СЕМ, Л-11
(2020-12-10)

Дефиниция. Две случайни величини са равни тогава и само тогава, когато имат еднакви функции на разпределение $X = Y \Leftrightarrow F_X(t) = F_Y(t)$, за всички $t \in \mathbb{R}$. Ако X, Y са непрекъснати случайни величини, то от дясната страна на $\Leftrightarrow$ може да сложим и равенство на плътностите $f_X = f_Y$.

Твърдение. Нека $Z_1, Z_2, \ldots, Z_n$ са независими стандартно нормално разпределени случайни величини. $Z_i \in \mathcal{N}(0,1),\ \forall 1 \le i \le n$. Тогава :

$$X = Z_1^2 + Z_2^2 + \ldots + Z_n^2 = \sum_{i=1}^{n} Z_i^2 \in \mathcal{X}^2(n).$$

Доказателство: Ще докажем, че $Z_1^2 \in \mathcal{X}^2(1) = \Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$. Понеже Z_i^2 са независими и еднакво разпределени, то от Л-10

$$\Rightarrow \sum_{i=1}^{n} Z_i^2 \in \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$$, което по дефиниция е $\mathcal{X}^2(n)$.

Тоест трябва да докажем само, че $Z_1^2 = \mathcal{X}^2(1)$. Този факт не може да се докаже директно със смяна, защото функцията, която имаме $Y_1 = Z_1^2 = g(Z_1),\ g(x) = x^2$, която функция g не е монотонна и еднозначна. Ще трябва да използваме директен подход.

Ще се опитаме да пресметнем плътността на Y_1 и от нейната производна да намерим функцията на разпределението.

$x \ge 0,\ \mathbb{P}(Y_1 < x) = \mathbb{P}(Z_1^2 < x) = \mathbb{P}(-\sqrt{x} < Z_1 < \sqrt{x}) =$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{x}}^{\sqrt{x}} e^{-\frac{y^2}{2}} dy = \frac{2}{\sqrt{2\pi}} \int_{0}^{\sqrt{x}} e^{-\frac{y^2}{2}} dy\,.$$

$$f_{Y_1}(x) = \frac{\partial}{\partial x} \frac{2}{\sqrt{2\pi}} \int_{0}^{\sqrt{x}} e^{-\frac{y^2}{2}} dy = \frac{1}{\sqrt{2\pi}} \times \frac{1}{2\sqrt{x}} \times e^{-\frac{x}{2}} = \frac{x^{-\frac{1}{2}}}{\sqrt{2\pi}} \times e^{-\frac{x}{2}} = \frac{x^{-\frac{1}{2}} e^{-\frac{x}{2}}}{\Gamma\left(\frac{1}{2}\right)},$$

което е плътността на $\mathcal{X}^2(1)$.

(E) t-разпределение (Student's t-distribution)

t-разпределението (разпределение на Стюдънт) е вероятностно разпределение, което възниква при оценката на средната стойност на нормално разпределена популация, когато размерът на извадката е малък и стандартното отклонение на популацията е неизвестно.

„Разпределението на Стюдънт“ е кръстено на британския статистик и химик Уилям Сили Госет (1876–1937) 8 . Госет публикувал работите си под псевдонима „Стюдънт“ (Student), тъй като работодателят му, пивоварната Guinness (Дъблин), имала политика служителите да не разкриват имената си в публикации.

Дефиниция. Нека $X_1, X_2, \ldots, X_n$ са независими случайни величини от нормално разпределение $\mathcal{N}(\mu, \sigma^2)$. Тогава статистиката:

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

Следва t-разпределение с $\nu = n - 1$ степени на свобода, където:

- $\overline{X} = \frac{1}{n}\Sigma_{i=1}^{n} X_i$ е средна стойност на извадката
- $S^2 = \frac{1}{n-1}\Sigma_{i=1}^{n}(X_i - \overline{X})^2$ е извадковата дисперсия
- n е размерът на извадката

Плътност на вероятността:

Функцията на плътност за t-разпределение с ν степени на свобода е:

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}, -\infty < t < \infty$$

където Γ е гама функцията.

Функция на рапределение:

Функцията на разпределение се изразява чрез непълна бета функция:

$$F(t) = \mathbb{P}(T \leq t) = \frac{1}{2} + \frac{t \cdot \Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \cdot {}_2F_1\left(\frac{1}{2}, \frac{\nu+1}{2}; \frac{3}{2}; -\frac{t^2}{\nu}\right)$$

където ${}_2F_1$ е хипергеометрична функция

Доказателство за произхода на t-разпределението:

Нека $Z \sim \mathcal{N}(0, 1)$ и $W \sim \chi^2_\nu$ са независими случайни величини. Дефинираме:

$$T = \frac{Z}{\sqrt{W/\nu}}$$

Тогава функцията на плътност на T е:

$$f_T(t) = \int_0^\infty f_Z(t\sqrt{w/\nu}) \cdot f_V(w) \cdot \sqrt{w/\nu}\, \mathrm{d}\, w,$$

където $f_Z(z) = \dfrac{1}{\sqrt{2\pi}} e^{-z^2/2}$ и $f_W(w) = \dfrac{1}{2^{\nu/2}\Gamma(\nu/2)} w^{\nu/2-1} e^{-w/2}$.

