Теория вероятности и математическая статистика, 2 курс, 2 семестр

@defunator

18 апреля 2020 г.

Содержание

1 Сходимости случайных величин

Определение 1. Последовательность случайных величин ξ_n сходится к случайной величине ξ:

1. *Почти наверное*($\xi_n \xrightarrow{\text{п.н.}} \xi$), если
$$P\left(\lim_{n\to\infty} \xi_n = \xi\right) = 1$$

2. *По вероятности*($\xi_n \xrightarrow{\text{p}} \xi$), если
$$\forall \varepsilon > 0 \ \lim_{n\to\infty} P\left(|\xi_n - \xi| \geq \varepsilon\right) = 0$$

3. *По распределению*($\xi_n \xrightarrow{\text{d}} \xi$), если
$$\lim_{n\to\infty} F_{\xi_n}(x) = F_\xi(x)$$
для любых x, в которых непрерывна F_ξ

Теорема 1. *(Эквивалетное определение сходимости по распределению)* $\xi_n \xrightarrow{d} \xi \Leftrightarrow \forall g$ *– непрерывна и ограничена на* $\mathbb{R}$ *верно* $\lim\limits_{n\to\infty} \mathbb{E}g(\xi_n) = \mathbb{E}g(\xi)$

Доказательство. $\Rightarrow$

Пусть t – точка непрерывности $F_\xi(t)$. Заметим, что $F_\xi(t) = P(\xi \in (-\infty, t]) =$
$= \mathbb{E}\mathrm{Ind}_{(-\infty,t]}(\xi)$.
В силу:
(1) $\mathbb{E}\mathrm{Ind}_{(a_i,b_i]}(\xi_n) = F_{\xi_n}(b_i) - F_{\xi_n}(a_i) \xrightarrow{n\to\infty} F_\xi(b_i) - F_\xi(a_i) = \mathbb{E}\mathrm{Ind}_{(a_i,b_i]}(\xi)$
(2) Линейность предела (с какими-то коэффициентами c_i)
Верна следующая сходимость:
$$\mathbb{E}\sum_{i=1}^{N} c_i \cdot \mathrm{Ind}_{(a_i,b_i]}(\xi_n) \to \mathbb{E}\sum_{i=1}^{N} c_i \cdot \mathrm{Ind}_{(a_i,b_i]}(\xi)$$

Теперь нам бы хотелось от непрерывной ограниченной функции на прямой перейти к функции на отрезке, а там мы уже сможем ее приблизить ступенчатой и воспользоваться предыдудщим утверждением и все доказать. Мы знаем, что
$\forall \varepsilon > 0 \ \exists A$: $P(-A < \xi \leq A) > 1 - \varepsilon$ (потому что $P(\xi \in \mathbb{R}) = 1$). Тогда получаем:
$$F_{\xi_n}(A) - F_{\xi_n}(-A) \xrightarrow{n\to\infty} F_\xi(A) - F_\xi(-A) = P(-A < \xi \leq A) > 1 - \varepsilon$$

То есть для $\forall \varepsilon > 0 \ \exists N \ \forall n > N$ верно:
$$|(F_{\xi_n}(A) - F_{\xi_n}(-A)) - (F_\xi(A) - F_\xi(-A))| < \varepsilon$$

Комбинируя последние два утверждения, получаем для $\forall \varepsilon > 0 \ \exists A \ \exists N \ \forall n > N$:
$$F_{\xi_n}(A) - F_{\xi_n}(-A) > F_\xi(A) - F_\xi(-A) - \varepsilon > 1 - 2\varepsilon$$

Из чего следует $\forall \varepsilon > 0 \ \exists A \ \exists N \ \forall n > N$:
$$P(-A \leq \xi_n \leq A) \geq F_{\xi_n}(A) - F_{\xi_n}(-A) > 1 - \varepsilon$$

Теперь возьмем любую непрерывную ограниченную функцию g, приблизим ее на отрезке $[-A, A]$ ступенчатой функцией g_ε, что $|g(x) - g_\varepsilon(x)| < \varepsilon$, а вне отрезка положим $g_\varepsilon = 0$. Имеем $\forall \varepsilon > 0 \ \exists A \ \forall n$:
$$|\mathbb{E}g(\xi_n) - \mathbb{E}g(\xi)| \leq \left|\mathbb{E}\left(1 - \mathrm{Ind}_{[-A,A]}(\xi_n)\right) \cdot g(\xi_n) - \mathbb{E}\left(1 - \mathrm{Ind}_{[-A,A]}(\xi)\right) g(\xi)\right| +$$

$$+\left|\mathbb{E}\mathrm{Ind}_{[-A,A]}(\xi_n)\cdot g(\xi_n)-\mathbb{E}\mathrm{Ind}_{[-A,A]}(\xi)\cdot g(\xi)\right|$$

Ясно, что первый модуль $< C\cdot 2\varepsilon$ (из ограниченности g $\forall x_1,\ x_2$: $|g(x_1)-g(x_2)|<C$ и так как $P(|\xi|>A),\ P(|\xi_n|>A)<\varepsilon$). А во втором модуле g заменим на g_ε с погрешностью ε, то есть он $<2\varepsilon+\left|\mathbb{E}\mathrm{Ind}_{[-A,A]}(\xi_n)\cdot g_\varepsilon(\xi_n)-\mathbb{E}\mathrm{Ind}_{[-A,A]}(\xi)\cdot g_\varepsilon(\xi)\right|$. А про оставшийся модуль мы уже знаем, что он сходится к 0, так как ступенчатая функция(то есть $<\varepsilon$ для $\forall n>N$). В итоге имеем: $|\mathbb{E}g(\xi_n)-\mathbb{E}g(\xi)|<\varepsilon$. То есть $\lim\limits_{n\to\infty}\mathbb{E}g(\xi_n)=\mathbb{E}g(\xi)$

$\Leftarrow$

Пусть t − точка непрерывности $F_\xi(t)$. Мы знаем, что $F_\xi(t)=\mathbb{E}\mathrm{Ind}_{(-\infty,t]}(\xi)$. Для $\forall\delta>0$ определим функции:

$$g_{-\delta}(x)=\begin{cases}1 & x<t-\delta\\ \frac{1}{\delta}\cdot(t-x) & t-\delta\le x\le t\\ 0 & t<x\end{cases}$$

$$g_{+\delta}(x)=\begin{cases}1 & x<t\\ \frac{1}{\delta}\cdot(t-x) & t\le x\le t+\delta\\ 0 & t+\delta<x\end{cases}$$

Заметим, что $\forall x$:

$$\mathrm{Ind}_{(-\infty,t-\delta]}(x)\le g_{-\delta}(x)\le \mathrm{Ind}_{(-\infty,t]}(x)\le g_{+\delta}(x)\le \mathrm{Ind}_{(\infty,t+\delta]}(x)$$

Взяв матожидания $(x=\xi_n)$ от второго и третьего неравенств, получим:

$$\mathbb{E}g_{-\delta}(\xi_n)\le F_{\xi_n}(t)\le \mathbb{E}g_{+\delta}(\xi_n)$$

Теперь устремим $n\to\infty$:

$$\mathbb{E}g_{-\delta}(\xi)\le \inf\lim_{n\to\infty}F_{\xi_n}(x)\le \sup\lim_{n\to\infty}F_{\xi_n}(x)\le \mathbb{E}g_{+\delta}(\xi)$$

Рассмотрим первое и последнее неравенство той цепочки($x=\xi$ и возьмем то него матожидание), получим:

$$F_\xi(t-\delta)\le \mathbb{E}g_{-\delta}(\xi)\le \mathbb{E}g_{+\delta}(\xi)\le F_\xi(t+\delta)$$

Теперь $\delta\to 0$:

$$F_\xi(t)=\mathbb{E}g_{-\delta}(\xi)=inf\lim_{n\to\infty}F_{\xi_n}(t)=sup\lim_{n\to\infty}F_{\xi_n}(t)=\mathbb{E}g_{+\delta}(\xi)$$

Получаем: $F_\xi(t)=\lim\limits_{n\to\infty}F_{\xi_n}(t)$

■

Теорема 2. $\xi_n\xrightarrow{п.н.}\xi\Rightarrow\xi_n\xrightarrow{p}\xi$

Доказательство. Знаем $P\left(\lim\limits_{n\to\infty}\xi_n=\xi\right)=1$. Заметим вложенность следующих событий для $\forall\varepsilon>0$: $\left\{\lim\limits_{n\to\infty}\xi_n=\xi\right\}\Rightarrow\bigcup\limits_{N=1}^{\infty}\bigcap\limits_{n=N}^{\infty}\{|\xi_n-\xi|<\varepsilon\}$ (это по сути и есть определение предела, что для $\forall\varepsilon$, начиная с некоторого N выполнятется условие). То есть:

$$1=P\left(\lim_{n\to\infty}\xi_n=\xi\right)\le P\left(\bigcup_{N=1}^{\infty}\bigcap_{n=N}^{\infty}\{|\xi_n-\xi|<\varepsilon\}\right)=1$$

Так как последовательность множеств $A_N=\bigcap\limits_{n=N}^{\infty}\{|\xi_n-\xi|<\varepsilon\}$ расширяющаяся ($A_N\subseteq A_{N+1}$), в объединении они дают событие вероятности 1, значит по теореме непрерывности $\lim\limits_{N\to\infty}P(A_N)=1$.

Теперь заметим: $P(A_N) \le P(|\xi_N - \xi| < \varepsilon)$. То есть: $\lim\limits_{N\to\infty} P(|\xi_N - \xi| < \varepsilon) = 1$. Или что то же самое: $\lim\limits_{N\to\infty}(1 - P(|\xi_N - \xi| < \varepsilon)) = \lim\limits_{N\to\infty} P(|\xi_N - \xi| \ge \varepsilon) = 0$ — определение сходимости по вероятности. ■

Теорема 3. *(Теорема Лебега о мажорируемой сходимости) $\xi_n \xrightarrow{p} \xi$ и $|\xi_n|, |\xi| \le \eta$ п. н. (где η — случайная величина, что $\mathbb{E}\eta < \infty$), то $\mathbb{E}\xi_n \to \mathbb{E}\xi$*

Доказательство. Докажем теорему в частном случае, когда $\eta \equiv C$. $\forall\varepsilon > 0 \forall n$ $|\mathbb{E}\xi_n - \mathbb{E}\xi| \le \mathbb{E}|\xi_n - \xi| = \mathbb{E}|\xi_n - \xi| \cdot \mathrm{Ind}_{|\xi_n - \xi| \ge \varepsilon} + \mathbb{E}|\xi_n - \xi| \cdot \mathrm{Ind}_{|\xi_n - \xi| < \varepsilon} \le 2C \cdot P(|\xi_n - \xi| \ge \varepsilon) + \varepsilon \cdot 1$. Так как $\xi_n \xrightarrow{\text{p}} \xi$, то $\exists N \forall n > N$: $P(|\xi_n - \xi| \ge \varepsilon) < \varepsilon$. Тогда получаем: $|\mathbb{E}\xi_n - \mathbb{E}\xi| < 2C \cdot \varepsilon + \varepsilon$. То есть $\mathbb{E}\xi_n \to \mathbb{E}\xi$. ■

Предложение 1. *$\xi_n \xrightarrow{p} \xi \Rightarrow$ для $\forall g$ — непрерывная, $g(\xi_n) \xrightarrow{p} g(\xi)$*

Доказательство. Знаем для любой случайной величины $\forall\varepsilon > 0 \exists A$: $P\left(|\xi| > \dfrac{A}{2}\right) < \varepsilon$. $\exists N \forall n > N$:

$$P(|\xi_n| > A) \le P(|\xi - \xi_n| + |\xi| > A) \le P\left(|\xi - \xi_n| > \frac{A}{2} \vee |\xi| > \frac{A}{2}\right) \le P\left(|\xi - \xi_n| > \frac{A}{2}\right) + P\left(|\xi| > \frac{A}{2}\right) < \varepsilon$$

. Теперь возьмем g, она равномерно непрерывна на $[-A, A]$:

$$\forall\varepsilon > 0 \exists\delta > 0: \ |x - y| < \delta \Rightarrow |g(x) - g(y)| < \varepsilon \ \forall x, y \in [-A, A]$$

Докажем $g(\xi_n) \xrightarrow{\text{p}} g(\xi)$:

$$P(|g(\xi_n) - g(\xi)| \ge \varepsilon) = P(|g(\xi_n) - g(\xi)| \ge \varepsilon \mid \xi_n, \xi \in [-A, A]) \cdot P(\xi_n, \xi \in [-A, A]) +$$

$$+ P(|g(\xi_n) - g(\xi)| \ge \varepsilon \mid \xi_n, \xi \notin [-A, A]) \cdot P(\xi_n, \xi \notin [-A, A])$$

Посмотрев на определение равномерное непрерывности, заметим, что:

$$P(|g(\xi_n) - g(\xi)| \ge \varepsilon \mid \xi_n, \xi \in [-A, A]) \le P(|\xi_n - \xi| \ge \delta)$$

А это уже, так как у нас есть сходимость по вероятности $\to 0$. И заметим, что

$$P(\xi_n, \xi \notin [-A, A]) \le P(|\xi_n| > A) + P(|\xi| > A) < 2\varepsilon$$

То есть $\to 0$ при $n, A \to \infty$ Все, получили, что $P(|g(\xi_n) - g(\xi)| \ge \varepsilon) \to 0$ при $n, A \to \infty$. ■

Следствие 1. *$\xi_n \xrightarrow{p} \xi \Rightarrow \xi_n \xrightarrow{d} \xi$*

Доказательство. Берем предыдущее предложение. Потом используем теорему Лебега для произвольной непрерывной ограниченной g и вспоминаем эквивалентное опеределение сходимости по распределению. ■

Теорема 4. *(Эквивалетное опеределение сходимости почти наверное)*

$$\xi_n \xrightarrow{\text{п.н.}} \xi \Leftrightarrow \forall\varepsilon > 0: \ \lim_{n\to\infty} P\left(\sup_{k \ge n} |\xi_k - \xi| > \varepsilon\right) = 0$$

Доказательство. Рассмотрим следующие события:

$$A_k^\varepsilon = \{|\xi_k - \xi| > \varepsilon\}$$

$$A^\varepsilon = \bigcap_{n=1}^{\infty} \bigcup_{k\geq n} A_k^\varepsilon$$

Заметим, что:

$$\left\{\sup_{k\geq n}|\xi_k - \xi| > \varepsilon\right\} = \{\exists k \geq n: \ |\xi_k - \xi| > \varepsilon\} = \bigcup_{k\geq n} A_k^\varepsilon$$

$$\left\{\lim_{n\to\infty} \xi_n \neq \xi\right\} = \{\exists \varepsilon > 0 \forall n \exists k \geq n: \ |\xi_k - \xi| \geq \varepsilon\} = \bigcup_{m=1}^{\infty} \bigcap_{n}^{\infty} \bigcup_{k\geq n} A_k^{\varepsilon=\frac{1}{m}} = \bigcup_{m=1}^{\infty} A^{\frac{1}{m}}$$

