\sigma} \geqslant -Z_{\alpha/2}\right\} \cup \left\{\frac{\sqrt{n}\left(\overline{X}-\theta_0\right)}{\sigma} \leqslant Z_{\alpha/2}\right\}.$$

Заметим, что

$$\mathcal{X}^+_{1,\,\alpha/2} \coloneqq \left\{\frac{\sqrt{n}\left(\overline{X}-\theta_0\right)}{\sigma} \geqslant -Z_{\alpha/2}\right\}$$

– оптимальный критерий уровня $\dfrac{\alpha}{2}$ гипотезы против правосторонней альтернативы $H_1^+\colon\ \theta > \theta_0$, а

$$\mathcal{X}^-_{1,\,\alpha/2} \coloneqq \left\{\frac{\sqrt{n}\left(\overline{X}-\theta_0\right)}{\sigma} \leqslant Z_{\alpha/2}\right\}$$

– оптимальный критерий уровня $\dfrac{\alpha}{2}$ гипотезы против левосторонней альтернативы $H_1^-\colon\ \theta < \theta_0$. Можно показать, что этот критерий будет РНМ несмещённым критерием.

Замечание 16.2. Рассмотрим случай неизвестного значения σ^2.

Сформулируем правило: по выборке X из $\mathcal{N}(\theta, \sigma^2)$, где параметры $\theta \in \mathbb{R}^1$ и $\sigma^2 > 0$ неизвестны, отвергаем гипотезу $H_0\colon\ \theta = \theta_0$ против двусторонней альтернативы $H_1\colon\ \theta \neq \theta_0$, если $\left|\dfrac{\sqrt{n}\left(\overline{X} - \theta_0\right)}{S}\right| > -t_{\alpha/2}(n-1)$, где $t_{\alpha/2}(n-1)$ – квантиль уровня $\dfrac{\alpha}{2}$ распределения Стьюдента $t(n-1)$, а $S^2 = \dfrac{1}{n-1}\sum\limits_{i=1}^{n}\left(X_i - \overline{X}\right)^2$.

Задача 16.2. Пусть $X = (X_1, \ldots, X_n)$ и $Y = (Y_1, \ldots, Y_m)$ – две независимые выборки, $X_1 \sim \mathcal{N}(a, \sigma^2)$, $Y_1 \sim \mathcal{N}(b, \sigma^2)$, где a, b и $\sigma^2 > 0$ – неизвестные параметры. Проверить гипотезу $H_0\colon\ a = b$ против альтернативы $H_1\colon\ a \neq b$, для этого:

а) построить доверительный интервал для $a - b$ уровня $1 - \alpha$;

б) построить критическое множество уровня α для проверки гипотезы $H_0\colon\ a = b$ против альтернативы $H_1\colon\ a \neq b$.

Решение:

Объединим все наблюдения в один вектор:

$$\begin{pmatrix} X_1 \\ \vdots \\ X_n \\ Y_1 \\ \vdots \\ Y_m \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_{n+m} \end{pmatrix},$$

где $\{\varepsilon_i\}$ – зависимые одинаково распределённые случайные величины $\mathcal{N}(0, \sigma^2)$.

а) В модели линейной регрессии оптимальной оценкой является ОНК.

$$\sum_{i=1}^{n}(X_i - a)^2 + \sum_{i=1}^{n}(Y_i - b)^2 \to \min_{a,\, b};$$

$$\hat{a} = \overline{X}, \qquad \hat{b} = \overline{Y}, \qquad \hat{\sigma}^2 = \frac{1}{n+m-2}\left(\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 + \sum_{i=1}^{m}\left(Y_i - \overline{Y}\right)^2\right).$$

Таким образом, $\overline{X} - \overline{Y}$ – оптимальная оценка для $a - b$, $\overline{X} - \overline{Y} \sim \mathcal{N}\left(a - b,\ \dfrac{\sigma^2}{n} + \dfrac{\sigma^2}{m}\right)$.