След интегрирането се получава:

$$f_T(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

<u>Граници</u>:

- При $\nu \to \infty$:

  $$\lim_{\nu\to\infty} f(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$$

  което е плътността на стандартното нормално разпределение.

- При $\nu \to 1$:

  $$f(t) = \frac{1}{\pi(1 + t^2)}$$

  което е плътността на разпределението на Коши.

<u>Приложение</u>:

t-разпределението се използва предимно в:

- t-тестове за проверка на хипотези за средни стойности
- Доверителни интервали за средна стойност при неизвестна дисперсия
- Регресионен анализ и анализ на коефициентите в линейни модели

Видове сходимост на случайни величини

$X_n : \Omega \to \mathbb{R}, n \geq 1$ и $X : \Omega \to \mathbb{R}$ са случайни величини във вероятностното пространство $V = (\Omega, \mathscr{A}, \mathbb{P})$ (тоест имаме едно единствено вероятностно пространство и $X_i, i = \overline{1, n}, X$ са функции на елементарни събития в реалните числа).

1. Сходимост по вероятност (Почти сигурна сходимост).

- **Означение:** $X_n \xrightarrow{\text{п.с.}} X$
- **Интуиция:** С вероятност 1, реалните реализации (пътеките) на случайните величини X_n се доближават до тези на X и остават близки завинаги.

  Ако проведем „експеримента“ (генератора на случайности ω) веднъж, получаваме една реална последователност от числа $X_1(\omega), X_2(\omega), \ldots$. При почти сигурна сходимост, за почти всеки такъв експеримент, тази конкретна числова редица ще се доближава и ще остане близо до границата $X(\omega)$.
- **Формално:** $\mathbb{P}\left(\omega \in \Omega : \lim\limits_{n\to\infty} X_n(\omega) = X(\omega)\right) = 1,$

  Това означава, че за **почти всички** елементарни събития ω (тоест за множество с вероятност 1), редицата от числа $\{X_n(\omega)\}_{n=1}^{\infty}$ клони като обикновена числова редица към числото $X(\omega)$.
- **Аналогия:** Функционална сходимост. Вземете конкретния резултат от случайния експеримент (ω) - за почти всички такива резултати, числовата редица $\{X_n(\omega)\}_{n=1}^{\infty}$ клони към $X(\omega)$.
- **Пример:** Законът за големите числа на Колмогоров.
- **Втори пример:** Нека имаме вероятностното пространство $(\Omega, \mathscr{A}, \mathbb{P})$, където $\Omega = [0,1]$ е единичния интервал. $\mathscr{A}$ е бореловата σ-алгебра (можем да измерваме интервали), а $\mathbb{P}$ е **мярката на Льобег** (Анри Льобег) (дължина на интервал) $\mathbb{P}([a, b]) = b - a$. Това съответства на равномерно разпределение в $[0,1]$.

  Дефинираме редица от случайни величини по следния начин: за всяко $n \in \mathbb{N}, (n = 1, 2, 3, \ldots)$: $X_n(\omega) = \dfrac{\omega}{n}$, $\omega \in [0,1]$ и целевата (граничната) слузайна величина $X(\omega) = 0$ за всички $\omega \in [0,1]$. Тоест X е константна случайна величина, равна на 0.

  Представете си: ω е фиксирана точка в $[0,1]$ (например 0.49):
  - $X_1(\omega) = 0.49$
  - $X_2(\omega) = 0.245$
  - $X_3(\omega) \approx 0.163$
  - $X_{10}(\omega) = 0.049$
  - $X_{100}(\omega) = 0.0049$

Започваме от права линия $X_1(\omega) = \omega$ и постепенно я „сплескваме“ към нулата.

Фиксираме конкретна точка ω_0. За нея, редицата $\{X_n(\omega_0)\}$ е просто някаква числова редица: $X_n(\omega_0) = \frac{\omega_0}{n}$, за която $\lim_{n\to\infty} \frac{\omega_0}{n} = 0$ за всяко фиксирано ω_0. Това е вярно за всяко реално число ω_0, така че определено е вярно и за $\omega_0 \in [0,1]$. Следователно множеството на сходимост е $A = \left\{\omega \in [0,1] : \lim_{n\to\infty} \frac{\omega}{n} = 0\right\} = [0,1]$ и вероятността за това множество е $\mathbb{P}(A) = \mathbb{P}([0,1]) = 1 - 0 = 1$.

Следователно $\mathbb{P}\left(\{\omega : \lim_{n\to\infty} X_n(\omega) = 0\}\right) = 1$.

2. Сходимост по вероятност.

- **Означение:** $X_n \xrightarrow{\mathbb{P}} X$
- **Интуиция:** Не ни интересува дали за *един* експеримент редицата се сближава. Интересува ни, че вероятността X_n да се отклони от X с повече от произволно малко ϵ става пренебрежимо малка с нарастване на n. Това е **сходимост на вероятностите** не на самите реализации.
- **Формално:** $\forall \epsilon > 0 : \lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$
- **Аналогия:** Мерна сходимост. Не ни интересува дали *всички* реализации се доближават, а само че „лошите“ реализации (тези, които се отклоняват с $> \varepsilon$) стават все по-редки.
- **Връзка:** Почти сигурна (п.с.) сходимост $\Rightarrow$ сходимост **по вероятност** (обратното не е вярно)
- **Пример:** Слабият закон за големите числа (разгледан и доказан по долу в следващата точка), оценители в статистиката (например сходимост на извадкова средна).