Тогда имеем:

$$\xi_n \xrightarrow{\text{п.н.}} \xi \Leftrightarrow P\left(\lim_{n\to\infty} \xi_n \neq \xi\right) = 0 \Leftrightarrow P\left(\bigcup_{m=1}^{\infty} A^{\frac{1}{m}}\right) = 0 \Leftrightarrow \forall m \ P\left(A^{\frac{1}{m}}\right) = 0 \Leftrightarrow \forall \varepsilon > 0 \ P\left(A^\varepsilon\right) = 0 \Leftrightarrow$$

Теперь заметим вложенность последовательности событий $B_n^\varepsilon = \bigcup_{k\geq n} A_k^\varepsilon$ и, взглянув на определение A^ε, по теореме о непрерывности вероятностной меры продолжаем цепочку:

$$\Leftrightarrow \lim_{n\to\infty} P\left(\bigcup_{k\geq n} A_k^\varepsilon\right) = 0 \Leftrightarrow \lim_{n\to\infty} P\left(\sup_{k\geq n}|\xi_k - \xi| > \varepsilon\right) = 0$$

■

Теперь приведем некоторые примеры, опровергающие остальные следствия сходимостей

Пример 1.

$$\xi_n \xrightarrow{\text{d}} \xi \not\Rightarrow \xi_n \xrightarrow{\text{P}} \xi$$

Пусть $\Omega = \{\omega_1, \omega_2\}$, $P(\{\omega_i\}) = \dfrac{1}{2}$. Определим $\forall n \ \xi_n(\omega_i) = (-1)^i$. Положим $\xi = -\xi_n$. Тогда $\forall f$ непрерыной и ограниченной:

$$\mathbb{E} f(\xi_n) = \frac{f(1) + f(-1)}{2} = \mathbb{E} f(\xi)$$

То есть $\xi_n \xrightarrow{\text{d}} \xi$. Однако $\forall n \ |\xi_n - \xi| = 2$, то есть $\xi_n \overset{\text{P}}{\not\to} \xi$

Пример 2.

$$\xi_n \xrightarrow{\text{P}} \xi \not\Rightarrow \xi_n \xrightarrow{\text{п.н.}} \xi$$

Возьмем $\Omega = [0, 1]$, $\xi_{2^n+p} = \text{Ind}_{\left[\frac{p}{2^n}, \frac{p+1}{2^n}\right]}$, $0 \leq p < 2^n$. Ясно, что $\xi_n \xrightarrow{\text{P}} \xi$, так как

$$\lim_{n\to\infty} P(\xi_n > 0) = \lim_{n\to\infty} \frac{1}{2^{\lfloor \log n \rfloor}} = 0$$

, но $\xi_n \overset{\text{п.н.}}{\not\longrightarrow} 0$, так как для любого исхода ω существует бесконечно много n, что $\xi_n(\omega) = 1$. Теперь осталось посмотреть на теорему 4 и все станет ясно.

Пример 3.

$$\xi_n \xrightarrow{\text{d}} \xi \not\Rightarrow \xi_n \xrightarrow{\text{п.н.}} \xi$$

Если бы следствие имело место, то отсюда вытекало бы, что:

$$\xi_n \xrightarrow{\text{d}} \xi \not\Rightarrow \xi_n \xrightarrow{\text{P}} \xi$$

противоречие

2 Характеристические функции

Определение 2. *Характеристической функцией* случайной величины ξ называется функция:

$$\varphi_\xi(t) = \mathbb{E}e^{it\xi} = \mathbb{E}(\cos(t\xi) + i\sin(t\xi))$$

Предложение 2. *(Свойства характеристических функций)*

1. $\varphi_\xi(0) = 1$, $|\varphi_\xi(t)| \leq 1\ \forall t \in \mathbb{R}$
2. $\varphi_{a\xi+b}(t) = e^{itb}\varphi_\xi(at)$
3. *если* $\xi_1, \cdots \xi_n$ *— независимые случайные величины и* $S = \xi_1 + \cdots + \xi_n$, *то*

$$\varphi_S(t) = \varphi_{\xi_1}(t)\cdots\varphi_{\xi_n}(t)$$

Доказательство. .

1. Понятно, так как в матожидании могут быть только комплексные числа с модулем ≤ 1.
2. $\varphi_{a\xi+b}(t) = \mathbb{E}e^{it(a\xi+b)} = e^{itb}\mathbb{E}e^{i(at)\xi} = e^{itb}\varphi_\xi(at)$
3. $\varphi_S(t) = \mathbb{E}\left(e^{it\xi_1}\cdots e^{it\xi_n}\right) = \mathbb{E}e^{it\xi_1}\cdots\mathbb{E}e^{it\xi_n} = \varphi_{\xi_1}(t)\cdots\varphi_{\xi_n}(t)$

■

Пример 4. Вычислим $\varphi_\xi(t)$, где $\xi \sim N(0,1)$:

$$\varphi_\xi(t) = \mathbb{E}e^{it\xi} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\cos(tx)\,e^{-\frac{x^2}{2}}dx + \frac{i}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\sin(tx)\,e^{-\frac{x^2}{2}}dx$$

Заметим, что $\displaystyle\int_{-\infty}^{+\infty}\sin(tx)\,e^{-\frac{x^2}{2}}dx = 0$, так как это интеграл от нечетной функции по симметричному промежутку. Тогда имеем:

$$\varphi_\xi(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\cos(tx)\,e^{-\frac{x^2}{2}}dx$$

Возьмем производную по t:

$$\varphi'_\xi(t) = -\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}x\sin(tx)\,e^{-\frac{x^2}{2}}dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\sin(tx)\,d\left(e^{-\frac{x^2}{2}}\right) =$$

$$= \frac{1}{\sqrt{2\pi}}\sin(tx)\,e^{-\frac{x^2}{2}}\Big|_{-\infty}^{+\infty} - \frac{t}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\cos(tx)\,e^{-\frac{x^2}{2}}dx = 0 - \frac{t}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}\cos(tx)\,e^{-\frac{x^2}{2}}dx = -t\varphi_\xi(t)$$

Теперь надо решить дифференциальное уравнение:

$$\varphi'_\xi(t) = -t\varphi_\xi(t)$$

$$\frac{\varphi'_\xi(t)}{\varphi_\xi(t)} = -t$$

Интегрируем обе части:

$$\int\frac{d(\varphi_\xi(t))}{\varphi_\xi(t)} = \ln|\varphi_\xi(t)| + C = \int -tdt = -\frac{t^2}{2}$$

$$\varphi_\xi(t) = C'e^{-\frac{t^2}{2}}$$

$$\varphi_\xi(0) = 1 = C' \Rightarrow C' = 1$$

(Ответ никак нулевым быть не может, поэтому, когда мы поделили на хар функцию ничего плохого не произошло)

Предложение 3. *Пусть случайная величина ξ обладает конечным k-тым моментом, то есть $\mathbb{E}|\xi|^k < \infty$. Тогда φ имеет непрерывную k-тую производную и $\varphi^{(k)}(0) = i^k \mathbb{E}\xi^k$*

Доказательство. Заметим, что:

$$\left|\frac{e^{i\Delta t\xi}-1}{\Delta t}\right| \le \frac{\sqrt{(\cos(\Delta t\xi)-1)^2+\sin^2(\Delta t\xi)}}{|\Delta t|} = \frac{\sqrt{2-2\cos(\Delta t\xi)}}{|\Delta t|} = \frac{2\left|\sin\left(\frac{\Delta t\xi}{2}\right)\right|}{\Delta t} \le \frac{2\cdot\frac{|\Delta t\xi|}{2}}{|\Delta t|} = |\xi|$$

Возьмем вместо Δt последовательность $a_n \to 0$, Получим, что последовательность случайных величин, сходящихся почти наверное и ее предел ($i\xi$) ограничены случайной величиной (ξ)с конечным ожиданием, получаем по теорема Лебега:

$$\mathbb{E}\frac{e^{ia_n\xi}-1}{a_n} \to \mathbb{E}i\xi$$

Теперь посчитаем производную хар фукнции:

$$\varphi'(t) = \lim_{\Delta t\to 0}\frac{\varphi(t+\Delta t)-\varphi(t)}{\Delta t} = \lim_{\Delta t\to 0}\mathbb{E}e^{it\xi}\cdot\frac{e^{i\Delta t\xi}-1}{\Delta t} = i\mathbb{E}e^{it\xi}\xi$$

Теперь очевидна непрерывность первой производной и $\varphi'(0) = i\mathbb{E}\xi$, для производных высших порядков аналогично

■

Теорема 5.

$$\xi_n \xrightarrow{d} \xi \Leftrightarrow \forall t \lim_{n\to\infty}\varphi_{\xi_n}(t) = \varphi_\xi(t)$$

Доказательство. $\Rightarrow$
Очевидно по Теореме 1
$\Leftarrow$
Докажем при условиии $\sup\limits_n \mathbb{E}\xi_n^2 \le C < \infty$. По неравенству Чебышева:

$$P(|\xi_n \ge A|) \le \frac{\mathbb{E}\xi^2}{A} \le \frac{C}{A}$$

Пусть f – ограниченная непрерывная функция и $M = \sup|f|$. Из записанного неравенства Чебышева следует, что $\forall\varepsilon > 0\exists A$:

$$P(|\xi_n| \ge A) \le \varepsilon,\ P(|\xi| \ge A) \le \varepsilon$$

Пусть непрерывная ограниченная f_ε совпадает с f на $[-A, A]$, потом от $-A-1$ до $-A$ и от $A+1$ до A она будет прямой из 0 в $f(-A)$ и $f(A)$ соответственно, а дальше будет повторять этот шаблон (периодическая). Заметим, что:

$$|\mathbb{E}f_\varepsilon(\xi_n) - \mathbb{E}f(\xi_n)| < 2M\varepsilon,\ |\mathbb{E}f_\varepsilon(\xi) - \mathbb{E}f(\xi)| < 2M\varepsilon$$

Теперь равномерно приблизим f_ε комбинацией sin, cos (знаем с матана, что периодическую можно так приблизить). А из сходимости хар функции мы знаем, что:

$$\lim_{n\to\infty}\mathbb{E}\sin(\xi_n) = \mathbb{E}\sin(\xi)$$

$$\lim_{n\to\infty}\mathbb{E}\cos(\xi_n) = \mathbb{E}\cos(\xi)$$

То есть получаем $\lim\limits_{n\to\infty}\mathbb{E}f_\varepsilon(\xi_n) = \mathbb{E}f_\varepsilon(\xi)$. В итоге, вспоминая те неравенства с $2M\varepsilon$ и устремляя $n \to \infty$:

$$\mathbb{E}f(\xi) - 4M\varepsilon \le \inf\lim_{n\to\infty}\mathbb{E}f(\xi_n) \le \sup\lim_{n\to\infty}\mathbb{E}f(\xi_n) \le \mathbb{E}f(\xi) + 4M\varepsilon$$

Устремляя $\varepsilon \to 0$, получаем $\lim\limits_{n\to\infty}\mathbb{E}f(\xi_n) = \mathbb{E}f(\xi)$, что доказывает сходимость по распределению.

■

Следствие 2. $\varphi_\xi \equiv \varphi_\eta \Rightarrow F_\xi \equiv F_\eta$

Доказательство. Предыдущая теорема + Теорема 1. ∎

Теорема 6. *(Центральная предельная теорема)*
Пусть ξ_n – последовательность независимых одинаково распределенных случайных величин, причем $\mathbb{E}\xi_1 = \mu$, $\mathbb{D}\xi_1 = \sigma^2$. Тогда $\forall x$:

$$\lim_{n\to\infty} P\left(\frac{\xi_1 + \dots + \xi_n - n\mu}{\sigma\sqrt{n}} \leq t\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{x^2}{2}} dx$$

(справа записана $F(t)$ – функция распределения случайной величины с распределением $N(0,1)$)

Доказательство. Переходя к случайным величинам $\xi_i = \dfrac{\xi_i - \mu}{\sigma}$ дальше будем считать, что $\mathbb{E}\xi_i = 0$ и $\mathbb{D}\xi_i = 1$. Пусть φ – хар функция случаной величины ξ_1. Тогда хар функция случаной величины

$$\frac{\xi_1 + \dots + \xi_n}{\sqrt{n}}$$

равна

$$\varphi_n(t) = \left(\varphi\left(\frac{t}{\sqrt{n}}\right)\right)^n$$

Разложим $\varphi\left(\dfrac{t}{\sqrt{n}}\right)$ в ряд Тейлора в 0 (при $n \to \infty$), помним предложение 3:

$$\varphi\left(\frac{t}{\sqrt{n}}\right) = \varphi(0) + x\varphi'(0) + \dots = 1 + 0 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)$$

Следовательно получаем:

$$\varphi_n(t) = \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n = \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^{-\frac{2n}{t^2}\cdot\left(-\frac{t^2}{2}\right)} \xrightarrow{n\to\infty} e^{-\frac{t^2}{2}}$$

Получили характеристическую функцию нормального распределения, то есть и функции распределения должны совпадать. ∎

3 Неравенство типа Хефдинга-Чернова

Теорема 7. *(Неравенство Хефдинга-Чернова)*
Пусть случайные величины $\xi_1, \cdots \xi_n$ *независимы и* $a_i \le \xi_i \le b_i$. *Тогда для случайной величины* $S_n = \xi_1 + \cdots + \xi_n$ *и для каждого* $t > 0$ *выполнено*

$$P\left(|S_n - \mathbb{E}S_n| \ge t\right) \le 2\exp\left(-\frac{t^2}{4\sum_{i=1}^n (b_i - a_i)^2}\right)$$

Доказательство. Пусть $\eta_i = \xi_i - \mathbb{E}\xi_i$, тогда $|\eta_i| \le b_i - a_i$. Заметим, что для каждого $\lambda > 0$ (просто домножили и взяли экспоненту):

$$P\left(S_n - \mathbb{E}S_n \ge t\right) = P\left(\sum_{i=1}^n \eta_i \ge t\right) = P\left(e^{\lambda\sum\eta_i} \ge e^{\lambda t}\right)$$

Теперь применим неравенство Маркова:

$$P\left(e^{\lambda\sum\eta_i} \ge e^{\lambda t}\right) \le e^{-\lambda t}\mathbb{E}e^{\lambda\sum\eta_i}$$

Вспомним, что $\eta_1, \cdots \eta_n$ независимы:

$$e^{-\lambda t}\mathbb{E}e^{\lambda\sum\eta_i} = e^{-\lambda t}\prod \mathbb{E}e^{\lambda\eta_i}$$