Построим центральную статистику. Так как $\hat{a}$ и $\hat{b}$ не зависят от $\hat{\sigma}^2$, получаем:

$$\frac{\overline{X} - \overline{Y} - (a - b)}{\sqrt{\dfrac{1}{n} + \dfrac{1}{m}} \cdot \hat{\sigma}} \sim t(n + m - 2).$$

Построим доверительный интервал для $a-b$ уровня $1-\alpha$:

$$I_{1-\alpha}(X,Y)=\left\{a-b\colon\ \left|\frac{\overline{X}-\overline{Y}-(a-b)}{\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}\cdot\hat{\sigma}}\right|<-t_{\alpha/2}(n+m-2)\right\}.$$

б) Построим критическое множество уровня для проверки гипотезы $H_0\colon\ a-b=0$ против альтернативы $H_1\colon\ a-b\neq 0$:

$$\mathcal{X}_{1\alpha}=\left\{X,Y\colon\ \frac{|\overline{X}-\overline{Y}|}{\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}\cdot\hat{\sigma}}\geqslant -t_{\alpha/2}(n+m-2)\right\}=$$

$$=\left\{X,Y\colon\ \frac{\left(\overline{X}-\overline{Y}\right)^2}{\left(\dfrac{1}{n}+\dfrac{1}{m}\right)\hat{\sigma}^2}\geqslant t^2_{\alpha/2}(n+m-2)\right\}$$

$\square$

Определение 16.1. $F(k_1,k_2)\overset{d}{=}\dfrac{\frac{1}{k_1}\chi^2(k_1)}{\frac{1}{k_2}\chi^2(k_2)}$ – *F-распределение Фишера*, где $\chi^2(k_1)$ и $\chi^2(k_2)$ независимы.

Замечание 16.3. Так как $t(k_2)\overset{d}{=}\dfrac{\xi_0}{\sqrt{\frac{1}{k_2}\chi^2(k_2)}}$, где ξ_0 не зависит от $\chi^2(k_2)$, то $F(1,k_2)\overset{d}{=}t^2(k_2)$. Отсюда можно показать, что $t^2_{\alpha/2}(n+m-2)=f_{1-\alpha}(1,\,n+m-2)$.

Задача 16.3. Доказать, что $t^2_{\alpha/2}(k)=f_{1-\alpha}(1,k)$.

Решение:

$$1-\alpha=P\left(F(1,k)\leqslant f_{1-\alpha}(1,k)\right);$$

$$\alpha=P\left(F(1,k)>f_{1-\alpha}(1,k)\right)=P\left(\frac{\xi^2}{\frac{1}{k}\chi^2(k)}>f_{1-\alpha}(1,k)\right),$$

где $\xi\sim\mathcal{N}(0,1)$ и не зависит от $\chi^2(k)$.

$$\alpha = P\left(\left|\frac{\xi}{\sqrt{\frac{1}{k}\chi^2(k)}}\right| > \sqrt{f_{1-\alpha}(1,k)}\right) = P\left(|t(k)| > \sqrt{f_{1-\alpha}(1,k)}\right) =$$

$$= 2P\left(t(k) < -\sqrt{f_{1-\alpha}(1,k)}\right);$$

$$P\left(t(k) < -\sqrt{f_{1-\alpha}(1,k)}\right) = \frac{\alpha}{2}; \qquad -\sqrt{f_{1-\alpha}(1,k)} = t_{\alpha/2}(k); \qquad t^2_{\alpha/2}(k) = f_{1-\alpha}(1,k).$$

$\square$

Семинар 17

Нецентральные распределения

Определение 17.1. $t(k,\mu) \stackrel{d}{=} \dfrac{\xi_0+\mu}{\sqrt{\dfrac{1}{k}\left(\xi_1^2+\ldots+\xi_k^2\right)}}$ – *распределение Стьюдента с числом степеней свободы k и параметром нецентральности μ*, где $\{\xi_i\}$ – независимые одинаково распределённые случайные величины $\mathcal{N}(0,1)$.

Определение 17.2. $\chi^2(k,\Delta^2) \stackrel{d}{=} (\xi_1+a_1)^2+\ldots+(\xi_k+a_k)^2$ – *распределение χ^2 с числом степеней свободы k и параметром нецентральности Δ^2*, где $\{\xi_i\}$ – независимые одинаково распределённые случайные величины $\mathcal{N}(0,1)$, $\Delta^2 = a_1^2+\ldots+a_k^2$.

Замечание 17.1. Заметим, что в определении (17.2) в равенстве слева указана зависимость только от квадрата длины вектора $a = \begin{pmatrix} a_1 \\ \vdots \\ a_k \end{pmatrix}$, а справа – от компонент этого вектора. Покажем, что определение корректно.

Пусть $\xi = \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_k \end{pmatrix}$, тогда $\xi \sim \mathcal{N}_k(0, I_k)$. Пусть C – ортогональная матрица размера $k\times k$, тогда

$$C\xi \sim \mathcal{N}_k(C\cdot 0,\, CI_kC^T) = \mathcal{N}_k(0, I_k).$$

Таким образом, $\xi \stackrel{d}{=} C\xi$.