3. Сходимост по разпределение.

- **Означение:** $X_n \xrightarrow{D} X$ или $X_n \rightsquigarrow X$
- **Интуиция:** Функцията на разпределение $F_n(x)$ на X_n клони към функцията на разпределение на $F(x)$ на X във всички точки на непрекъснатост на F.
- **Формално:** $\lim_{n\to\infty} F_n(x) = F(x)$ за всяко x, в което F е непрекъсната.
- **Важно:** Това е **най-слабият** вид сходимост. Тук случайните величини могат да бъдат дефинирани на напълно различни вероятностни пространства. Сходимостта се отнася само за закона на разпределение, а не за самите стойности.
- **Връзка:** Сходимост **по вероятност** $\Rightarrow$ сходимост **по разпределение** (обратното не е вярно, освен ако граничната случайна величина не е константа).

- **Пример: Централна гранична теорема** (CLT: Central Limit Theorem) - суми от независими случайни величини, стаандартизирани подходящо, се сближават по разпределение към стандартно нормално разпределение.

Дефиниция. Ако F_X е функция на разпределение, то с C_{F_X} означаваме всички точки x, за които F е непрекъсната в x. $C_{F_X} = \{x \in \mathbb{R} : \ F_X$ е непрекъсната в $x\}$ и $x \in C_{F_X} \Leftrightarrow \mathbb{P}(X = x) = 0$.

⊕ Ако X е непрекъсната случайна величина, то $\mathbb{P}(X = x) = 0, \forall x \in \mathbb{R} \Rightarrow C_{F_X} = \mathbb{R}$.

Алтернативна дефиниция (Сходимост по разпределение). Казваме, че $X_n \xrightarrow{d} X \Leftrightarrow \lim\limits_{n\to\infty} F_{X_n}(x) = F_X(x), \forall x \in C_{F_X}$ (тук никъде не споменаваме вероятностно пространство).

Теорема. Нека $\{X_n\}_{n=1}^{\infty}$ е редица от случайни величини и X е случайна величина.

А) Ако $\{X_n\}$ почти сигурно схожда по вероятност към X, то $\{X_n\}$ схожда по вероятност към X. Тоест: $X_n \xrightarrow[n\to\infty]{\text{п.с.}} X \Rightarrow X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$

Б) Ако $\{X_n\}$ схожда по вероятност към X, то $\{X_n\}$ схожда по разпределение към X. Тоест: $X_n \xrightarrow[n\to\infty]{\mathbb{P}} \Rightarrow X_n \xrightarrow[n\to\infty]{d} X$

В) Обратните индикации на А) и Б) не са верни.

Преди да преминем към доказателствата на връзките между видовете сходимост на случайни величини от теоремата както и примерите от видовете сходимост, ще разгледаме някой неравенства, които ще използваме.

Неравенство на Марков (Андрей Андреевич Марков (1856-1922), Русия)

Нека X е неотрицателна случайна величина (тоест $\mathbb{P}(X \geq 0) = 1$) с крайно математическо очакване $\mathbb{E}X < \infty$. Тогава за всяко $a > 0$ е е изпълнено:

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}X}{a}$$

Доказателство:

I. *Дискретна* случайна величина.

$\mathbb{E}[X] = \sum_i x_i p_i$, където сумирането е по всички въчможни стойности $x_i \geq 0$.

Разделяме сумата на две части: сума за $x_i < a$ и сума за $x_i \geq a$:

$$\mathbb{E}[X] = \sum_{x_i<a} x_i p_i + \sum_{x_i \geq a} x_i p_i$$

Първата сума е неотрицателна (тъй като $x_i \geq 0$), следователно:

$$\mathbb{E}[X] \geq \sum_{x_i \geq a} x_i p_i$$

За втората сума използваме, че сме в случая, в който $x_i \geq a$ и следователно:

$$\sum_{x_i \geq a} x_i p_i \geq \sum_{x_i \geq a} a p_i = a \sum_{x_i \geq a} p_i$$

Но $\sum_{x_i \geq a} p_i = \mathbb{P}(X \geq a)$. Комбинираме и получаваме $\mathbb{E}[X] \geq a \cdot \mathbb{P}(X \geq a)$.

Разделяме на $a > 0$ и получаваме исканото неравенство $\mathbb{P}(X \geq a) \leq \dfrac{\mathbb{E}[X]}{a}$.