Оценим каждый множитель $\mathbb{E}e^{\lambda\eta_i}$ отдельно. Разложим его в ряд Тейлора:

$$\mathbb{E}e^{\lambda\eta_i} = 1 + \lambda\mathbb{E}\eta_i + \lambda^2\mathbb{E}\eta_i^2 + \sum_{k=3}^{\infty}\frac{1}{k!}\lambda^k\eta^k \le 1 + \lambda^2(b_i - a_i)^2 + \sum_{k=3}^{\infty}\frac{1}{k!}\lambda^k(b_i - a_i)^k$$

Докажем, что при $R > 0$:

$$1 + \frac{1}{2}R^2 + \sum_{k=3}^{\infty}\frac{1}{k!}R^k \le e^{R^2}$$

Если $R > 1$:

$$1 + \frac{1}{2}R^2 + \sum_{k=3}^{\infty}\frac{1}{k!}R^k = 1 + \frac{1}{2}R^2 + \sum_{k=2}^{\infty}\frac{1}{k!}R^{2k}\left(\frac{k!}{(2k-1)!}R^{-1} + \frac{n!}{(2n)!}\right) \le 1 + \frac{1}{2}R^2 + \sum_{k=2}^{\infty}\frac{1}{k!}R^{2k} = e^{R^2}$$

Если же $R \le 1$:

$$1 + \frac{1}{2}R^2 + \sum_{k=3}^{\infty}\frac{1}{k!}R^k \le 1 + \frac{1}{2}R^2 + \sum_{k=3}^{\infty}\frac{1}{2^{k-1}}R^2 = 1 + R^2 \le e^{R^2}$$

Таким образом:

$$P\left(S_n - \mathbb{E}S_n \ge t\right) \le \exp\left(-\lambda t + \lambda^2\sum_{i=1}^n (b_i - a_i)^2\right)$$

Взяв $\lambda = \dfrac{t}{2\sum(a_i - b_i)^2}$ получим желаемое неравенство. Для $P\left(S_n - \mathbb{E}S_n \ge t\right)$ получим такую же оценку, и, объединяя их, получим оценку на модуль, только придется домножить оценку на 2. ■

Следствие 3. *Пусть* $\xi_i \sim Bern(p)$ *– набор* n *независимых случайных величин,* $S_n = \xi_1 + \cdots + \xi_n$, *тогда*

$$P\left(\left|\frac{S_n}{n} - p\right| \ge t\right) \le 2e^{-\frac{nt^2}{4}}$$

Доказательство. Разделим каждую случайную величину на n, тогда $\mathbb{E}\frac{S_n}{n} = p$, а $\sum_{i=1}^{n}(a_i - b_i)^2 = n \cdot \frac{1}{n^2} = \frac{1}{n}$. Подставляем и получаем, нужное неравенство ■

Пример 5. Пусть в ящике какое-то количество черных и белых шаров. Каким должен быть размер выборки, чтобы оценить долю белых шаров с малой погрешностью? Пусть ξ_i – бернуллевская случайная величина, равная 1, если шар белого цвета и 0, если цвет черный. Мы хотим оценить вероятность успеха p. По неравенству выше:

$$P\left(\left|\frac{S_n}{n} - p\right| \geq t\right) \leq 2e^{-\frac{nt^2}{4}} \leq \varepsilon$$

Тогда при размере выборки $n = O\left(\frac{\ln\frac{1}{\varepsilon}}{t^2}\right)$ выборочное среднее приближает реальную долю белых шаров с точностью t с вероятностью более $1-\varepsilon$(то есть вероятность, что наша оценка верна $\geq 1-\varepsilon$)

4 Теоремы непрерывности

Для применения ЦПТ на практике важную роль играют так называемые теоремы о непрерывности.

Предложение 4. *Если последовательность случайных величин* $\xi_n \xrightarrow{d} \xi$, *то для всякой непрерывной* g $g(\xi_n) \xrightarrow{d} g(\xi)$

Доказательство. Следует из Теоремы 1(эквивалентное определение сходимости по распределению). ■

Лемма 1. *Пусть* X, Y, Z *– случайные величины. Тогда*

$$P(X+Y \le t) \le P(X+Z \le t+\varepsilon) + P(|Y-Z| \ge \varepsilon), \ \forall t \in \mathbb{R}, \forall \varepsilon > 0$$

Доказательство. Заметим, что

$$P(X+Y \le t) \le P(X+Y \le t, \ |Y-Z| \le \varepsilon) + P(X+Y \le t, \ |Y-Z| \ge \varepsilon) \le$$

$$\le P(X+Z-\varepsilon \le t) + P(|Y-Z| \ge \varepsilon)$$

■

Предложение 5. *Если* $\xi_n \xrightarrow{P} a = const$ *и* $\eta_n \xrightarrow{d} \eta$, *то* $\xi_n\eta_n \xrightarrow{d} a\eta$, $\xi_n + \eta_n \xrightarrow{d} a + \eta$

Доказательство. Докажем утверждение для суммы. Пусть $\varepsilon > 0$, тогда по лемме имеем:

$$P(\xi_n + \eta_n \le t) \le F_{\eta_n}(t-a+\varepsilon) + P(|\xi_n - a| \ge \varepsilon)$$

и

$$P(\xi_n + \eta_n \le t) \ge F_{\eta_n}(t-a-\varepsilon) - P(|\xi_n - a| \ge \varepsilon)$$

Устремляя сначала $n \to \infty$, а затем $\varepsilon \to 0$. Из сходимости по вероятности получаем, что $P(|\xi_n - a| \ge \varepsilon) \to 0$. Тогда в итоге имеем:

$$\lim \left(F_{\eta_n}(t-a-\varepsilon) - P(|\xi_n - a| \ge \varepsilon)\right) \le \lim F_{\xi_n+\eta_n}(t) \le \lim \left(F_{\eta_n}(t-a+\varepsilon) + P(|\xi_n - a| \ge \varepsilon)\right)$$

$$\lim F_{\xi_n+\eta_n}(t) = F_\eta(t-a) = F_{a+\eta}(t)$$

Теперь докажем утверждение для произведения. Пусть $a = 0$ (случай, когда $a \ne 0$ выводится из суммы $(\xi_n - a)\eta_n$ и $a\eta_n$). Теперь заметим, что $\forall \varepsilon > 0 \forall C > 0$ верно включение:

$$\{|\xi_n\eta_n| > \varepsilon\} \subseteq \{|\eta_n| > C\} \bigcup \left\{|\xi_n| > \frac{\varepsilon}{C}\right\}$$

(Пояснение: это верно, так как пересечение отрицания обоих событий точно приводит к противоречию). Тогда, переходя к вероятностям, получаем:

$$P(|\xi_n\eta_n| > \varepsilon) \le 1 - F_{\eta_n}(C) + F_{\eta_n}(-C) + P\left(|\xi_n| > \frac{\varepsilon}{C}\right)$$

Устремляя сначала $n \to 0$, а затем $C \to \infty$, получаем, что $\xi_n\eta_n \xrightarrow{\mathrm{P}} 0 \Rightarrow \xi_n\eta_n \xrightarrow{\mathrm{d}} 0$. ■

Пример 6. (Выборочная дисперсия)
Пусть задана последовательность независимых и одинаково распределенных случайных величин ξ_i, причем $\mathbb{E}\xi_i = \mu$ и $\mathbb{D}\xi_i = \sigma^2$. Тогда последовательность случайных величин

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{\infty}\left(\xi_i - \overline{\xi_n}\right)^2 \xrightarrow{\text{P}} \sigma^2$$

где $\overline{\xi_n} = \dfrac{\xi_1 + \cdots + \xi_n}{n}$ (умножаем на $\dfrac{1}{n-1}$, а не на $\dfrac{1}{n}$, чтобы $\mathbb{E}s_n^2 = \sigma^2$, то есть таким образом посчитанная диспресия по вероятности сходится к именно тому, чему и надо, в другом случае будет небольшое смещение в $\dfrac{n-1}{n}$ раз, но с увеличение n разница в любом случае будет стрираться). Действительно:

$$s_n^2 = \frac{1}{n-1}\left(\sum_{i=1}^{\infty}\xi_i^2 - 2n\overline{\xi_n}\cdot\frac{1}{n}\sum_{i=1}^{\infty} + n\overline{\xi_n}^2\right) = \frac{1}{n-1}\left(\sum_{i=1}^{\infty}\xi_i^2 - n\overline{\xi_n}^2\right) = \frac{n}{n-1}\cdot\frac{1}{n}\sum_{i=1}^{\infty}\xi_i^2 - \frac{n}{n-1}\overline{\xi_n}^2$$

Теперь заметим, что по ЗБЧ:

$$\frac{1}{n}\sum_{i=1}^{\infty}\xi_i^2 \xrightarrow{\text{P}} \sigma^2 + \mu,\ \overline{\xi_n} \xrightarrow{\text{P}} \mu$$

Получили искомую сходимость

Пример 7. Пусть задана последовательность независимых и одинаково распределенных случайных величин ξ_i, причем $\mathbb{E}\xi_i = \mu$ и $\mathbb{D}\xi_i = \sigma^2 > 0$. Тогда из ЦПТ следует, что:

$$\frac{\xi_1 + \cdots + \xi_n - n\mu}{\sigma\sqrt{n}} = \frac{\sqrt{n}\left(\overline{\xi_n} - \mu\right)}{\sigma} \xrightarrow{\text{d}} \xi \sim N(0,1)$$

Более того, так как $s_n^2 \xrightarrow{\text{P}} \sigma^2 > 0$, то имеет место сходимость по распределению величин:

$$\frac{\sqrt{n}\left(\overline{\xi_n} - \mu\right)}{s_n^2} \xrightarrow{\text{d}} \xi \sim N(0,1)$$

Предложение 6. *Пусть $a, h_n \in \mathbb{R}$, $\lim\limits_{n\to\infty} h_n = 0$ и f — непрерывно дифференцируемая функция на $\mathbb{R}$. Если последовательность случайных величин ξ_n сходится по распределению к ξ, то:*

$$\frac{f(a + h_n\xi_n) - f(a)}{h_n} \xrightarrow{d} f'(a)\,\xi$$

Доказательство. Имеет место равенство:

$$\frac{f(a + h_n\xi_n) - f(a)}{h_n} = \frac{f(a + 1\cdot h_n\xi_n) - f(a + 0\cdot h_n\xi_n)}{h_n} = \frac{1}{h_n}\int_0^1 f(a + th_n\xi_n)\,d(a + th_n\xi_n) =$$

$$= \xi_n\int_0^1 f(a + th_n\xi_n)\,dt$$

Из предложения 5(самый конец) получаем, что $h_n\xi_n \xrightarrow{P} 0$. Также заметим, что функция

$$g(y) = \int_0^1 f(a + ty)dt$$

непрерывна. Следовательно по предложению 4:

$$g(h_n\xi_n) = \int_0^1 f(a + th_n\xi_n)\,dt \xrightarrow{P} g(0) = f'(a)$$

Теперь снова используя предложение 5, получаем нужную сходимость. ∎

Пример 8. Пусть задана последоватеность независисмых и одинаково распределенных случайных величин ξ_i, причем $\mathbb{E}\xi_i = \mu$ и $\mathbb{D}\xi_i = \sigma^2 > 0$. Если h – непрерывно дифференцируемая функция, то

$$\sqrt{n}\left(h\left(\overline{\xi_n}\right) - h\left(\mu\right)\right) \xrightarrow{d} \xi \sim N\left(0, q^2\right), \ q = \sigma h'\left(\mu\right)$$

Действительно, имеем равенство

$$\sqrt{n}\frac{\left(h\left(\overline{\xi_n}\right) - h\left(\mu\right)\right)}{\sigma} = \frac{1}{\sigma} \cdot \frac{h\left(\mu + n^{-\frac{1}{2}}\sigma\zeta_n\right) - h\left(\mu\right)}{n^{-\frac{1}{2}}}$$

где(сходимость по ЦПТ)

$$\zeta_n = \frac{\sqrt{n}\left(\overline{\xi_n} - \mu\right)}{\sigma} \xrightarrow{d} \xi \sim N\left(0, 1\right)$$

Используем предложение 6 и получаем требуемое.

Пример 9. Пусть $\xi_1, \cdots \xi_n$ положительные независимые одинаково распределенные случайные величины, $\mathbb{E}\xi_1 = \mu$, $0 < \mathbb{D}\xi_1 = \sigma^2 < \infty$. Рассмотрим случайную величину $S_n = \xi_1 + \cdots + \xi_n$ и найдем предел в смысле сходимости по распределению у последовательности случайных величин $\sqrt{n}\left(\frac{n}{S_n} - \frac{1}{\mu}\right)$.

Первый способ:

$$\sqrt{n}\left(\frac{n}{S_n} - \frac{1}{\mu}\right) = -\frac{1}{\mu}\frac{n}{S_n}\sqrt{n}\left(\frac{S_n}{n} - \mu\right)$$

По ЦПТ

$$\sqrt{n}\left(\frac{S_n}{n} - \mu\right) \xrightarrow{d} \xi \sim N\left(0, \sigma^2\right)$$

По ЗБЧ

$$\frac{n}{S_n} \xrightarrow{P} \frac{1}{\mu}$$

Таким образом имеем:

$$\sqrt{n}\left(\frac{n}{S_n} - \frac{1}{\mu}\right) \xrightarrow{d} -\frac{1}{\mu}^2 \xi \sim N\left(0, \frac{\sigma^2}{\mu^4}\right)$$

Второй способ:

Пусть $h\left(x\right) = \frac{1}{x}$, тогда

$$\sqrt{n}\left(\frac{n}{S_n} - \frac{1}{\mu}\right) = \sqrt{n}\left(h\left(\frac{S_n}{n}\right) - h\left(m\right)\right) \xrightarrow{d} \xi \sim N\left(0, \sigma^2\left(h'\left(\mu\right)\right)^2\right) = N\left(0, \frac{\sigma^2}{\mu^4}\right)$$

5 Многомерная характеристическая функция и ЦПТ

Определение 3. *Характеристическая функция случайного вектора* $\xi = (\xi_1, \cdots, \xi_m)^T$ определяется равенством

$$\varphi_\xi(x) = \mathbb{E}e^{ix\xi} = \mathbb{E}e^{i\sum_{i=1}^m x_i\xi_i}$$

Предложение 7. $\varphi_\xi \equiv \varphi_\eta \Leftrightarrow \xi$ *и* η *имеют одинаковые распределения*

Доказательство. Заметим, что

$$F_\xi(x_1, \cdots, x_m) = \mathbb{E}I_{\le x_1}(\xi_1)\cdots I_{\le x_m}(\xi_m)$$

По аналогии с одномерным случаем, нам достаточно доказать, что

$$\mathbb{E}g_1(\xi_1)\cdots g_m(\xi_m) = \mathbb{E}g_1(\eta_1)\cdots g_m(\eta_m)$$

для непрерывных периодических функций $g_k(u)$. Такие функции приближаются линейными комбинациями функций вида $e^{i\mu_k u}$. Значит, достаточно проверять совпадение выражений вида

$$\mathbb{E}\exp(i\mu_1\xi_1 + \cdots + i\mu_m\xi_m) = \mathbb{E}\exp(i\mu_1\eta_1 + \cdots + i\mu_m\eta_m)$$