Так как при умножении вектора на ортогональную матрицу его длина не меняется, то получаем:

$$(\xi_1+a_1)^2+\ldots+(\xi_k+a_k)^2 = \|\xi+a\|^2 = \|C\xi+Ca\|^2 \stackrel{d}{=} \|\xi+Ca\|^2.$$

Выберем C так, чтобы выполнялось $Ca = \begin{pmatrix} \|a\| \\ \vdots \\ 0 \end{pmatrix}$. Таким образом, видно, что распределение зависит только от квадрата длины вектора a.

Определение 17.3. $F(k_1,k_2,\Delta^2) \stackrel{d}{=} \dfrac{\dfrac{1}{k_1}\chi^2(k_1,\Delta^2)}{\dfrac{1}{k_2}\chi^2(k_2)}$ – *распределение Фишера с числом степеней свободы k_1, k_2 и параметром нецентральности Δ^2*, где $\chi^2(k_1,\Delta^2)$ и $\chi^2(k_2)$ – независимые случайные величины.

Стохастическая упорядоченность

Определение 17.4. $\xi \underset{st}{\leqslant} \eta$ *(ξ стохастически не больше, чем η)*, если

$$P(\xi > x) \leqslant P(\eta > x) \qquad \forall x \in \mathbb{R}.$$

Определение 17.5. $\xi \underset{st}{<} \eta$ *(ξ стохастически меньше, чем η)*, если

$$P(\xi > x) < P(\eta > x) \qquad \forall x \in \mathbb{R}.$$

Утверждение 17.1. *$t(k, \mu)$ стохастически упорядочены по μ.*

Утверждение 17.2. *$\chi^2(k, \Delta^2)$ стохастически упорядочены по $\Delta^2 \geqslant 0$.*

Утверждение 17.3. *$F(k_1, k_2, \Delta^2)$ стохастически упорядочены по $\Delta^2 \geqslant 0$.*

Задача 17.1. Доказать, что $\chi^2(k, \Delta^2)$ стохастически упорядочены по $\Delta^2 \geqslant 0$.

Решение:

Вспомним следующее утверждение:

Утверждение 17.4. *Если ξ и η – независимые случайные величины и $E|\varphi(\xi, \eta)|$, где φ – борелевская функция, то $E\varphi(\xi, \eta) = E\left(E\varphi(\xi, y)\big|_{y=\eta}\right)$.*

Выберем добавочный вектор с ненулевой только первой компонентой, получим:

$$P\left(\chi^2(k, \Delta^2) > x\right) = P\left((\xi_1 + \Delta)^2 + \xi_2^2 + \ldots + \xi_k^2 > x\right),$$

где $\{\xi_i\}$ – независимые одинаково распределённые случайные величины $\mathcal{N}(0, 1)$. Перейдём к математическому ожиданию:

$$P\left(\chi^2(k, \Delta^2) > x\right) = E\left(EI\left((\xi_1 + \Delta)^2 > x - x_2^2 - \ldots - x_k^2\right)\big|_{(x_2, \ldots, x_k) = (\xi_2, \ldots, \xi_k)}\right).$$

Обозначим $y \coloneqq x - x_2^2 - \ldots - x_k^2$ и $g_y(\Delta) \coloneqq EI\left((\xi_1 + \Delta)^2 > y\right)$. Если $y \leqslant 0$, то $g_y(\Delta) = 1$. Далее рассмотрим случай $y > 0$:

$$\begin{aligned}
&g_y(\Delta) = P\left(|\xi_1 + \Delta| > \sqrt{y}\right) = P\left(\xi_1 > \sqrt{y} - \Delta\right) + P\left(\xi_1 < -\sqrt{y} - \Delta\right) = \\
&= \Phi\left(\Delta - \sqrt{y}\right) + \Phi\left(-\Delta - \sqrt{y}\right); \\
&g'_y(\Delta) = \varphi\left(\Delta - \sqrt{y}\right) - \varphi\left(-\Delta - \sqrt{y}\right); \\
&g'_y(0) = 0, \qquad g'_y(\Delta) > 0 \quad \text{при} \quad \Delta > 0.
\end{aligned}$$

Отсюда следует, что $P(\chi^2(k, \Delta^2) > x) \leqslant P(\chi^2(k, \tilde{\Delta}^2) > x)$ при $\Delta < \tilde{\Delta}$. $\square$

Задача 17.2. Для критерия из задачи (16.2)

$$\mathcal{X}_{1\alpha} = \left\{ X, Y\colon\ \frac{\left(\overline{X}-\overline{Y}\right)^2}{\sqrt{\dfrac{1}{n}+\dfrac{1}{m}\hat{\sigma}^2}} > f_{1-\alpha}(1,\, n+m-2) \right\}$$

вычислить мощность $W(a-b)$ и доказать несмещённость теста.