$\square$

II. *Непрекъсната* случайна величина.

Нека $f(x)$ е плътността на X (за $x \geq 0$). Математическото очакване е $\mathbb{E}[X] = \int_0^\infty x f(x) dx$. Разделяме интеграла на две части:

$$\mathbb{E}[X] = \int_0^a x f(x) dx + \int_a^\infty x f(x) dx$$

Първият интеграл е неотрицателен (тъй като $x \geq 0$), следователно:

$$\mathbb{E}[X] \geq \int_a^\infty x f(x) dx$$

За втория интеграл изполчваме, че сме в случая, в който $x \geq a$ и следователно:

$$\int_a^\infty x f(x) dx \geq \int_a^\infty a f(x) dx = a \int_a^\infty f(x) dx$$

Но $\int_a^\infty f(x) dx = \mathbb{P}(X \geq a)$. Комбинираме и получаваме $\mathbb{E}[X] \geq a \cdot \mathbb{P}(X \geq a)$.

Разделяме на $a > 0$ и получаваме исканото неравенство $\mathbb{P}(X \geq a) \leq \dfrac{\mathbb{E}[X]}{a}$.

$\square$

Неравенство на Чебишов (Пафнутий Львович Чебишов (1821–1894), Русия)

Нека X е случайна величина с крайно математическо очакване $\mu = \mathbb{E}[X] < \infty$ и с крайна дисперсия $\sigma^2 = \mathrm{Var}[X] < \infty$. Тогава за всяко $k > 0$:

$$\mathbb{P}\left(|X - \mu| \geq k\sigma\right) \leq \frac{1}{k^2}$$

Или еквивалентно за всяко $a > 0$:

$$\mathbb{P}\left(|X - \mu| \geq a\right) \leq \frac{\sigma^2}{a^2}$$

Доказателство:

I. Чрез дисперсия и индикаторна функция.

Дефинираме индикатор:

$$I = \begin{cases} 1, & \text{ако } |X - \mu| \geq a, \\ 0, & \text{иначе} \end{cases}$$

- Ако $I = 1$, тогава $|X - \mu| \geq a$, откъдето $(X - \mu)^2 \geq a^2 = a^2 \cdot 1$
- Ако $I = 0$, тогава $(X - \mu)^2 \geq 0 = a^2 \cdot 0$

Следоивателно $(X - \mu)^2 \geq a^2 \cdot I$

Вземаме математическо очакване:

$$\mathbb{E}\left[(X - \mu)^2\right] \geq \mathbb{E}[a^2 \cdot I] = \varepsilon^2 \mathbb{E}[I] = a^2 \mathbb{P}(|X - \mu| \geq a)$$

Но $\mathbb{E}\left[(X - \mu^2\right] = \sigma^2$, следователно:

$$\sigma^2 \geq a^2 \cdot \mathbb{P}\left(|X - \mu| \geq a\right)$$

Преобразуваме:

$$\mathbb{P}\left(|X - \mu| \geq a\right) \leq \frac{\sigma^2}{a^2}$$

$\square$

Неравенството на Чебишов може лесно да се изведе от неравенството на Марков, чрез трансформация на проблема.

Доказателство:

Дефинираме **нова неотрицателна случайна величина**. Разглеждаме отклонението на X от средната стойност на квадрат:

$$Y = (X - \mu)^2$$

Величината Y е **неотрицателна** ($Y \geq 0$), тъй като е квадрат. Нейното математическо очакване е:

$$\mathbb{E}[Y] = \mathbb{E}\left[(X - \mu)^2\right] = \mathrm{Var}[X] = \sigma^2$$

Забелязваме връзката между събитията. Събитието, което ни интересува в неравенството на Чебишов, е:

$$\{\,|X-\mu| \geq a\}$$

Но това е еквивалентно на:

$$\{(X-\mu)^2 \geq a^2\} = \{Y \geq a^2\}$$

Прилагаме неравенството на Марков за Y и праг $c > 0$:

$$\mathbb{P}(Y \geq c) \leq \frac{\mathbb{E}[Y]}{c}$$

В нашия случай избираме $c = a^2$. Тогава:

$$\mathbb{P}(Y \geq a^2) \leq \frac{\mathbb{E}[Y]}{a^2} = \frac{\sigma^2}{a^2}$$

Връщаме се към първоначалното събитие. Тъй като $\mathbb{P}(Y \geq a^2) = \mathbb{P}\big((X-\mu)^2 \geq a^2\big) = \mathbb{P}\big(\,|X-\mu| \geq a\big)$, получаваме $\mathbb{P}\big(\,|X-\mu| \geq a\big) \leq \dfrac{\sigma^2}{a^2}$, което искахме да докажем.

$\square$

Доказване на връзките между видовете сходимост.

А) Сходимост почти сигутно ⇒ Сходимост по вероятност

Дадено: $X_n \xrightarrow{\text{п.с.}} X$, тоест $\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1$.
Трябва да докажем: За всяко $\varepsilon > 0$, $\lim_{n\to\infty} \mathbb{P}\big(\,|X_n - X| \geq \varepsilon\big) = 0$.

Доказателство:

Дефинираме събитието на сходимост:

$$A = \left\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}$$

По условие $\mathbb{P}(A) = 1$. За фиксирано $\varepsilon > 0$, дефинираме събитята:

$$B_m = \bigcap_{n=m}^{\infty} \{|X_n - X| < \varepsilon\}$$

Това е събитието „за всички $n \geq m$, $|X_n - X| < \varepsilon$“.