А это у нас есть(это хар функции). ∎

Следствие 4. *Случайные величины* $\xi_1, \cdots, \xi_m$ *независимы тогда и только тогда, когда*

$$\varphi_\xi(y_1, \cdots, y_m) = \varphi_{\xi_1}(y_1)\cdots\varphi_{\xi_m}(y_m)$$

Доказательство. $\Rightarrow$

$$\varphi_\xi(y_1, \cdots, y_m) = \mathbb{E}e^{i(\xi_1 y_1 + \cdots + \xi_m y_m)} = \mathbb{E}e^{i\xi_1 y_1}\cdots e^{i\xi_m y_m} =$$

В силу независимости ξ_i

$$= \mathbb{E}e^{i\xi_1 y_1}\cdots\mathbb{E}e^{i\xi_m y_m} = \varphi_{\xi_1}(y_1)\cdots\varphi_{\xi_m}(y_m)$$

$\Leftarrow$
Сделаем новый вектор $(\eta_1, \cdots, \eta_m)$, так что:

1. Распр $\eta_i =$ распр $\xi_i \quad \forall i \in \{1, \cdots, m\}$
2. $\eta_1, \cdots, \eta_m$ независимы

Определим $F(y) = F_{\eta_1}(y_1)\cdots F_{\eta_m}(y_m)$. Тогда мы знаем, что существует вектор, у которого такая функция распределения, из чего непременно следует независимость $\eta_1, \cdots, \eta_m$
Посчитаем хар. функцию $\eta = (\eta_1, \cdots, \eta_m)$

$$\varphi_\eta(y) = \begin{bmatrix} \text{в силу} \\ \text{нез - сти } \eta_i \end{bmatrix} = \varphi_{\eta_1}(y_1)\cdots\varphi_{\eta_m}(y_m) = \varphi_{\xi_1}(y_1)\cdots\varphi_{\xi_m}(y_m) = \varphi_\xi(y)$$

По предыдущему предложению и независимости $\eta_1, \cdots, \eta_m \Rightarrow \xi_1, \cdots, \xi_m$ независимы. ∎

Теорема 8. *Пусть последовательность независимых одинаково распределенных случайных векторов* $\xi^n = (\xi_1^n, \cdots, \xi_m^n) \in \mathbb{E}$ *имеют конечные*

$$\mathbb{E}\xi_i^1 = \mu_i,\ r_{ij} = cov\left(\xi_i^1, \xi_j^1\right)$$

Тогда величины

$$\eta_i^n = \frac{\xi_i^1 + \cdots + \xi_i^n - n\mu_i}{\sqrt{n}}$$

таковы, что последовательность векторов $\eta^n = (\eta_1^n, \cdots, \eta_m^n) \xrightarrow{d} \eta$*, характеристическая функция которого имеет вид*

$$\varphi_\eta(y) = \exp\left(-\frac{\langle yR, y\rangle}{2}\right),\ R = (r_{ij})$$

Доказательство. В векторной форме:

$$\eta^n = \frac{\sum\limits_{s=1}^{n} \xi^s - \mu n}{\sqrt{n}}$$

Запишем хар. функцию:

$$\varphi_{\eta^n}(t) = \varphi_{(\xi^1-\mu+\cdots+\xi^n-\mu)/\sqrt{n}}(t) = \varphi_{\xi^1-\mu+\cdots+\xi^n-\mu}\left(\frac{t}{\sqrt{n}}\right) =$$

В силу независимости ξ_i и их одинаковой распределенности

$$= \left(\varphi_{\xi^1-\mu}\left(\frac{t}{\sqrt{n}}\right)\right)^n$$

Напоминание. Пусть $F : \mathbb{R}^n \to \mathbb{R}^m$
$x = (x_1, x_2, \cdots x_n) \in \mathbb{R}^n$ и $a = (a_1, a_2, \cdots a_n) \in \mathbb{R}^n$. Тогда ряд Тейлора функции F в точке a это

$$F(x) = F(a) + \sum_{i=1}^{n} \frac{dF(a)}{dx_i}(x_i - a_i) + \frac{1}{2}\sum_{i,j=1}^{n} \frac{d^2F(a)}{dx_i dx_j}(x_i - a_i)(x_j - a_j)$$

$$+ \cdots + \frac{1}{k!}\sum_{i_1,\cdots,i_k=1}^{n} \frac{d^kF(a)}{dx_{i_1}\cdots dx_{i_k}}(x_{i_1} - a_{i_1})\cdots(x_{i_k} - a_{i_k}) + R_k(x-a, a)$$

Разложим в ряд Тейлора:

$$\left(\varphi_{\xi^1-\mu}\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1 + \left\langle \triangledown\varphi_{\xi^1-\mu}(0), \frac{t}{\sqrt{n}}\right\rangle + \frac{1}{2}\cdot D^2\varphi_{\xi^1-\mu}(0)\left\langle\frac{t}{\sqrt{n}}, \frac{t}{\sqrt{n}}\right\rangle + o\left(\frac{1}{n}\right)\right)^n$$

Рассмотрим на 2-мерном случае (на других аналогично):

$$\varphi_\xi(t_1, t_2) = \mathbb{E}e^{it_1\xi_1 + it_2\xi_2}$$

Первая производная:

$$\frac{d}{dt_j}\varphi_\xi(t) = i\mathbb{E}\xi_j e^{it_1\xi_1+it_2\xi_2};\ \frac{d}{dt_j}\varphi_\xi(0) = i\mathbb{E}\xi_j$$

В нашем случае нетрудно понять, что $\triangledown\varphi_{\xi^1-\mu}(0) = 0$, так как у нас ξ это $\xi^1 - \mu$, а $i\mathbb{E}(\xi_i^1 - \mu_i) = 0$
Вторая производная:

$$\frac{d^2}{dt_j t_s}\varphi_\xi(t) = -\mathbb{E}\xi_j\xi_s e^{it_1\xi_1+it_2\xi_2};\ \frac{d^2}{dt_j t_s}\varphi_\xi(0) = -\mathbb{E}\xi_j\xi_s$$

В нашем случае $-\mathbb{E}(\xi_j^1 - \mu_j)(\xi_s^1 - \mu_s) = -r_{js}$
Получаем:

$$\left(1 - \frac{1}{2}\langle Rt, t\rangle \frac{1}{n} + o\left(\frac{1}{n}\right)\right)^n \xrightarrow{n\to\infty} e^{-\frac{1}{2}\langle Rt,t\rangle}$$

■

6 Многомерное нормальное распределение

Определение 4. Случайный вектор $\xi \in \mathrm{Mat}_{m\times 1}$ имеет *нормальное распределение* или является *гауссовским*, если $\forall x \in \mathbb{R}^m$

$$\varphi_\xi(x) = \mathbb{E}e^{ix\xi} = e^{ix\mu - \frac{1}{2}xRx^T}$$

где $\mu = (\mu_1, \cdots, \mu_m)^T$, $R \in \mathrm{Mat}_{m\times m}$ – симметричная неотрицательно определенная. Далее кратко пишем

$$\xi \sim N(\mu, R)$$

Определение 5. Пусть $\xi \in \mathrm{Mat}_{m\times 1}$ случайный вектор. Матрица $R \in \mathrm{Mat}_{m\times m}$ с компонентами $r_{ij} = \mathrm{cov}(\xi_i, \xi_j)$ называется *ковариационной матрицей* вектора ξ. Можно еще написать, что

$$R = \mathrm{cov}(\xi, \xi) = \mathbb{E}(\xi - \mathbb{E}\xi)(\xi - \mathbb{E}\xi)^T$$

Лемма 2. *Симметричная неотрицательно определенная матрица $R \in Mat_{m\times m}$ является ковариционной матрицей случайного вектора-столбца $\xi \in Mat_{m\times 1}$ тогда и только тогда, когда $\forall x, y \in \mathbb{R}^m$*

$$cov(x\xi, y\xi) = x\,cov(\xi, \xi)\,y^T = xRy^T$$

Доказательство. $\Rightarrow$
Распишем по определению ковариации двух случайных величин (у нас именно они):

$$\mathrm{cov}(x\xi, y\xi) = \mathbb{E}\left[(x\xi - \mathbb{E}x\xi)(y\xi - \mathbb{E}y\xi)\right] = \mathbb{E}\left[x(\xi - \mathbb{E}\xi)\,y(\xi - \mathbb{E}\xi)\right] =$$

Транспонируя скаляр, получаем тот же скаляр:

$$= \mathbb{E}\left[x(\xi - \mathbb{E}\xi)\left(y(\xi - \mathbb{E}\xi)\right)^T\right] = \mathbb{E}\left[x(\xi - \mathbb{E}\xi)(\xi - \mathbb{E}\xi)^T y^T\right] = x\mathbb{E}\left[(\xi - \mathbb{E}\xi)(\xi - \mathbb{E}\xi)^T\right] y^T = x\mathrm{cov}(\xi, \xi)\,y^T$$

$\Leftarrow$
Возьмем $x = e_i$, $y = e_j$ (базисные единичные вектора). Тогда из данного равенства получим:

$$\mathrm{cov}(e_i\xi, e_j\xi) = \mathrm{cov}(\xi_i, \xi_j) = e_iRe_j = r_{ij}$$

Следовательно $R = \mathrm{cov}(\xi, \xi)$ по определению.

■

Предложение 8. $\forall A \in Mat_{m\times m}$ $\forall b \in Mat_{m\times 1}$ *и* $\xi \in Mat_{m\times 1}$ *– случайного вектора, верно:*

$$cov(A\xi + b, A\xi + b) = AR_\xi A^T$$

Доказательство. Распишем по определению:

$$\mathrm{cov}(A\xi + b, A\xi + b) = \mathbb{E}\left[(A\xi + b - \mathbb{E}(A\xi + b))(A\xi + b - \mathbb{E}(A\xi + b))^T\right] =$$

$$= \mathbb{E}\left[(A\xi + b - b - \mathbb{E}A\xi)(A\xi + b - b - \mathbb{E}A\xi)^T\right] = \mathbb{E}\left[(A\xi - \mathbb{E}A\xi)(A\xi - \mathbb{E}A\xi)^T\right] =$$

$$= A\mathbb{E}\left[(\xi - \mathbb{E}\xi)(\xi - \mathbb{E}\xi)^T\right]A^T = AR_\xi A^T$$

■

Следствие 5. *Если вектор $\xi \sim N(\mu, R)$, то вектор $A\xi + b \sim N\left(A\mu + b, AR_\xi A^T\right)$*

Теорема 9. *Вектор $\xi \in Mat_{m\times 1}$ имеет нормальное распределение тогда и только тогда, когда $\forall x \in \mathbb{R}^m$ случайная величина $x\xi$ имеет нормальное распределение*

Доказательство. $\Rightarrow$
Если ξ нормальный вектор, то

$$\varphi_{x\xi}(t) = \mathbb{E}e^{itx\xi} = \varphi_\xi(tx) = \exp\left(-\frac{1}{2}txR(tx)^T + itx\mu\right) =$$

$$= \exp\left(-\frac{1}{2}t^2 xRx^T + itx\mu\right)$$

Получили хар функцию нормального распределения $x\xi \sim N\left(x\mu, xRx^T\right)$.
$\Leftarrow$
В обратную сторону:

$$\varphi_\xi(x) = \mathbb{E}e^{ix\xi} = \varphi_{x\xi}(1) = \exp\left(-\frac{1}{2}\mathbb{D}x\xi + i\mathbb{E}x\xi\right) = \exp\left(-\frac{1}{2}\mathrm{cov}(x\xi, x\xi) + ix\mu\right) =$$

$$= \exp\left(-\frac{1}{2}xRx^T + ix\mu\right)$$

, где $R = \mathrm{cov}(\xi, \xi)$, $\mu = \mathbb{E}\xi$. Последний переход вытекает из леммы2. ■

Следствие 6. *Если $\xi \sim N(\mu, R)$, то $R = cov(\xi, \xi)$, $\mu = \mathbb{E}\xi$.*

Следствие 7. *Если вектор $\xi = (\xi_1, \xi_2)$ имеет нормальное распределение и $cov(\xi_1, \xi_2) = 0$, то случайные величины ξ_1 и ξ_2 независимы.*

Доказательство. Пусть

$$\mu = \mathbb{E}\xi = (\mu_1, \mu_2)$$

Заметим, что

$$R = \mathrm{cov}(\xi, \xi) = \begin{pmatrix} \mathrm{cov}(\xi_1, \xi_1) & 0 \\ 0 & \mathrm{cov}(\xi_2, \xi_2) \end{pmatrix}$$

Теперь запишем харфункцию ξ:

$$\varphi_\xi(x_1, x_2) = \exp\left(-\frac{1}{2}\left(x_1^2\mathbb{D}\xi_1 + x_2^2\mathbb{D}\xi_2\right) + i(x_1\mu_1 + x_2\mu_2)\right) =$$

$$= \exp\left(-\frac{1}{2}x_1^2\mathbb{D}\xi_1 + ix_1\mu_1\right) \cdot \exp\left(-\frac{1}{2}x_2^2\mathbb{D}\xi_2 + ix_2\mu_2\right) = \varphi_{\xi_1}(x_1)\varphi_{\xi_2}(x_2)$$

Теперь из следствия4 вытекает независимость ξ_1 и ξ_2 ■

Следствие 8. *Если $\xi \sim N(\mu, R)(\in Mat_{m\times 1})$, то $\exists A \in Mat_{m\times k}$, что $\xi = A\eta + \mu$, где $\eta = (\eta_1, \cdots, \eta_k)^T$, η_i – независимые $N(0,1)$ случайные величины. Причем $AA^T = R$*

Доказательство. Пусть

$$\xi' = \xi - \mu = (\xi_1', \cdots, \xi_m')^T$$

Свели задачу к задачи нахождения ортонормированного базиса $\eta = (\eta_1, \cdots, \eta_k)^T$ в подпространстве $\langle \xi_1', \cdots, \xi_m' \rangle$ со скалярным произведением $(X, Y) = \mathbb{E}XY$. Эта задача решается методом Грама-Шмидта. Получили матрицу перехода A, что

$$\xi - \mu = \xi' = A\eta$$

То есть

$$\xi = A\eta + \mu$$

Осталось пояснить $AA^T = R$:

$$R = \mathrm{cov}(\xi, \xi) = \mathrm{cov}(\xi - \mu, \xi - \mu) = \mathrm{cov}(A\eta, A\eta) = A\mathrm{cov}(\eta, \eta)A^T = AEA^T = AA^T$$

$\mathrm{cov}(\eta, \eta) = E$, так как это ортонормированный базис. ■

Теорема 10. *Если $\xi \sim N(\mu, R)$ (в этой теореме сделаем $\mu := \mu^T \in R^m$) и $detR \neq 0$, случайный вектор ξ имеет плотность*

$$\rho(x) = \frac{1}{(2\pi)^{\frac{m}{2}}\sqrt{detR}} e^{-2^{-1}(x-\mu)R^{-1}(x-\mu)^T}$$