Решение:

Так как $\overline{X}-\overline{Y} \sim \mathcal{N}\left(a-b,\ \left(\dfrac{1}{n}+\dfrac{1}{m}\right)\sigma^2\right)$, то получаем:

$$\frac{\overline{X}-\overline{Y}}{\sigma\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}} \sim \mathcal{N}\left(\frac{a-b}{\sigma\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}},\, 1\right);$$

$$\left(\frac{\overline{X}-\overline{Y}}{\sigma\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}}\right)^2 \sim \mathcal{N}^2\left(\frac{a-b}{\sigma\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}},\, 1\right) = \chi^2(1, \Delta^2).$$

Так как $\dfrac{\hat{\sigma}^2}{\sigma^2} \sim \dfrac{\chi^2(n+m-2)}{n+m-2}$ и случайные величины независимы, то получаем:

$$\frac{\left(\overline{X}-\overline{Y}\right)^2}{\left(\dfrac{1}{n}+\dfrac{1}{m}\right)\hat{\sigma}^2} \sim F(1, n+m-2, \Delta^2),$$

где $\Delta^2 = \dfrac{(a-b)^2}{\left(\dfrac{1}{n}+\dfrac{1}{m}\right)\sigma^2}$.

Выпишем функцию мощности:

$$W(a-b) = P_{a,\,b,\,\sigma^2}\left(\frac{\left(\overline{X}-\overline{Y}\right)^2}{\left(\dfrac{1}{n}+\dfrac{1}{m}\right)\hat{\sigma}^2} > f_{1-\alpha}(1,\, n+m-2)\right) =$$
$$= 1 - F_{1,\,n+m-2}\left(f_{1-\alpha}(1,\, n+m-2);\ \Delta^2\right).$$

Так как $\dfrac{\left(\overline{X}-\overline{Y}\right)^2}{\left(\dfrac{1}{n}+\dfrac{1}{m}\right)\hat{\sigma}^2} \sim F(1,\, n+m-2,\, \Delta^2)$, то в силу стохастической упорядоченности F-распределения $W(a-b)$ строго возрастает по параметру Δ^2. Так как

$W(0) = \alpha$ и W возрастает по параметру $|a-b|$, то $W(0)$ – наименьшее значение функции мощности. Так как случай $a-b=0$ соответствует гипотезе $H_0\colon\ a=b$, то тест несмещённый. $\square$

Построение доверительного множества в модели линейной гауссовской регрессии

Зафиксируем, как мы строили доверительные интервалы для некоторой линейной комбинации параметров.

Рассмотрим модель линейной гауссовской регрессии: $X = Z\theta + \varepsilon$, где $\varepsilon \sim \mathcal{N}_n(0,\, \sigma^2 I_n)$, $\operatorname{rank} Z = k < n$, параметры $\theta \in \mathbb{R}^k$ и $\sigma^2 > 0$ неизвестны. Построим доверительное множество для $\beta = A\theta$, где A – матрица размера $m \times k$, $\operatorname{rank} A = m \leqslant k$.

$\hat{\theta} = \left(Z^T Z\right)^{-1} Z^T X$ – ОНК для θ. Тогда $\hat{\beta} = A\hat{\theta}$ – оптимальная оценка для β. Так как $\hat{\theta} \sim \mathcal{N}_k\left(\theta,\, \sigma^2\left(Z^TZ\right)^{-1}\right)$, то $\hat{\beta} \sim \mathcal{N}_m\left(\beta,\, \sigma^2 A\left(Z^TZ\right)^{-1}A^T\right)$. Обозначим $D \coloneqq A\left(Z^TZ\right)^{-1}A^T$. Тогда получаем:

$$\left\| D^{-\frac{1}{2}}\left(\hat{\beta}-\beta\right)\frac{1}{\sigma}\right\|^2 \sim \|\mathcal{N}_m(0, I_m)\|^2 = \chi^2(m);$$

$$\frac{\frac{1}{m}\left(\hat{\beta}-\beta\right)^T D^{-1}\left(\hat{\beta}-\beta\right)}{\hat{\sigma}^2} \sim F(m,\, n-k).$$

Запишем доверительное множество:

$$I_{1-\alpha}(X) = \left\{\beta \in \mathbb{R}^m\colon\ \left(\hat{\beta}-\beta\right)^T D^{-1}\left(\hat{\beta}-\beta\right) < m\hat{\sigma}^2 f_{1-\alpha}(m,\, n-k)\right\}$$

– доверительный эллипсоид для β уровня $1-\alpha$.