Забелязваме, че ако $\omega \in A$, тогава съществува m (зависещо от ω и ε) такова, че $\omega \in B_m$. Следователно:

$$A \subseteq \bigcup_{m=1}^{\infty} B_m$$

Освен това, последователността $\{B_m\}$ е **не-намаляваща** (тъй като $B_1 \subseteq B_2 \subseteq \dots$), следователно:

$$\mathbb{P}\left(\bigcup_{m=1}^{\infty} B_m\right) = \lim_{m\to\infty} \mathbb{P}(B_m)$$

От $A \subseteq \bigcup_{m=1}^{\infty} B_m$ и $\mathbb{P}(A) = 1$, имаме:

$$1 = \mathbb{P}(A) \leq \mathbb{P}\left(\bigcup_{m=1}^{\infty} B_m\right) = \lim_{m\to\infty} \mathbb{P}(B_m) \leq 1$$

Следователно:

$$\lim_{n\to\infty} \mathbb{P}(B_m) = 1$$

Сега, $B_m \subseteq \left\{|X_m - X| < \varepsilon\right\}$ (всъщност B_m е **по-силно** събитие от $\left\{|X_m - X| < \varepsilon\right\}$).
Знаем, че $\left\{|X_m - X| \geq \varepsilon\right\} = \left\{|X_m - X| < \varepsilon\right\}^c \subseteq B_m^c$.

Тогава:

$$\mathbb{P}\big(|X_m - X| \geq \varepsilon\big) \leq \lim_{m\to\infty} [1 - \mathbb{P}(B_m)] = 1 - 1 = 0$$

Тъй като вероятността е неотрицателна, получаваме:

$$\lim_{n\to\infty} \mathbb{P}\big(|X_n - X| \geq \varepsilon\big) = 0$$

Това точно е определението за $X_n \xrightarrow{\mathbb{P}} X$, което искахме да докажем.

$\square$

Б) Сходимост по вероятност ⇒ Сходимост по разпределение

Дадено: $X_n \xrightarrow{\mathbb{P}} X$.
Трябва да докажем: За всяка точка на непрекъснатост на x на функцията на разпределение $F_X(t) = \mathbb{P}(X \leq t)$, имаме

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

Където $F_{X_n}(t) = \mathbb{P}(X_n \leq t)$ е функцията на разпределението.

Доказателство:

Нека x е точка на непрекъснатост на F_X. Фиксираме $\varepsilon > 0$. Разглеждаме събитията и техните вероятности:
За всяко n:

$$F_{X_n}(x) = \mathbb{P}(X_n \leq x)$$

Искаме да сравним $\mathbb{P}(X_n \leq x)$ с $\mathbb{P}(X \leq x)$.
Забелязваме, че:

$$\{X_n \leq x\} = \underbrace{\{X_n \leq x, X \leq x + \varepsilon\}}_{\text{част 1}} \cup \underbrace{\{X_n \leq x, X > x + \varepsilon\}}_{\text{част 2}}$$

Част 2 имплицира, че $|X_n - X| > \varepsilon$, защото ако $X_n \leq x$ и $X > X + \varepsilon$, тогава $X - X_n > \varepsilon$.
Следователно:

$$\mathbb{P}(X_n \leq x) \leq \mathbb{P}(X \leq x + \varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon)$$

Тъй като $\{X_n \leq x, X \leq x + \varepsilon\} \subseteq \{X \leq x + \varepsilon\}$ и $\mathbb{P}(\text{част 2}) \leq \mathbb{P}(|X_n - X| > \varepsilon)$.

Вземаме горна граница при $n \to \infty$:

$$\limsup_{n\to\infty} F_{X_n}(x) \leq \limsup_{n\to\infty} \left[\mathbb{P}(X \leq x + \varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon)\right]$$

От $X_n \xrightarrow{\mathbb{P}} X$, $\mathbb{P}(|X_n - X| > \varepsilon) \to 0$, така че:

$$\limsup_{n\to\infty} F_{X_n}(x) \leq \mathbb{P}(X \leq x + \varepsilon) = F_X(x + \varepsilon)$$

Сега, от друга страна, нека разгледаме събитието $\{X \leq x - \varepsilon\}$:

$$\{X \leq x - \varepsilon\} = \underbrace{\{X \leq x - \varepsilon, X_n \leq x\}}_{\text{част А}} \cup \underbrace{\{X \leq x - \varepsilon, X_n > x\}}_{\text{част Б}}$$

Част Б имплицира, че $|X_n - X| > \varepsilon$ (тъй като $X_n > x$ и $X \leq x - \varepsilon$ дава $X_n - X > \varepsilon$).