Доказательство. Так как $\xi = A\eta + \mu$, причем $\exists A^{-1}$, то

$$P(\xi \in B) = P(A\eta + \mu \in B) = \frac{1}{\left(\sqrt{2\pi}\right)^m} \int_{A\eta+\mu\in B} e^{-2^{-1}xx^T} dx =$$

$$= \frac{1}{(2\pi)^{\frac{m}{2}}} \int_B e^{-2^{-1}\left(A^{-1}(x-\mu)\right)\left(A^{-1}(x-\mu)\right)^T} d\left(A^{-1}(x-\mu)\right) = \frac{1}{(2\pi)^{\frac{m}{2}} \det A} \int_B e^{-2^{-1}\left(A^{-1}(x-\mu)\right)\left(A^{-1}(x-\mu)\right)^T} dx$$

Остается заметить, что

$$\det AA^T = (\det A)^2 = \det R$$

и что

$$A^{-1}(x-\mu)\left(A^{-1}(x-\mu)\right)^T = (x-\mu)A^{-1}\left(A^{-1}\right)^T (x-\mu)^T =$$
$$= (x-\mu)\left(AA^T\right)^{-1}(x-\mu)^T = (x-\mu)R^{-1}(x-\mu)^T$$

■

Пример 10. Пусть $\xi = (\xi_1, \cdots, \xi_n)^T$, где $\xi_i \sim N\left(0, \sigma^2\right)$ и независимы между собой (или, что то же самое $\xi \sim N\left(0, \sigma^2 E\right)$). Положим

$$\overline{\xi} = \frac{\xi_1 + \cdots + \xi_n}{n},\ \zeta = \left(\xi_1 - \overline{\xi}\right)^2 + \cdots + \left(\xi_n - \overline{\xi}\right)^2$$

(ζ – выборочная диспресия). Покажем, что $\overline{\xi}$ и ζ независимы.
Пусть $U \in \mathrm{Mat}_{n\times n}$ – ортогональная матрица ($UU^T = E$), первая строка которой имеет вид $\left(n^{-\frac{1}{2}}, \cdots, n^{-\frac{1}{2}}\right)$. Тогда координаты вектора $u = U\xi \sim N\left(0, U\sigma^2 EU^T\right) = N\left(0, \sigma^2 E\right)$ являются независимыми. Заметим, что $u_n = \overline{\xi}\sqrt{n}$ и что

$$u^T u = u_1^2 + \cdots + u_n^2 = n\overline{\xi^2} + u_2^2 + \cdots + u_n^2 = \xi^T U^T U \xi = \xi^T \xi = \xi_1^2 + \cdots + \xi_n^2$$

Иначе говоря

$$u_2^2 + \cdots + u_n^2 = \xi_1^2 + \cdots + \xi_n^2 - n\overline{\xi^2}$$

Теперь заметим

$$\zeta = \xi_1^2 + \cdots + \xi_n^2 - 2\sum_{i=1}^n \xi_i \overline{\xi} + n\overline{\xi^2} = \xi_1^2 + \cdots + \xi_n^2 - n\overline{\xi^2} = u_2^2 + \cdots + u_n^2$$

Так как $u_1, \cdots, u_n$ – независимы, то и $\frac{u_1}{\sqrt{n}} = \overline{\xi}$ и $u_2^2 + \cdots + u_n^2 = \zeta$ тоже независимы.
Распределение величины $\chi = \eta_1^2 + \cdots + \eta_n^2$, где η_i независимые с распределением $N(0,1)$, называют распределением хи-квадрат с n степенями свободы и обозначают через χ_n^2. Найдем плотсноть распределения χ_n^2:

$$P(\chi \leq t) = (2\pi)^{-\frac{n}{2}} \int_{\eta_1^2+\cdots+\eta_n^2\leq t} e^{-\frac{x_1^2+\cdots+x_n^2}{2}} dx =$$

Делаем сферическую замену (w_n – площадь n-мерной единичной сферы):

$$= (2\pi)^{-\frac{n}{2}} w_n \int_0^{\sqrt{t}} r^{n-1} e^{-\frac{r^2}{2}} dr$$

Тогда

$$\rho(t) = (2\pi)^{-\frac{n}{2}} w_n \cdot \frac{1}{2} t^{-\frac{1}{2}} t^{\frac{n-1}{2}} e^{-\frac{t}{2}} = \frac{1}{2}(2\pi)^{-\frac{n}{2}} w_n t^{\frac{n-2}{2}} e^{-\frac{t^2}{2}} \mathrm{Ind}_{t>0}$$

7 Условные математические ожидания: дискретный случай

Предположим, что задана дискретная случайная величина

$$\xi(w) = \sum_{i=1}^{n} x_i \mathrm{Ind}_{A_i}(w)$$

Рассмотрим следующую задачу: найти математическое ожидание ξ, если достоверно известно, что произошло событие B, $P(B) > 0$. Поскольку мы знаем, что событие B произошло, то надо пересчитать вероятности A_k с учетом новой информации, а именно, заменить $P(A_k)$ на $P(A_k|B)$. Таким образом, надо вычислить математическое ожидание не относительно исходной вероятностной меры P, а относительно условной вероятности $P(\cdot|B)$.

Определение 6. Имеем:

$$\mathbb{E}(\xi|B) = \sum_{i=1}^{n} x_i P(A_i|B) = \sum_{i=1}^{n} x_i \frac{\mathbb{E}(\mathrm{Ind}_{A_i}\mathrm{Ind}_B)}{P(B)} = \frac{\mathbb{E}(\xi \mathrm{Ind}_B)}{P(B)}$$

Это выражение будем называть *условным математическим ожиданием относительно события* B.

Пусть теперь имеется разбиение

$$\Omega = \bigcup_{k=1}^{N} B_k,\ \ B_k \bigcap B_m = \emptyset,\ \ P(B_k) > 0$$

Обозначим это разбиение $\{B_k\}$ через $\mathcal{B}$. Удобно собрать вместе значения условных математических ожиданий $\mathbb{E}(\xi|B_k)$.

Определение 7. Рассмотрим случайную величину:

$$\Lambda(w) = \sum_{i=1}^{N} \mathrm{Ind}_{B_i}(w)\,\mathbb{E}(\xi|B_i)$$

Если $w \in B_i$, то эта случайная величина выдает среднее значение ξ при условии, что произошло событие B_i. Величину $\Lambda(w)$ называют *условным математическим ожиданием относительно разбиения* $\mathcal{B}$ и обозначают через $\mathbb{E}(\xi|\mathcal{B})$.

Случайную величину

$$P(A|\mathcal{B}) = \mathbb{E}(\mathrm{Ind}_A|\mathcal{B})$$

называют условной вероятностью события A относительно разбиения $\mathcal{B}$. Ясно, что

$$\mathbb{E}(\xi|\mathcal{B}) = \sum_{i=1}^{N} \mathrm{Ind}_{B_i}(w)\,\mathbb{E}(\xi|B_i) = \sum_{i=1}^{N} \mathrm{Ind}_{B_i}(w) \sum_{j=1}^{n} x_j P(A_j|B_i) = \sum_{j=1}^{n}\sum_{i=1}^{N} \mathrm{Ind}_{B_i}(w)\, x_j P(A_j|B_i) =$$

$$= \sum_{j=1}^{n} x_j \sum_{i=1}^{N} \mathrm{Ind}_{B_i}(w)\, P(A_j|B_i) = \sum_{j=1}^{n} x_j \sum_{i=1}^{n} \mathrm{Ind}_{B_i}(w)\,\mathbb{E}\left(\mathrm{Ind}_{A_j}|B_i\right) = \sum_{j=1}^{n} x_j \mathbb{E}\left(\mathrm{Ind}_{A_j}|\mathcal{B}\right) =$$

$$= \sum_{j=1}^{n} x_j P(A_j|\mathcal{B})$$

Пример 11. Рассмотрим важный пример, когда $\mathcal{B} = \left\{B, \overline{B}\right\}$. Тогда

$$P\left(A|\mathcal{B}\right) = \mathrm{Ind}_B\left(w\right) P\left(A|B\right) + \mathrm{Ind}_{\overline{B}}\left(w\right) P\left(A|\overline{B}\right)$$

Если $w \in B$, то $P\left(A|\mathcal{B}\right)(w) = P\left(A|B\right)$

Теорема 11. *Имеют место следующие свойства условного математического ожидания:*

(1) *(линейность)* $\mathbb{E}\left(\alpha\xi + \beta\eta|\mathcal{B}\right) = \alpha\mathbb{E}\left(\xi|\mathcal{B}\right) + \beta\mathbb{E}\left(\eta|\mathcal{B}\right)$

(2) *(монотонность) из* $\xi \le \eta$ *следует* $\mathbb{E}\left(\xi|\mathcal{B}\right) \le \mathbb{E}\left(\eta|\mathcal{B}\right)$

(3) *(аналог формулы полной вероятности)* $\mathbb{E}\left(\mathbb{E}\left(\xi|\mathcal{B}\right)\right) = \mathbb{E}\xi$

(4) *(независимость) если случаная величина* ξ *не зависит от разбиения* $\mathcal{B}$, *т.е. случайные величины* ξ *и* Ind_{B_k} *независимы, то* $\mathbb{E}\left(\xi|\mathcal{B}\right) = \mathbb{E}\xi$

(5) *для всякой случайной величины* $\eta = \sum\limits_{k=1}^{N} c_k Ind_{B_k}$ *верно равенство* $\mathbb{E}\left(\eta\xi|\mathcal{B}\right) = \eta\mathbb{E}\left(\xi|\mathcal{B}\right)$

Доказательство. Свойства (1) и (2) следуют из того, что они верны отдельно для каждого B_k. Свойство (3) проверяется непосредственной подстановкой:

$$\mathbb{E}\left(\mathbb{E}\left(\xi|\mathcal{B}\right)\right) = \mathbb{E}\left(\sum_{i=1}^{N} \mathrm{Ind}_{B_i}\mathbb{E}\left(\xi|B_i\right)\right) = \mathbb{E}\left(\sum_{i=1}^{N} \mathrm{Ind}_{B_i}\frac{\mathbb{E}\left(\xi\mathrm{Ind}_{B_i}\right)}{P\left(B\right)}\right) = \sum_{i=1}^{N}\mathbb{E}\left(\mathrm{Ind}_{B_i}\right)\frac{\mathbb{E}\left(\xi\mathrm{Ind}_{B_i}\right)}{P\left(B_i\right)} =$$

$$= \sum_{i=1}^{N}\mathbb{E}\left(\xi\mathrm{Ind}_{B_i}\right) = \mathbb{E}\xi$$

Обоснуем пункт (4). Так как ξ и Ind_{B_k} независимы, то

$$\mathbb{E}\left(\xi|B_k\right) = \frac{\mathbb{E}\left(\xi\mathrm{Ind}_{B_k}\right)}{P\left(B_k\right)} = \frac{\mathbb{E}\xi\mathbb{E}\mathrm{Ind}_{B_k}}{P\left(B_k\right)} = \mathbb{E}\xi$$

Следовательно,

$$\mathbb{E}\left(\xi|\mathcal{B}\right) = \sum_{k=1}^{N}\mathrm{Ind}_{B_k}\left(w\right)\mathbb{E}\left(\xi|B_k\right) = \sum_{k=1}^{N}\mathrm{Ind}_{B_k}\left(w\right)\mathbb{E}\xi = \mathbb{E}\xi$$

Для обоснования (5) достаточно заметить, что

$$\mathbb{E}\left(\eta\xi|B_k\right) = c_k\mathbb{E}\left(\xi|B_k\right)$$

■

Наиболее типична ситуация, когда разбиение $\mathcal{B}$ появляется посредством некоторой случайной величины

$$\eta = \sum_{i=1}^{N} y_i\mathrm{Ind}_{B_i},$$

где y_i − разлиные числа и $P\left(B_i\right) > 0$.

Определение 8. В этом случае $B_i = \{w : \eta\left(w\right) = y_i\}$ и условное математическое ожидание $\mathbb{E}\left(\xi|\mathcal{B}\right)$ обозначают через $\mathbb{E}\left(\xi|\eta\right)$ и называют *условным математическим ожиданием относительно* η.

Несложно предъявить функцию F (это можно сделать несколькими способами), что

$$\mathbb{E}\left(\xi|\eta\right)(w) = F\left(\eta\left(w\right)\right)$$

Легко видеть, что $F\left(y_i\right) = \mathbb{E}\left(\xi|B_i\right)$.
Можно воспринимать $\mathbb{E}\left(\xi|\eta\right)$ как проекцию ξ на η, а $\mathbb{E}\xi\eta$ как их скаляное произведение.

Лемма 3. *Для условного математического ожидания выполнено*

$$\mathbb{E}\left(\xi f\left(\eta\right)\right) = \mathbb{E}\left[f\left(\eta\right)\mathbb{E}\left(\xi|\eta\right)\right]$$

для произвольной функции f. Кроме того, если для какой-то случайной величины $\zeta = g\left(\eta\right)$ выполнено

$$\mathbb{E}\left(\xi f\left(\eta\right)\right) = \mathbb{E}\left(f\left(\eta\right)\zeta\right),$$

то $\zeta = \mathbb{E}\left(\xi|\eta\right)$ п.н.

Доказательство. По (5) и (3) из теоремы 11:

$$\mathbb{E}\left[f\left(\eta\right)\mathbb{E}\left(\xi|\eta\right)\right] = \mathbb{E}\left[\mathbb{E}\left(f\left(\eta\right)\xi|\eta\right)\right] = \mathbb{E}\left(f\left(\eta\right)\xi\right)$$

Докажем вторую часть:

$$\mathbb{E}\left(f\left(\eta\right)\zeta\right) = \mathbb{E}\left(\xi f\left(\eta\right)\right) = \mathbb{E}\left[f\left(\eta\right)\mathbb{E}\left(\xi|\eta\right)\right]$$

$$\mathbb{E}\left[f\left(\eta\right)\left(\zeta - \mathbb{E}\left(\xi|\eta\right)\right)\right] = 0$$

Так как ζ и $\mathbb{E}\left(\xi|\eta\right)$ – функции от η, то возьмем $f\left(\eta\right) = \zeta - \mathbb{E}\left(\xi|\eta\right)$ и получим:

$$\mathbb{E}\left(\zeta - \mathbb{E}\left(\xi|\eta\right)\right)^2 = 0,$$

то есть $\zeta = \mathbb{E}\left(\xi|\eta\right)$ п.н. ■

Теперь докажем, что $\mathbb{E}\left(\xi|\eta\right)$ и правда является проекцией ξ на η.