Если мы хотим проверить линейную гипотезу $H_0:\ \beta = \beta_0$ в модели линейной регрессии, то критическим множеством уровня α для её проверки будет

$$\mathcal{X}_{1\alpha} = \left\{X\colon\ \left(\hat{\beta}-\beta_0\right)^T D^{-1}\left(\hat{\beta}-\beta_0\right) \geqslant m\hat{\sigma}^2 f_{1-\alpha}(m,\, n-k)\right\}.$$

Можно вычислить функцию мощности этого критерия:

$$W(\beta, \mathcal{X}_{1\alpha}) = 1 - F_{m,n-k}\left(f_{1-\alpha}(m,\, n-k);\ \Delta^2\right),$$

где $\Delta^2 = \dfrac{(\beta-\beta_0)^T D^{-1}(\beta-\beta_0)}{\sigma^2}$. Так как распределение Фишера стохастически упорядочено по параметру Δ^2, то критерий несмещённый и функция мощности растёт по параметру Δ^2.

Задача 17.3. Пусть дана модель линейной гауссовской регрессии: $X = Z\theta + \varepsilon$, где

$\varepsilon \sim \mathcal{N}_n(0,\, \sigma^2 I_n)$, где параметр $\theta \in \mathbb{R}^k$. Пусть $\theta = \begin{pmatrix}\theta^{(1)}\\ \theta^{(2)}\end{pmatrix}$, где размерности $\theta^{(1)}$ и $\theta^{(2)}$ равны k_1 и k_2 соответственно. Проверить гипотезу $H_0\colon\ \theta^{(2)} = 0$.

Решение:

$\hat\theta = \left(Z^T Z\right)^{-1} Z^T X \sim \mathcal{N}_k\left(\theta,\, \sigma^2 \left(Z^T Z\right)^{-1}\right)$ – ОНК для θ.

Тогда $\hat\theta^{(2)} \sim \mathcal{N}_{k_2}\left(\theta^{(2)},\, \sigma^2 D\right)$ – оптимальная оценка для $\theta^{(2)}$, где D – правый нижний блок размера $k_2 \times k_2$ матрицы $\left(Z^T Z\right)^{-1}$.

Построим критическое множество:

$$\mathcal{X}_{1\alpha} = \left\{X\colon\ \frac{\hat\theta^{(2)\,T} D^{-1} \hat\theta^{(2)}}{k_2 \hat\sigma^2} > f_{1-\alpha}(k_2,\, n-k)\right\}$$

– критерий уровня α для проверки гипотезы $H_0:\ \ \theta^{(2)} = 0$. Построенный тест несмещённый. □

Обсудим, как вычислять тестовую статистику простым способом.

Мы вычисляем ОНК, минимизируя выражение $RSS(\theta) \coloneqq \|X - Z\theta\|^2$. Тогда

$$\hat\theta^{(2)\,T} D^{-1} \hat\theta^{(2)} = RSS(\hat\theta_{H_0}) - RSS(\hat\theta),$$

причём $\hat\sigma^2 = \dfrac{RSS(\hat\theta)}{n-k}$.

Так как $H_0:\ \ \theta^{(2)} = 0$, то $\hat\theta_{H_0} = \begin{pmatrix}\hat\theta^{(1)}_{H_0}\\ 0\end{pmatrix}$. Разобьём матрицу Z на блоки $Z^{(1)}$, содержащий левые k_1 столбцов, и $Z^{(2)}$, содержащий правые k_2 столбцов.

Тогда в модели линейной регрессии $X = Z\theta + \varepsilon = Z^{(1)}\theta^{(1)} + Z^{(2)}\theta^{(2)} + \varepsilon$ ОНК получается следующим образом: $\hat\theta = \left(Z^T Z\right)^{-1} Z^T X$.

При гипотезе H_0 получаем: $X = Z^{(1)}\theta^{(1)} + \varepsilon$. Тогда $\hat\theta^{(1)}_{H_0} = \left(Z^{(1)\,T} Z^{(1)}\right)^{-1} Z^{(1)\,T} X$.

МЕХАНИКО-
МАТЕМАТИЧЕСКИЙ
ФАКУЛЬТЕТ
МГУ ИМЕНИ
М.В. ЛОМОНОСОВА

teach-in
ЛЕКЦИИ УЧЕНЫХ МГУ