Следователно:

$$F_X(x - \varepsilon) = \mathbb{P}(X \leq x - \varepsilon) \leq \mathbb{P}(X_n \leq x) + \mathbb{P}(|X_n - X| > \varepsilon)$$

Пренареждаме:

$$F_X(x - \varepsilon) - \mathbb{P}(|X_n - X| > \varepsilon) \leq F_{X_n}(x)$$

Вземаме долна граница при $n \to \infty$:

$$\liminf_{n\to\infty} F_{X_n}(x) \geq F_X(x - \varepsilon) - \lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = F_X(x - \varepsilon)$$

Така полуихме за всяко $\varepsilon > 0$:

$$F_X(x - \varepsilon) \leq \liminf_{n\to\infty} F_{X_n}(x) \leq \liminf_{n\to\infty} F_{X_n} \leq F_X(x + \varepsilon)$$

Сега, тъй като x е точка на непрекъснатост на F_X, когато $\varepsilon \to 0^+$, имаме:

$$F_X(x - \varepsilon) \to F_X(x) \text{ и } F_X(x + \varepsilon) \to F_X(x)$$

Следователно:

$$\liminf_{n\to\infty} F_{X_n}(x) = \limsup_{n\to\infty} F_{X_n}(x) = F_X(x)$$

Което означава:

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

Това е точно определението за $X_n \xrightarrow{d} X$.

$\square$

В) Обратните индикации на А) и Б) не са верни. Тоест:

$$X_n \xrightarrow[n\to\infty]{\mathbb{P}} X \not\Rightarrow X_n \xrightarrow[n\to\infty]{\text{п.с.}} \qquad \text{и} \qquad X_n \xrightarrow[n\to\infty]{d} X \not\Rightarrow X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$$

Най-лесният начин да се покажат неверностите на обратните индикации са чрез конрапримери.

$$X_n \xrightarrow[n\to\infty]{\mathbb{P}} X \not\Rightarrow X_n \xrightarrow[n\to\infty]{\text{п.с.}} X$$

Доказателство:

Нека вероятностното пространство е $\Omega = [0,1]$ с равномерно разпределение (дължина на интервалите = вероятност). Дефинираме редица случайни величини $X_1, X_2, X_3, \ldots$ по следния начин:

Стъпка 1: Подреждаме интервалите в редица

Първо: $X_1 = \mathbf{1}_{[0,1]}$ (целия интервал)
После: $X_2 = \mathbf{1}_{[0,\frac{1}{2}]}, X_3 = \mathbf{1}_{[\frac{1}{2},1]}$
После: $X_4 = \mathbf{1}_{[0,\frac{1}{3}]}, X_5 = \mathbf{1}_{[\frac{1}{3},\frac{2}{3}]}, X_6 = \mathbf{1}_{[\frac{2}{3},1]}$
После: $X_7 = \mathbf{1}_{[0,\frac{1}{4}]}, X_8 = \mathbf{1}_{[\frac{1}{4},\frac{2}{4}]}, X_9 = \mathbf{1}_{[\frac{2}{4},\frac{3}{4}]}, X_{10} = \mathbf{1}_{[\frac{3}{4},1]}$
и т.н.

Общо: На k-тия „блок“ (след първия) интервалите имат дължина $\frac{1}{k}$, покриват целия $[0,\ 1]$ без пропуски и припокривания (освен границите).

Стъпка 2: Проверка за сходимост по вероятност към 0

Фиксираме $\varepsilon = 0.5$.
Трябва да проверим $\mathbb{P}(|X_n - 0| > \varepsilon) = \mathbb{P}(X_n = 1)$.

Но $\mathbb{P}(X_n = 1)$ = дължината на интервала, на който X_n е индикатор. С нарастване на n, интервалите стават все по-къси (дължина $\frac{1}{2}$, после $\frac{1}{3}$, после $\frac{1}{4}, \ldots$). Въпреки че понякога се връщаме към по-дълги интервали (защото подредбата не е строго намаляваща), като цяло за всяко n, има безкрайно много членове на редицата с произволно малка дължина, и затова:

За всяко фиксирано $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|X_n| > \varepsilon) = 0.$$

(Защо? Защото без значение колко е голямо n, ще имаме интервали с дължина $\frac{1}{m}$ за всяко m, и когато m расте, вероятността $X_n = 1$ клони към нула.)

Точно доказателство за сходимост по вероятност:

В нашата подредба, за $n \geq 2$, ако X_n е индикатор на интервал с дължина ℓ_n, то ℓ_n приема стойности $\frac{1}{2}, \frac{1}{3}, \frac{1}{3}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \ldots$
Вижда се, че $\ell_n \to 0$ (защото за всяко $\varepsilon > 0$ съществува N, такова че за всички $n > N$, $\ell_N < \varepsilon$).

Следователно $\mathbb{P}(|X_n| > \varepsilon) = \ell_n \to 0$. Значи $X_n \xrightarrow{\mathbb{P}} 0$.

Стъпка 3: Проверка за липса на почти сигурна сходимост

Взимаме което и да е $\omega \in [0, 1]$. В нашата подредба интервалите покриват целия $[0, 1]$ многократно, и дължините им не стават нулеви за всички след някой индекс, а периодично се появяват интервали, съдържащи ω.

По-точно:

- На първи етап: $X_1(\omega) = 1$ за всочло ω.
- На втори етап: един от X_2, X_3 е 1 за даденото ω.
- На трети етап: един от X_4, X_5, X_6 е 1 за даденото ω.
  И така, за всяко ω има безброй много n, за които $X_n(\omega) = 1$, и също безкрайно много n, за които $X_n(\omega) = 0$ (защото интервалите са различни и не покриват само една точка).