Предложение 9. *Пусть $\mathbb{E}\xi^2 < \infty$. Условное математическое ожидание $\mathbb{E}\left(\xi|\eta\right)$ среди всех случайных величин вида $f\left(\eta\right)$ является лучшим среднеквадратическим приближением для ξ, т.е.*

$$\min_{\zeta:\zeta = f(\eta)} \mathbb{E}\left(\xi - \zeta\right)^2 = \mathbb{E}\left[\xi - \mathbb{E}\left(\xi|\eta\right)\right]^2$$

Доказательство. Пусть $\zeta = f\left(\eta\right)$. Так как $\left(\mathbb{E}\left(\xi|\eta\right) - \zeta\right.$ – функция от $\left.\eta\right)$

$$\mathbb{E}\left[\left(\xi - \mathbb{E}\left(\xi|\eta\right)\right)\left(\mathbb{E}\left(\xi|\eta\right) - \zeta\right)\right] = 0,$$

то

$$\mathbb{E}\left(\xi - \zeta\right)^2 = \mathbb{E}\left[\left(\xi - \mathbb{E}\left(\xi|\eta\right)\right) + \left(\mathbb{E}\left(\xi|\eta\right) - \zeta\right)\right]^2 =$$

$$\mathbb{E}\left[\xi - \mathbb{E}\left(\xi|\eta\right)\right]^2 + 2\underbrace{\mathbb{E}\left[\left(\xi - \mathbb{E}\left(\xi|\eta\right)\right)\left(\mathbb{E}\left(\xi|\eta\right) - \zeta\right)\right]}_{=0} + \mathbb{E}\left[\mathbb{E}\left(\xi|\eta\right) - \zeta\right]^2 \geq \mathbb{E}\left[\xi - \mathbb{E}\left(\xi|\eta\right)\right]^2$$

Последнее неравенство достигается взятием $\zeta = \mathbb{E}\left(\xi|\eta\right)$ ■

8 Условные математические ожидания: общий случай

Определение 9. ξ, η – случаные величины. $\mathbb{E}|\xi| < \infty$. Тогда случайная величина вида $F(\eta)$ называется *условным математическим ожиданием* $\mathbb{E}(\xi|\eta)$, если

$$\mathbb{E}[\xi f(\eta)] = \mathbb{E}[\mathbb{E}(\xi|\eta) f(\eta)]$$

для любой ограниченной f. Любые две случаные величины, удовлетворяющие этому условию почти наверное совпадают (лемма 3).

Из этого определения следует, что $\mathbb{E}(\xi|\eta)$ есть наименее отличающаяся от ξ случайная величина вида $F(\eta)$, то есть проекция ξ на η.

Предложение 10. *Предположим, что распределение случайной величины (ξ, η) задано совместной плотностью $\rho_{\xi\eta}(x, y)$. Тогда*

$$\mathbb{E}[g(\xi, \eta)|\eta = y] = \int_{-\infty}^{+\infty} g(x, y) \frac{\rho_{\xi\eta}(x, y)}{\rho_\eta(y)} dx$$

Доказательство. Имеет место цепочка равенств:

$$\mathbb{E}[g(\xi, \eta) f(\eta)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f(y) \rho_{\xi\eta}(x, y) dxdy =$$

$$= \int_{-\infty}^{+\infty} f(y) \underbrace{\int_{-\infty}^{+\infty} g(x, y) \frac{\rho_{\xi\eta}(x, y)}{\rho_\eta(y)}}_{=F(y)} \rho_\eta(y) dy = \mathbb{E}[F(\eta) f(\eta)]$$

Следовательно $F(\eta) = \mathbb{E}(\xi|\eta)$ по определению. ■

Определение 10. *Условной плотностью* случайной величины ξ при условии $\eta = y_0$ называется следующая величина

$$\rho_{\xi|\eta}(x|y_0) = \frac{\rho_{\xi,\eta}(x, y_0)}{\rho_\eta(y_0)}$$

Теперь докажем теорему 11, только для непрерывного случая

Теорема 12. *Имеют место следующие свойства условного математического ожидания:*

(1) (линейность) $\mathbb{E}(\alpha\xi + \beta\eta|\zeta) = \alpha\mathbb{E}(\xi|\zeta) + \beta\mathbb{E}(\eta|\zeta)$

(2) (монотонность) из $\xi \leq \eta$ следует $\mathbb{E}(\xi|\zeta) \leq \mathbb{E}(\eta|\zeta)$

(3) (аналог формулы полной вероятности) $\mathbb{E}(\mathbb{E}(\xi|\eta)) = \mathbb{E}\xi$

(4) (независимость) если случаные величины ξ и η, то $\mathbb{E}(\xi|\eta) = \mathbb{E}\xi$

(5) для всякой случайной величины $\zeta = g(\eta)$ верно равенство $\mathbb{E}(\zeta\xi|\eta) = \zeta\mathbb{E}(\xi|\eta)$

Доказательство. Докажем (1). По определению для любой ограниченной f имеем:

$$\mathbb{E}[f(\zeta) \mathbb{E}(\alpha\xi + \beta\eta|\zeta)] = \mathbb{E}[f(\zeta)(\alpha\xi + \beta\eta)] = \alpha\mathbb{E}[f(\zeta)\xi] + \beta\mathbb{E}[f(\zeta)\eta] =$$

$$= \mathbb{E}[f(\zeta)(\alpha\mathbb{E}(\xi|\zeta) + \beta\mathbb{E}(\eta|\zeta))]$$

Теперь взяв $f(\zeta) = \mathbb{E}(\alpha\xi + \beta\eta|\zeta) - (\alpha\mathbb{E}(\xi|\zeta) + \beta\mathbb{E}(\eta|\zeta))$, получим нужное равенство почти наверное.
Во втором, если перенести все вправо, то по сути надо доказать

$$\xi \geq 0 \Rightarrow \mathbb{E}(\xi|\eta) \geq 0$$

Возьмем функцию

$$f(\eta) = 1 - \operatorname{sgn}(\mathbb{E}(\xi|\eta)) \geq 0$$

Тогда по определению

$$\mathbb{E}[f(\eta)\mathbb{E}(\xi|\eta)] = \mathbb{E}[f(\eta)\xi] \geq 0$$

В то же время

$$\mathbb{E}[f(\eta)\mathbb{E}(\xi|\eta)] = \mathbb{E}[\mathbb{E}(\xi|\eta) - |\mathbb{E}(\xi|\eta)|] \leq 0$$

Следовательно

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

почти наверное, следовательно она почти наверное ≥ 0.
В (3) возьмем $f(\eta) \equiv 1$ и запишем определение.
(4):

$$\mathbb{E}[f(\eta)\xi] = \mathbb{E}f(\eta)\mathbb{E}\xi = \mathbb{E}[f(\eta)\mathbb{E}\xi] \Rightarrow \mathbb{E}(\xi|\eta) = \mathbb{E}\xi$$

(5):

$$\mathbb{E}[f(\eta)\xi\zeta] = \mathbb{E}[f(\eta)\mathbb{E}(\xi\zeta|\eta)]$$

С другой стороны:

$$\mathbb{E}[f(\eta)\xi\zeta] = \mathbb{E}[f(\eta)g(\eta)\zeta] = \mathbb{E}[f(\eta)g(\eta)\mathbb{E}(\xi|\eta)]$$
$$\Rightarrow \mathbb{E}(g(\eta)\xi|\eta) = g(\eta)\mathbb{E}(\xi|\eta)$$

почти наверное

■

Теорема 13. *(Аналог формулы Байеса)*

$$\mathbb{E}(g(\eta)|\xi = x) = \frac{\mathbb{E}\left[g(\eta)\rho_{\xi|\eta}(x,\eta)\right]}{\mathbb{E}\rho_{\xi|\eta}(x,\eta)}$$

для любой ограниченной g

Доказательство. Для любой ограниченной f:

$$\mathbb{E}[f(\xi)g(\eta)] = \mathbb{E}[g(\eta)\mathbb{E}(f(\xi)|\eta)] = \mathbb{E}\int_{-\infty}^{+\infty} f(x)\rho_{\xi|\eta}(x,\eta)g(\eta)\,dx = \int_{-\infty}^{+\infty} f(x)\mathbb{E}\left[\rho_{\xi|\eta}(x,\eta)g(\eta)\right]dx$$

С другой стороны

$$\mathbb{E}[f(\xi)g(\eta)] = \mathbb{E}[f(\xi)\mathbb{E}(g(\eta)|\xi)] = \int_{-\infty}^{+\infty} f(x)\rho_\xi(x)\mathbb{E}[g(\eta)|\xi = x]\,dx$$

(опять подставляя вместо f разность этих величин) получаем равенство почти наверное:

$$\mathbb{E}\left[\rho_{\xi|\eta}(x,\eta)g(\eta)\right] = \rho_\xi(x)\mathbb{E}[g(\eta)|\xi = x]$$

Подставим $g \equiv 1$:

$$\mathbb{E}\left[\rho_{\xi|\eta}(x,\eta)\right] = \rho_\xi(x)$$

В итоге получаем:

$$\mathbb{E}\left[g(\eta)\rho_{\xi|\eta}(x,\eta)\right] = \mathbb{E}\left[\rho_{\xi|\eta}(x,\eta)\right]\mathbb{E}[g(\eta)|\xi = x]$$

■

Пример 12. Пусть (ξ, η) – нормальный вектор. Посчитаем $E[\xi|\eta]$ (помним, что это как проекция ξ на η). Центрируем случайные величны

$$X := \xi - \mathbb{E}\xi$$

$$Y := \eta - \mathbb{E}\eta$$

Найдем ортогональную Z проекцию X на Y (должно быть $X = Z + \mathbb{E}[\xi|\eta]$):

$$Z = X - \frac{\operatorname{cov}(X, Y)}{\operatorname{cov}(Y, Y)} Y = X - \frac{\operatorname{cov}(\xi, \eta)}{\mathbb{D}\eta} Y$$

Выразим ξ:

$$\xi = \mathbb{E}\xi - \frac{\operatorname{cov}(\xi, \eta)}{\mathbb{D}\eta}(\eta - \mathbb{E}\eta) + Z$$

Теперь будем считать условное матожидание (условно от константы или $f(\eta)$ она сама, условное от независимой – его ожидание):

$$\mathbb{E}[\xi|\eta] = \mathbb{E}\xi - \frac{\operatorname{cov}(\xi, \eta)}{\mathbb{D}\eta}(\eta - \mathbb{E}\eta) + \mathbb{E}Z = \mathbb{E}\xi - \frac{\operatorname{cov}(\xi, \eta)}{\mathbb{D}\eta}(\eta - \mathbb{E}\eta)$$

Так как (Y, Z) – нормальный вектор и $\operatorname{cov}(Y, Z) = 0$, то Y и Z независимые случайные величины. $Z \sim N\left(0, \frac{\mathbb{D}\xi\mathbb{D}\eta - [\operatorname{cov}(\xi, \eta)]^2}{\mathbb{D}\eta}\right)$, так как это линейная комбинация случайных величин нормального вектора.
Теперь найдем условную плотность:

$$\mathbb{E}[f(X)|Y = y] = \mathbb{E}\left[f\left(Z + \frac{\operatorname{cov}(\xi, \eta)}{\mathbb{D}\eta} Y\right)|Y = y\right] = \int_{-\infty}^{+\infty} f\left(z + \frac{\operatorname{cov}(\xi, \eta)}{\mathbb{D}\eta} y\right) \frac{\rho_{Z,Y}(z, y)}{\rho_Y(y)} dz =$$

Совместная плотность раскладывается в произведение:

$$= \int_{-\infty}^{+\infty} f\left(z + \frac{\operatorname{cov}(\xi, \eta)}{\mathbb{D}\eta} y\right) \rho_Z(z)\, dz =$$

Делаем замену $u = z + \frac{\operatorname{cov}(\xi, \eta)}{\mathbb{D}\eta} y$:

$$= \int_{-\infty}^{+\infty} f(u)\, \rho_Z\left(u - \frac{\operatorname{cov}(\xi, \eta)}{\mathbb{D}\eta} y\right) du$$

В итоге получаем

$$\rho_{X|Y}(x, y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)}{2\sigma^2}} = \rho_Z(z)$$

, где

$$\mu = \frac{\operatorname{cov}(\xi, \eta)}{\mathbb{D}\eta} y$$

$$\sigma = \frac{\mathbb{D}\xi\mathbb{D}\eta - [\operatorname{cov}(\xi, \eta)]^2}{\mathbb{D}\eta}$$

Статистика

9 Оценки параметров и их свойства

Пусть есть X случайная величина. Мы знаем распределение случайной величины $F_\theta(t)$ с точностью до параметра $\theta \in \Theta$. Задача статистики заключается в том, чтобы оценить параметр θ.

Определение 11. Вектор $(X_1, \cdots, X_n)$ с независимыми компонентами, где все случайные величины $X_i \sim X$, называется *выборкой*. В теории вероятности выборка – набор случайных величин, в статистике выборка – величины, полученные в ходе эксперимента над этими случайными величинами (*простая выборка*, $(x_1, \cdots, x_n) = (X_1(w), \cdots, X_n(w))$).

Определение 12. Случайные величины вида $T_N(X_1, \cdots, X_n)$ называются *статистиками*. Если статистика оценивает θ, то она называется оценкой. Обозначение: $\widehat{\theta}_n(X_1, \cdots, X_n)$

Сформулируем свойства хороших оценок:

1. **Несмещенность**. $\mathbb{E}\widehat{\theta}(X_1, \cdots, X_n) = \theta$. Заметим, что матожидание здесь мы считаем по распределению F_θ, то есть хочется, чтобы оценка в среднем была верна.

   Пример 13. Пусть $\mathbb{E}X_i = \mu$, μ – единственный параметр. Возьмем оценку $\overline{X}$. Она будет несмещенной: $\mathbb{E}\overline{X} = \mu$. А если, например, параметризована дисперсия σ^2 и в качестве оценки берем смещенную выборочную дисперсию, то $\mathbb{E}s_n^2 = \dfrac{n}{n-1}\sigma^2 \neq \sigma^2$, получили смещенную оценку.

   Пример 14. Пусть $X \sim Bern(p)$. Тогда для нахождения несмещенной статистики, надо чтобы выполнялось равенство

   $$\mathbb{E}T(X_1, \cdots, X_n) = \sum_{k=0}^{n}\left[p^k(1-p)^{n-k}\sum_{k \text{ успехов}} T(\varepsilon_1, \cdots, \varepsilon_n)\right] = \theta.$$

   Если мы опеределим $\theta = \dfrac{1}{p}$, то несмещенных статистик не существует. Несмещенность должна выполняться $\forall\theta$, но, если $p \to 0$, то левая часть будет стремиться к некоторой константе, а правая $\theta \to +\infty$.

   Пример 15. Теперь приведем пример бессмыленной несмещенной оценки. Пусть $N = 1$, $X \sim Pois(\theta)$, где $0 < \theta < 1$. Тогда

   $$\mathbb{E}T(X_1) = e^{-\theta}\sum_{k=0}^{+\infty}\frac{T(k)\,\theta^k}{k!} = \theta$$

   $$\sum_{k=0}^{+\infty}\frac{T(k)\,\theta^k}{k!} = \theta e^\theta = \sum_{k=0}^{+\infty}\frac{\theta^{k+1}}{k!}$$

   Приравнивая соответственные члены, получаем, что $T(k) = k$, то есть $T(X_1) = X_1$. Наша оценка никак не пересекается с множеством параметров, она бессмысленна.

2. **Состоятельность**. Оценка называется состоятельной, если $\widehat{\theta}_n(X_1, \cdots, X_n) \xrightarrow{P} \theta$. Это значит, на больших объемах данных мы получаем более точную оценку (ЗБЧ).

   Утверждение 1. *Если $\widehat{\theta}_n$ – несмещенная оценка θ и $\mathbb{D}\widehat{\theta}_n \to 0$, то $\widehat{\theta}_n$ – состоятельная.*

Доказательство. Запишем неравенство Чебышева:

$$P\left(\left|\widehat{\theta}-\theta\right|\geq\varepsilon\right)\leq\frac{\mathbb{D}\widehat{\theta}_n}{\varepsilon}$$

Все ясно. ■

Записанное неравенство наталкивает на мысль, что при маленькой дисперсии, наши оценки будут более вероятно слабо разбросаны вокруг θ.

3. **Оптимальность**. Несмещенная оценка $\widehat{\theta}_n$ называется оптимальной, если $\mathbb{D}\widehat{\theta}_n \leq \mathbb{D}\theta_n^*$, $\forall$ несмещенных оценок θ_n^*

   Утверждение 2. *Не существует в общем случае самой оптимальной оценки в множестве всех оценок.*

   Доказательство. Допустим, такая оценка существует, что

   $$\mathbb{E}\left[\widehat{\theta}_n-\theta\right]\leq\mathbb{E}\left[\theta_n^*-\theta\right]$$

   для $\forall\ \theta\in\Theta$. Но если мы возьмем $\theta_n^*\equiv A$, то получим, что $\widehat{\theta}_n = A$ почти наверное по распределению F_A. Теперь возьмем $\theta_n^* = B\neq A$ и получим, что $\widehat{\theta}_n = B$ почти наверное по распределению F_B. Получили противоречие (противоречие получаем из-за оговорки в общем случае, что $\exists C : P_A\left(C\right),\ P_B\left(C\right)>0$) ■

4. **Сильная состоятельность**. $\lim\limits_{n\to+\infty}\widehat{\theta}_n\left(X_1,\cdots,X_n\right)=\theta$ почти наверное на распределении F_θ

5. **Асимптотическая нормальность**. Оценка называется асимптотически нормальной, если

   $$\sqrt{n}\left(\widehat{\theta}_n\left(X_1,\cdots,X_n\right)-\theta\right)\xrightarrow{d}N\left(0,\sigma\left(\theta\right)\right)$$

   Это условие влечет состоятельность и позволяет оценивать вероятности событий $\alpha<\widehat{\theta}_n\left(X_1,\cdots,X_n\right)<\beta$ через нормальные распределения.

   Предложение 11. *Если оценка асимптотически номальная, то она состоятельна.*

   Доказательство. $\widehat{\theta}_n\left(X_1,\cdots,X_n\right)-\theta\xrightarrow{d}0$, $\theta\xrightarrow{P}\theta$
   $\Rightarrow\widehat{\theta}_n\left(X_1,\cdots,X_n\right)\xrightarrow{P}\theta$ ■

Предложение 12. *Если эффективная оценка существует, то она единственна.*

Доказательство. Пусть θ_1^* и θ_2^* две эффективные оценки. Тогда оценка $\dfrac{\theta_1^*+\theta_2^*}{2}$ также является несмещенной. Кроме того,

$$\mathbb{D}\frac{\theta_1^*+\theta_2^*}{2}=\theta_1^*\mathbb{D}\theta_1^*=\mathbb{D}\theta_1^*=a=\frac{1}{4}a^2+\frac{1}{2}\mathrm{cov}\left(\theta_1^*,\theta_2^*\right)+\frac{1}{4}a^2\geq a^2$$

то есть $\mathrm{cov}\left(\theta_1^*,\theta_2^*\right)\geq a^2$. Тогда получим

$$\mathbb{D}\left[\theta_1^*-\theta_2^*\right]=2a^2-2\mathrm{cov}\left(\theta_1^*,\theta_2^*\right)\leq 0$$

Иначе говоря, $\theta_1^*=\theta_2^*$ п.н. ■

10 Метод моментов

У нас есть некоторая простая выборка $(x_1, \cdots, x_n)$, взятая из распределения F_θ, мы хотим по ней получить состоятельную оценку для θ.

Пусть g – непрерывная функция (методом моментов он называется, потому что обычно в качестве g берется степенная функция), что $\mathbb{E}_\theta |g(X)| < \infty$. Тогда введем функцию $f(\theta) := \mathbb{E}_\theta g(X)$. Матожидание является функцией от θ, потому что неизвестно нам только распределение, которое зависит от θ. Теперь допустим, что $\exists f^{-1}$ на Θ. Тогда легко найти θ:

$$\theta = f^{-1}(f(\theta)) = f^{-1}(\mathbb{E}_\theta g(X))$$

Но так как распределения мы не знаем, у нас есть только простая выборка из него, то поступаем так. Знаем, что по ЗБЧ

$$\frac{g(X_1) + \cdots + g(X_n)}{n} \xrightarrow{p} \mathbb{E}_\theta g(X) = f(\theta)$$

Тогда по теоремам непрерывности (предложение 1):

$$f^{-1}\left(\frac{g(X_1) + \cdots + g(X_n)}{n}\right) \xrightarrow{p} \theta$$

Тогда

$$\widehat{\theta}_n(X_1, \cdots, X_n) = f^{-1}\left(\frac{g(x_1) + \cdots + g(x_n)}{n}\right)$$

1. Оценка получилась состоятельная.

2. Покажем асимптотическую нормальность оценки. По ЦПТ знаем, что

$$\xi_n = \sqrt{n}\left(\frac{g(X_1) + \cdots + g(X_n)}{n} - \mathbb{E}_\theta g(X)\right) \xrightarrow{p} \xi \sim N(0, \mathbb{D}_\theta g(X))$$

Обозначим

$$a = \mathbb{E}_\theta g(X), \ \sigma^2 = \mathbb{D}_\theta g(X)$$

Тогда, предполагая, что g^{-1} непрерывно дифференцируема, по теореме непрерывности (предложение 6), получаем:

$$\sqrt{n}\left(\widehat{\theta}_n(X_1, \cdots, X_n) - \theta\right) = \sqrt{n}\left(g^{-1}\left(a + \frac{\sigma}{\sqrt{n}}\xi_n\right) - g^{-1}(a)\right) \xrightarrow{d} \xi\left(h^{-1}\right)'(a)$$

То есть, если $\left(h^{-1}\right)'(a) \neq 0$, то оценка $\widehat{\theta}_n(X_1, \cdots, X_n)$ асимптотически нормальна с коэффициентом $\sigma(\theta) = (h'(a))^2\sigma^2$

Пример 16. Возьмем выборку $(X_1, \cdots, X_n)$ из равномерного распределения $U_{[0,\theta]}$, где $\theta > 0$. Найдем оценку по методу моментов при $f(x) = x$:

$$\mathbb{E}_\theta f(X) = \frac{\theta}{2} \Rightarrow \theta = 2\mathbb{E}_\theta f(X)$$

Получаем оценку

$$\widehat{\theta}_n = 2\overline{X}$$

11 Информация Фишера и неравенство Рао-Крамера

Определение 13. *Несмещенной оценкой нуля* называется такая статистика, $U(X_1, \cdots, X_n)$, для которой верно $\mathbb{E}_\theta U = 0$, $\forall\theta$ (можно опять провести аналогии со скалярным произведением)

Утверждение 3. $\widehat{\theta}$ *– оптимальная (минимальная по дисперсии)* $\Leftrightarrow \mathbb{E}\left[\widehat{\theta} U\right] = 0$ $\forall U$ *– несмещеннных оценок нуля.*

Доказательство. $\Rightarrow$
Заметим, что $\forall\lambda$ верно $\mathbb{E}\left(\widehat{\theta} + \lambda U\right) = \widehat{\theta}$ и из оптимальности $\mathbb{D}\widehat{\theta} \le \mathbb{D}\left(\widehat{\theta} + \lambda U\right)$. Раскрыв дисперсию, получим

$$\mathbb{D}\widehat{\theta} \le \mathbb{D}\widehat{\theta} + \lambda^2 \mathbb{D}U + 2\lambda \text{cov}\left(\widehat{\theta}, U\right)$$

$$2\lambda \text{cov}\left(\widehat{\theta}, U\right) + \lambda^2 \mathbb{D}U \ge 0$$

Это верно только, когда $\text{cov}\left(\widehat{\theta}, U\right) = 0$ (потому что иначе есть λ, в которых неравенство не выполняется). Осталось записать

$$\text{cov}\left(\widehat{\theta}, U\right) = \mathbb{E}\left[\left(\widehat{\theta} - \theta\right) U\right] = \mathbb{E}\left[\widehat{\theta} U\right] = 0$$

$\Leftarrow$
Возьмем другую несмещенную оценку $\tilde{\theta}$. Рассмотрим ее дисперсию:

$$\mathbb{D}\tilde{\theta} = \mathbb{E}\left(\tilde{\theta} - \theta\right)^2 = \mathbb{E}\left(\left(\tilde{\theta} - \widehat{\theta}\right) + \left(\widehat{\theta} - \theta\right)\right)^2 = \mathbb{E}\left(\tilde{\theta} - \widehat{\theta}\right)^2 + 2\mathbb{E}\left[\left(\tilde{\theta} - \widehat{\theta}\right)\left(\widehat{\theta} - \theta\right)\right] + \mathbb{E}\left(\widehat{\theta} - \theta\right)^2$$

Заметим, что

$$\mathbb{E}\left[\left(\tilde{\theta} - \widehat{\theta}\right)\left(\widehat{\theta} - \theta\right)\right] = \mathbb{E}\left[\left(\tilde{\theta} - \widehat{\theta}\right)\widehat{\theta}\right] = 0$$

Последнее равенство вытекает, из условия, если взять $U = \tilde{\theta} - \widehat{\theta}$. Тогда получаем, что

$$\mathbb{D}\tilde{\theta} = \mathbb{D}\widehat{\theta} + \mathbb{D}\left(\tilde{\theta} - \widehat{\theta}\right) \ge \mathbb{D}\widehat{\theta}$$

То есть $\widehat{\theta}$ оптимальна.

∎

Пример 17. Пусть $X \sim Bern(\theta)$. Рассмотрим оценку $\overline{X}$. Она несмещенная и по ЗБЧ состоятельная. Докажем оптимальность. По доказанному утверждению, нужно убедиться, что $\mathbb{E}\overline{X}U = 0$ $\forall U$ – несмещенных оценок нуля. Запишем $\forall\theta$

$$0 = \mathbb{E}U(X_1, \cdots, X_n) = \sum_{\varepsilon_i \in \{0,1\}} U(\varepsilon_1, \cdots, \varepsilon_n)\, \theta^{\sum \varepsilon_i} (1-\theta)^{n - \sum \varepsilon_i} =$$

Сгруппируем

$$= \sum_{k=0}^{n} \left[\theta^k (1-\theta)^{n-k} \sum_{\varepsilon_i : \sum \varepsilon = k} U(\varepsilon_1, \cdots, \varepsilon_n)\right] = \sum_{k=0}^{n} \theta^k (1-\theta)^{n-k} Q_k = \frac{1}{\theta^n} \sum_{k=0}^{n} \left(\frac{\theta}{1-\theta}\right)^k Q_k = 0$$

При $\theta \in (0,1)$, $\dfrac{\theta}{1-\theta} \in (0, +\infty)$ – на этом промежутке многочлен равен нулю, тогда получается, что $Q_k = 0$. Теперь можем проверить $\overline{X}$ на оптимальность:

$$\mathbb{E}\overline{X}U = \sum_{k=0}^{n} \left[\frac{k}{n} \sum_{\varepsilon_i : \sum \varepsilon = k} U(\varepsilon_1, \cdots, \varepsilon_n)\, \theta^k (1-\theta)^{n-k}\right] = \sum_{k=0}^{n} \frac{k}{n} Q_k \theta^k (1-\theta)^{n-k} = 0$$

Получили оптимальность $\overline{X}$.

В методе максимального правдоподобия мы более подробно разберем понятие информации, а пока обойдемся сухими определениями.

Определение 14. Пусть $X = (X_1, \cdots, X_n)$ простая выборка из распределения с плотностью ρ_θ, тогда функция $p(x, \theta) = \rho_\theta(x_1) \cdots \rho_\theta(x_n)$ называется *функцией правдоподобия*, однако с произведением работать неудобно, поэтому прологарифмируем: $L(x, \theta) = \sum\limits_{j=1}^{n} \ln \rho_\theta(x_j)$

Определение 15. Величина $I(\theta) = \mathbb{E}_\theta \left[\dfrac{\partial L}{\partial \theta}\right]^2$ называется *информацией Фишера.*

В дальнейших рассуждениях предполагается выполнение условий регулярности для функции правдоподобия, что она непрерывна и дифференцируема по θ и что операции дифференцирования и интегрирования перестановочны ($\int f' dx = \left(\int f dx\right)'$)

Предложение 13. *(аддитивность информации Фишера) Верно равенство*

$$I(\theta) = ni(\theta),$$

где $i(\theta)$ – информация Фишера для выборки из одного элемента.

Доказательство. Заметим, что

$$\mathbb{E}\frac{\partial L}{\partial \theta} = \mathbb{E}\left[\sum_{j=1}^{n} \frac{\rho_\theta'(x_j)}{\rho_\theta(x_j)}\right] = \sum_{j=1}^{n} \int \frac{\rho_\theta'(x_j)}{\rho_\theta(x_j)} \rho_\theta(x_j)\, dx = \sum_{j=1}^{n} \left(\int_{-\infty}^{+\infty} \rho_\theta(x_j)\, dx\right)' = \sum_{j=1}^{n} (1)' = 0$$

Поэтому

$$I(\theta) = \mathbb{E}\left[\frac{\partial L}{\partial \theta}\right]^2 = \mathbb{D}\left[\frac{\partial L}{\partial \theta}\right] = \mathbb{D}\left[\sum_{j=1}^{n} \frac{\rho_\theta'(x_j)}{\rho_\theta(x_j)}\right] = \sum_{j=1}^{n} \mathbb{D}\left[\frac{\rho_\theta'(x_j)}{\rho_\theta(x_j)}\right] = \sum_{j=1}^{n} i(\theta) = ni(\theta)$$

∎

Рассмотрим неравенство, позволяющее оценивать оптимальные оценки.