Значи редицата $\{X_n(\omega)\}$ съдържа безкрайно много 1 и безкрайно много 0. Следователно тя не клони към 0 за нито едно ω.

Следователно:

$$\mathbb{P}\left(\lim_{n\to\infty} X_n = 0 \right) = 0,$$

Тоест няма сходимост почти сигурно към нула.

$\square$

$$X_n \xrightarrow[n\to\infty]{d} X \not\Rightarrow X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$$

Доказателство:

Нека $\Omega = \{0,1\}$ с $\mathbb{P}(0) = \mathbb{P}(1) = 1/2$. Нека $X(\omega) = \omega$ (тоест $X(0) = 0$ и $X(1) = 1$). Нека $X_n(\omega) = 1 - \omega$ (тоест $X_n(0) = 1$, $X_n(1) = 0$).
Тогава X_n има същото разпределение като X:
$\mathbb{P}(X_n = 0) = 1/2$, $\mathbb{P}(X_n = 1) = 1/2$.
Следователно $X_n \xrightarrow{d} X$.

Но $X_n - X = (1 - \omega) - \omega = 1 - 2\omega$, което е 1 при $\omega = 0$ и -1 при $\omega = 1$, значи $|X_n - X| = 1$ винаги.

Тогава $\mathbb{P}(|X_n - X| \geq 0.5) = 1$ за всяко n, значи няма сходимост по вероятност.

$\square$

Твърдение. Ако $X_n \xrightarrow{d} c$, то и $X_n \xrightarrow{\mathbb{P}} c$, където c е константа.

Доказателство:

Вземаме произволно $\varepsilon > 0$. Искаме да покажем, че $\lim_{n\to\infty} \mathbb{P}(|X_n - c| \geq \varepsilon) = 0$.

Забелязваме, че $\mathbb{P}(|X_n - c| \geq \varepsilon) = \mathbb{P}(X_n \leq c - \varepsilon) + \mathbb{P}(X_n \geq c + \varepsilon)$.

От сходимост по разпределение знаем, че функцията на разпределение $F_{X_n}(x)$ клони към $F_c(x)$, където:

$$F_c(x) = \begin{cases} 0 & \text{ако } x < c \\ 1 & \text{ако } x \geq c \end{cases}$$

и сходимостта е във всяка точка на непрекъснатост на F_c.

Избираме точките $c - \varepsilon$ и $c + \varepsilon$ на непрекъснатост на F_c (защото $\varepsilon > 0$, значи $c \pm \varepsilon \neq c$).

Следователно:

$$\lim_{n\to\infty} F_{X_n}(c - \varepsilon) = F_c(c - \varepsilon) = 0.$$

$$\lim_{n\to\infty} F_{X_n}(c + \varepsilon) = F_c(c + \varepsilon) = 1.$$

Преобразуваме вероятностите:

- $\mathbb{P}(X_n \leq c - \varepsilon) = F_{X_n}(c - \varepsilon) \to 0$.
- $\mathbb{P}(X_n \geq c + \varepsilon) = 1 - \mathbb{P}(X_n < c + \varepsilon)$.

За всяко $\delta > 0$ такова, че $c + \varepsilon - \delta$ е точка на непрекъснатост (например $\delta = \varepsilon/2$):

$$\mathbb{P}(X_n < c + \varepsilon) \geq \mathbb{P}(X_n \leq c + \varepsilon - \delta) = F_{X_n}(c + \varepsilon - \delta).$$

Но $c + \varepsilon - \delta > c$, значи:

$$F_{X_n}(c + \varepsilon - \delta) \to F_c(c + \varepsilon - \delta) = 1.$$

Следователно $\lim_{n \to \infty} \mathbb{P}(X_n < c + \varepsilon) = 1.$

Значи:

$$\mathbb{P}(X_n \geq c + \varepsilon) = 1 - \mathbb{P}(X_n < c + \varepsilon) \to 1 - 1 = 0.$$

Събираме:

$$\mathbb{P}(|X_n - c| \geq \varepsilon) = \underbrace{\mathbb{P}(X_n \leq c - \varepsilon)}_{\to 0} + \underbrace{\mathbb{P}(X_n \geq c + \varepsilon)}_{\to 0} \to 0.$$

Това е за всяко $\varepsilon > 0$, следователно $X_n \xrightarrow{\mathbb{P}} c$.

$\square$

Закон за големите числа (ЗГЧ)

1. **Слаб ЗГЧ**

   Ако проведем едни и същи независими вероятностни експерименти (хвърляне на зар, монета и т.н.) много, много пъти, **средната стойност** на резултатите ще се **доближи** до очакваната стойност (математическото очакване).

   - По-формално: Нека средното аритметично на независими, еднакво разпределени случайни величини $X_1, X_2, X_3, \ldots, X_n$ с крайно очакване μ е е $\overline{X_n}$. То **<u>схожда по вероятност</u>** към μ при $n \to \infty$. Тоест: $\lim_{n \to \infty} \mathbb{P}(|\overline{X_n} - \mu| > \varepsilon) = 0$
   - Интуитивно: При голям брой опити, вероятността средната стойност да се отклони значително от очакването става пренебрежимо малка.