Теорема 14. *(неравенство Рао-Крамера) Выполняются условия регулярности и $I(\theta) > 0$ (если $= 0$, то это значит, что мы от эксперимента не получаем никакой информации и все бессмысленно). Пусть $\theta_n(X)$ – произвольная несмещенная оценка статистика $\tau(\theta)$. Тогда верно*

$$\mathbb{D}\theta_n(X) \geq \frac{(\tau'(\theta))^2}{I(\theta)}$$

Доказательство. Из несмещенности знаем

$$\mathbb{E}\theta_n(X) = \int \theta_n(x_1, \cdots, x_n)\, p(x_1, \cdots, x_n, \theta)\, dx_1 \cdots dx_n = \tau(\theta) = A$$

Также знаем

$$\int p(x_1, \cdots, x_n, \theta)\, dx_1 \cdots dx_n = 1 = B$$

Рассмотрим выражение $A_\theta' - \theta B_\theta'$ ($= \tau'(\theta)$ по понятным причинам):

$$\tau'(\theta) = \int (\theta_n(x) - \theta)\, p'(x, \theta)\, dx = \int (\theta_n(x) - \theta) \frac{p'(x, \theta)}{p(x, \theta)} p(x, \theta)\, dx = \mathbb{E}\left[(\theta_n - \theta) \frac{\partial L}{\partial \theta}\right]$$

Из неравенства Коши-Буняковского получим:

$$\tau'(\theta) = \mathbb{E}\left[(\theta_n - \theta)\frac{\partial L}{\partial \theta}\right] \leq \sqrt{\mathbb{D}\theta_n}\sqrt{\mathbb{E}\left[\frac{\partial L}{\partial \theta}\right]^2}$$

$$(\tau'(\theta))^2 \leq I(\theta)\,\mathbb{D}\theta_n$$

■

Замечание 1. Можно переписать неравенство в виде

$$\mathbb{D}\theta_n \leq \frac{(\tau'(\theta))^2}{ni(\theta)}$$

То есть как бы мы не старались, точность оценки будет порядка $\frac{1}{n}$.

Метод максимального правдоподобия

Во-первых рассмотрим пример, объясняющий понятие информации.

Пример 18. Пусть есть монетка M с $P(M=0)=q$, $P(M=1)=p$. Подбрасывается монетка, и, в зависимости от исхода, генерируется случайное событие из распределение ρ_0 или ρ_1. Таким образом, получили случайную величину ξ. Изначально шансы, что случайная величина была взята из первого распределения против нулевого $\frac{p}{q}$. Несложно вычилисть шансы после одной генерации случайного события. Воспользовавшись аналогом формулы Байеса, получим

$$P(M=1|\xi=x)=\frac{\rho_1(x)\,p}{\rho_1(x)\,p+\rho_0(x)\,q},\ P(M=0|\xi=x)=\frac{\rho_0(x)\,q}{\rho_1(x)\,p+\rho_0(x)\,q}$$

То есть теперь шансы $\frac{\rho_1(x)\,p}{\rho_0(x)\,q}$. Они изменились. Пусть *информация* – это то, как изменились шансы. Как можно измерить информацию, полученную в ходе эксперимента? Например, возьмем разницу ln от шансов:

$$\ln\frac{\rho_1(x)\,p}{\rho_0(x)\,q}-\ln\frac{p}{q}=\ln\frac{\rho_1(x)}{\rho_0(x)}$$

Ну и чтобы для каждой величины не вычислять выражение, усредним по распределению ρ_1 (можно было и по ρ_0 усреднять) и будем называть это выражение информацией:

$$\tilde{I}=\int\ln\frac{\rho_1(x)}{\rho_0(x)}\rho_1(x)\,dx=\int\ln\rho_1(x)\,\rho_1(x)\,dx-\int\ln\rho_0(x)\,\rho_1(x)\,dx$$

Утверждение 4. *(неравенство информации) Пусть ρ_0,ρ_1 – вероятностные плотности. Тогда*

$$\int\ln\rho_1(x)\,\rho_1(x)\,dx\geq\int\ln\rho_0(x)\,\rho_1(x)\,dx,$$

то есть $\tilde{I}\geq 0$. Причем равенство выполняется $\Leftrightarrow \rho_0\equiv\rho_1$

Доказательство. Хотим доказать

$$\int\ln\frac{\rho_0(x)}{\rho_1(x)}\rho_1(x)\,dx\leq 0$$

Знаем, что $\ln x\leq x-1$. Тогда

$$\ln\frac{\rho_0(x)}{\rho_1(x)}\rho_1(x)\leq\rho_0(x)-\rho_1(x)$$

Подставляя в интеграл, получим, что

$$\int\ln\frac{\rho_0(x)}{\rho_1(x)}\rho_1(x)\,dx\leq 1-1=0$$

Теперь посмотрим, когда выполняется равенство. Оно выполнятется, когда $\ln x=x-1$, то есть

$$\int\left(\frac{\rho_0(x)}{\rho_1(x)}-1+\ln\frac{\rho_0(x)}{\rho_1(x)}\right)\rho_1(x)\,dx=0$$

Тогда

$$\frac{\rho_0(x)}{\rho_1(x)}-1+\ln\frac{\rho_0(x)}{\rho_1(x)}=0$$

почти наверное (так как это выражение ≥ 0), то есть $\rho_0=\rho_1$ почти наверное. ∎

То есть, если $\tilde{I} = 0$, то от новой информации ничего не меняется, она бессмысленна. Также добавим, что $\tilde{I}$ в реальности оценивает расстояние между распределениями, это называется энтропией распределения ρ_1 относительно ρ_0.

Применим полученные знания для построения оценки. Пусть $X = (X_1, \cdots, X_n)$ – выборка из распределения с параметром $\theta = \theta_1$. Рассмотрим функцию $W(\theta) = \mathbb{E}_{\theta_1} \ln \rho_\theta (X_1)$, ее называют истинным правдоподобием. Заметим из неравенства информации, что максимум истинного правдоподобия достигается только на $\theta = \theta_1$. Само матожидание вычислять мы не умеем, но по ЗБЧ значем, что

$$\frac{1}{n} L(X, \theta) = \frac{1}{n} \sum \ln \rho_\theta (x_j) \xrightarrow{P} W(\theta)$$

Тогда *оценкой максимального правдоподобия* называется оценка $\widehat{\theta} = \mathrm{argmax}_\theta L(X, \theta)$. Если посмотреть, что мы максимизируем функцию L, и вспомнить, что это по сути функция правдоподобия, то выходит, что мы максимизируем вероятность получения выборки X из распределения с параметром $\widehat{\theta}$. Еще заметим, что $W(\theta_1) - W(\theta)$ это величина из примера с монетой, то есть количество информации, которое дало одно наблюдение.

Утверждение 5. *(состоятельность оценки методом максимального правдоподобия) Дана выборка $X = (X_1, \cdots, X_n)$. Пусть $\theta \in (\alpha, \beta)$ и на этом промежутеке у функции $L(X_1, \cdots, X_n, \theta)$ $\exists! \widehat{\theta}_n$ – точка локального максимума, тогда $\widehat{\theta}_n \xrightarrow{P} \theta_1$, то есть оценка состоятельна.*

Доказательство. По ЗБЧ $\frac{1}{n} L(X, \theta) \xrightarrow{P} W(\theta)$. Хотим $\forall \delta > 0$: $\lim\limits_{n\to\infty} P\left(\left|\widehat{\theta}_n - \theta\right| \geq \delta\right) = 0$. Из информационного неравенства заметим, что $W(\theta_1 - \delta) < W(\theta_1) > W(\theta_1 + \delta)$. Знаем, что $\frac{1}{n} L(X, \theta_1 - \delta)$, $\frac{1}{n} L(X, \theta_1)$, $\frac{1}{n} L(X, \theta_1 + \delta)$ сходятся по вероятности соответственно к $W(\theta_1 - \delta)$, $W(\theta_1)$, $W(\theta_1 + \delta)$. Докажем, что сохраняются те же знаки неравенства.

Пусть $\varepsilon = W(\theta_1) - W(\theta_1 - \delta) > 0$. Знаем, что

$$P\left(\left|\frac{1}{n} L(X, \theta_1 - \delta) - W(\theta_1 - \delta)\right| \geq \frac{\varepsilon}{10}\right) \xrightarrow{n\to\infty} 0$$

$$P\left(\left|\frac{1}{n} L(X, \theta_1) - W(\theta_1)\right| \geq \frac{\varepsilon}{10}\right) \xrightarrow{n\to\infty} 0$$

Тогда

$$P\left(\frac{1}{n} L(X, \theta_1 - \delta) \geq \frac{1}{n} L(X, \theta_1)\right) \leq$$

$$\leq P\left(\left\{\left|\frac{1}{n} L(X, \theta_1 - \delta) - W(\theta_1 - \delta)\right| \geq \frac{\varepsilon}{10}\right\} \bigcup \left\{\left|\frac{1}{n} L(X, \theta_1) - W(\theta_1)\right| \geq \frac{\varepsilon}{10}\right\}\right) \leq$$

$$\leq P\left(\left|\frac{1}{n} L(X, \theta_1 - \delta) - W(\theta_1 - \delta)\right| \geq \frac{\varepsilon}{10}\right) + P\left(\left|\frac{1}{n} L(X, \theta_1) - W(\theta_1)\right| \geq \frac{\varepsilon}{10}\right) \xrightarrow{n\to\infty} 0$$

Допустим противное, что в первом неравенстве может стоять $< \frac{\varepsilon}{10}$ в обоих случаях. Тогда выходит

$$\varepsilon = W(\theta_1) - W(\theta_1 - \delta) \leq \frac{1}{n} L(X, \theta_1) + \frac{\varepsilon}{10} - W(\theta_1 - \delta) \leq \frac{1}{n} L(X, \theta_1) + \frac{\varepsilon}{10} - \frac{1}{n} L(X, \theta_1 - \delta) + \frac{\varepsilon}{10} \leq \frac{\varepsilon}{5}$$

Получили противоречие, то есть все ок. То есть вышло, что $\frac{1}{n} L(X, \theta_1 - \delta) < \frac{1}{n} L(X, \theta_1)$ почти наверное. Второе неравенство доказывается аналогично.

Получив желаемые неравенства, делаем вывод, что внутри отрезка $[\theta_1-\delta,\theta_1+\delta]$ есть локальный максимум, то есть $\left|\widehat{\theta}_n-\theta_1\right|<\delta$ (строгое неравенство, потому что мы доказывали строгие неравенства для L)

■

Замечание 2. Теперь обсудим, откуда возникает информация Фишера. У нас есть функция истинного правдоподобия $W(\theta)$. Разложим ее по Тейлору в точке θ_1 (предполагаем, что зависимость он θ позволяет дифференцировать под знаком интеграла). $W'(\theta)=\int\frac{\rho'_\theta(x)}{\rho_\theta(x)}\rho_{\theta_1}(x)\,dx$, соответственно $W'(\theta_1)=0$. $W''(\theta)=\int\frac{\rho''_\theta(x)}{\rho_\theta(x)}\rho_{\theta_1}(x)\,dx-\int\left(\frac{\rho'_\theta(x)}{\rho_\theta(x)}\right)^2\rho_{\theta_1}(x)\,dx$, соответственно $W''(\theta_1)=0-\int\left(\frac{\partial\ln\rho_\theta(x)}{\partial\theta}\right)^2\rho_{\theta_1}(x)\,dx=-I(\theta_1)$. В итоге получаем:

$$W(\theta)\approx W(\theta_1)-\frac{1}{2}I(\theta_1)(\theta-\theta_1)^2$$

То есть она выражает скорость изменения информации в окрестности максимума θ_1.

Замечание 3. Теперь вспомним неравенство Рао-Крамера. Равенство в неравенстве Коши-Буняковского выполняется только в случае пропорциональности $\theta_n(X)-\tau(\theta)$ и $\frac{\partial L}{\partial\theta}$. То есть, когда

$$\theta_n(X)-\tau(\theta)=C(\theta)\frac{\partial L}{\partial\theta}$$

Это знание в некоторых случаях помогает находить эффективные оценки

Замечание 4. Также отметим, что если существует несмещенная оценка $\theta_n(X)$ параметра θ, на которой достигается равенство в Рао-Крамере, то это обязательно оценка максимального правдоподобия. Действительно, верно равенство

$$\theta_n(X)-\tau(\theta)=C(\theta)\frac{\partial L}{\partial\theta}$$

Подставив вместо θ оценку максимального правдоподобия, получим

$$\theta_n(X)-\widehat{\theta}_n(X)=0$$

То есть они равны.

Замечание 5. И, наконец, сделаем небольшое замечание о том, откуда берется асмптотическа нормальность оценки максимального правдоподобия. Разложим $\frac{\partial L}{\partial\theta}(X,\theta)$ по Тейлору в θ_1 (примерно, без о малых):

$$\frac{\partial L}{\partial\theta}(X,\theta)\approx\frac{\partial L}{\partial\theta}(X,\theta_1)+\frac{\partial^2L}{\partial\theta^2}(X,\theta)(\theta-\theta_1)$$

Если мы подставим $\theta=\widehat{\theta}_n(X)$ – оценку максимального правдоподобия, то $\frac{\partial L}{\partial\theta}\left(X,\widehat{\theta}_n(X)\right)=0$ просто из определения оценки максимального правдоподобия. Тогда получаем:

$$\widehat{\theta}_n-\theta_1=-\frac{\frac{\partial L}{\partial\theta}}{\frac{\partial^2L}{\partial\theta^2}}$$

Домножим на $\sqrt{n}$, чтобы стало как в ЦПТ. Берем $\frac{\partial L}{\partial \theta} = \frac{1}{n}\sum\limits_{j=1}^{n} \frac{\partial}{\partial \theta} \ln \rho_{\theta_1}(x_j)$. Помним, что $\mathbb{E}\frac{\partial}{\partial \theta} \ln \rho_{\theta_1}(x_j) = 0 = \mu$. Теперь дисперсия:

$$\sigma^2 = \mathbb{D}\frac{\partial}{\partial \theta} \ln \rho_{\theta_1}(x_j) = \mathbb{E}\left[\frac{\partial}{\partial \theta} \ln \rho_{\theta_1}(x_j)\right]^2 = I(\theta_1)$$

Тогда из ЦПТ получаем

$$\frac{\partial L}{\partial \theta} \xrightarrow{P} N(0, I(\theta_1))$$

Теперь разберемся с

$$\frac{\partial^2 L}{\partial \theta^2} = \frac{1}{n}\sum_{j=1}^{n} \frac{\partial^2}{\partial \theta^2} \ln \rho_{\theta_1}(x_j) \xrightarrow{P} -I(\theta_1)$$

То есть $\mathbb{E}\frac{\partial^2 L}{\partial \theta^2} = -I(\theta_1)$. В итоге получаем, что $\sqrt{n}\left(\widehat{\theta}_n - \theta_1\right) \xrightarrow{P} N\left(0, \frac{1}{I(\theta_1)}\right)$