2. **Силен ЗГЧ**

   При същите условия, средната стойност $\overline{X_n}$ **<u>схожда почти сигурно</u>** към μ.

   - Това е по-силно твърдение. Означава, че вероятността това средно аритметично да се *сближи* до μ (а не само да е близо при конкретно n) е равна на 1. Разликата е техническа, но важна от математическа гледна точка.

Проблемът с класическия ЗГЧ е, че той изисква случайните величини да имат **еднакво разпределение** и/или да са независими. Какво се случва, ако те са различни, зависими или имат безкрайна дисперсия?

През 1930-1933 г. Колмогоров формулира и доказва най-общата и мощна версия на **Силния закон за големите числа**.

Теорема на Колмогоров за силния ЗГЧ (Андрей Николаевич Колмогоров (1903–1987), Русия)

Нека $X_1, X_2, X_3, \ldots, X_n$ са **независими** случайни величини (не е необходимо да са еднакво разпределени). Тогава, ако е изпълнено следното условие за техните дисперсии: $\sum_{i=1}^{n} \frac{\mathrm{Var}[X_i]}{i^2} < \infty$ (сумата от техните дисперсии, разделена на n^2, е крайна), то случайната величина $S_n - \mathbb{E}[S_n]$ клони към 0 **почти сигурно**, където $S_n = X_1 + X_2 + \ldots + X_n$.

С други думи:

- Отпада изискването за еднакво разпределение. Величините могат да са напълно различни.
- Условието $\sum_{i=1}^{n} \frac{\mathrm{Var}[X_i]}{i^2} < \infty$ е достатъчно, за да гарантира, че отклонението на средното аритметично от очакваната му стойност ще изчезне.
- Ако допълнително предположим, че всички величини имат едно и също очакване μ, тогава $\overline{X_n} = \frac{X_1 + X_2 + \ldots + X_n}{n} \to \mu$ почти сигурно.

Защо това е революционно?

- Премахва драстично ограничението за еднакво разпределение.
- Заменя го с минимално и точно условие върху дисперсиите.
- Демонстрира мощта на новата аксиоматична теория на вероятностите.
- Трансформира ЗГЧ от конкретен принцип за повторения в универсален инструмент за анализ на сложни, нестационарни случайни процеси.

Това не е просто „още една теорема“, а стратегическо разширение на самата сфера на действие на теорията на вероятностите. Тя превръща ЗГЧ от удобен инструмент за игри на хазарт в сериозен апарат за математически анализ на реалния свят.

Нека докажем Слабия ЗГЧ (версия на Чебишев):

Нека $X_1, X_2, \ldots, X_n$ са **независими, еднакво разпределени** случайни величини с крайно математическо очакване $\mathbb{E}[X_i] = \mu$ и крайна дисперсия $\mathrm{Var}[X_i] = \sigma^2$. Означаваме **средната стойност на извадката:** $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Тогава за всяко $\varepsilon > 0$, $\lim_{n\to\infty} \mathbb{P}\left(|\overline{X_n} - \mu| \geq \varepsilon\right) = 0$. Това се нарича **сходимост по вероятност:** $\overline{X_n} \xrightarrow{\mathbb{P}} \mu$.

Доказателство (стъпка по стъпка):

1) **Изчисляваме очакването на $\overline{X_n}$:**

$$\mathbb{E}[\overline{X_n}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{1}{n} \cdot n \cdot \mu = \mu.$$

2) **Изчисляваме дисперсията на $\overline{X_n}$:** Поради **независимостта**, дисперсията на думата е сума от дисперсиите:

$$\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{Var}[X_i] = n\sigma^2.$$

Следователно,

$$\mathrm{Var}[\overline{X_n}] = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}.$$

3) **Прилагаме неравенството на Чебишев към $\overline{X_n}$:** Неравенството на Чебишев за произволна случайна величина Y с крайна дисперсия:

$$\mathbb{P}\big(\,|Y - \mathbb{E}[Y]| \geq \varepsilon\big) \leq \frac{\mathrm{Var}[Y]}{\varepsilon^2}.$$

За $Y = \overline{X_n}$ имаме $\mathbb{E}[\overline{X_n}] = \mu$ и $\mathrm{Var}[\overline{X_n}] = \dfrac{\sigma^2}{n}$. Заместваме:

$$\mathbb{P}\big(\,|\overline{X_n} - \mu| \geq \varepsilon\big) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

4) **Граничен преход** при $n \to \infty$: Фиксираме $\varepsilon > 0$. Тъй като σ^2 и ε^2 са константи,

$$\lim_{n\to\infty} \frac{\sigma^2}{n\varepsilon^2} = 0$$

От неравенството на Чебишев следва, че

$$0 \leq \mathbb{P}\big(\,|\overline{X_n} - \mu| \geq \varepsilon\big) \leq \frac{\sigma^2}{n\varepsilon^2} \to 0.$$

Следователно по теоремата за двамата полицаи (squeeze theorem),

$$\lim_{n\to\infty} \mathbb{P}\big(\,|\overline{X_n} - \mu| \geq \varepsilon\big) = 0.$$

Точно това означава, че $\overline{X_n}$ схожда по веоятност към μ. **Слабият ЗГЧ е доказан.**

$\square$