СЕМ, лекция 12

(2020-12-17)

Централна Гранична Теорема (ЦГТ)

ЗГЧ: $\left(X_i\right)_{i=1}^{\infty}$ е редица от независими и еднакво разпределени (i.i.d.) случайни величини с $\mathbb{E}\,|\,X_1\,| < \infty$, $\mathbb{E}X_1 = \mu$ и $\sigma = \sqrt{\mathbb{D}X_1}$.

$$S_n = \sum_{j=1}^{n} X_j \Rightarrow \frac{S_n}{n} \xrightarrow[n\to\infty]{\text{п.с.}(\mathbb{P})} \mu; \qquad \frac{S_n}{n} - \mu \approx \frac{\sigma}{\sqrt{n}} H_n$$

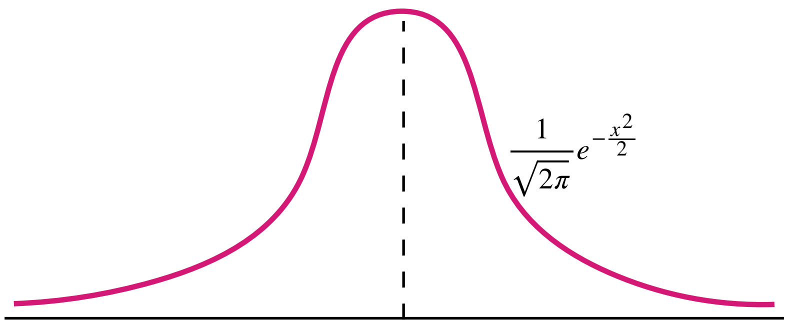
<u>Теорема</u>: (**ЦГТ**). Нека $X = \left(X_i\right)_{i=1}^{\infty}$ е редица от независими, еднакво разпределени случайни величини със $\sigma^2 = \mathbb{D}X_1 < \infty$ и $\mu = \mathbb{E}X_1$. Тогава

$$Z_n \overset{def.}{:=} \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\frac{S_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow[n\to\infty]{d} Z \in \mathcal{N}(0,1).$$

$$f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \mathbb{P}(Z < x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\, \mathrm{d}\, x.$$

$$\mathbb{P}(Z \geq x) = 1 - \Phi(x) = \overline{\Phi}(x).$$

ЦГТ измерва каква е грешката в ЗГЧ. Т.е. ние знаем, че $\frac{S_n}{n} \xrightarrow[n\to\infty]{} \mu$, почти сигурно (траекторно) и скалираме грешката по $\frac{\sigma}{\sqrt{n}}$ и виждаме, че се държи като нормално разпределение.



Често се дава за пример пресмятането на средно IQ за група хора или тяхна височина и т.н., но за да бъде IQ-то нормална разпределена случайна величина е необходимо да бъде средно аритметичен ефект от множество независими характеристики. А дали това е така?

Във физиката ЦГТ може да ни даде каква е грешката между акумулиран ефект и среден ефект, който сме измерили в някаква система.

Следствия.

$\oplus$ $\quad Z_n := \dfrac{S_n - n\mu}{\sigma\sqrt{n}}$, където $\mu = \mathbb{E}X_1$, $\sigma = \sqrt{\mathbb{D}X_1}$. ЦГТ гласи следното:

Ако се интересуваме от

$$\mathbb{P}\left(Z_n \in (a,b)\right) = \mathbb{P}\left(S_n \in (n\mu + a\sigma\sqrt{n}, n\mu + b\sigma\sqrt{n})\right) \sim \mathbb{P}\left(Z \in (a,b)\right) =$$

$$= \Phi(b) - \Phi(a) = \frac{1}{\sqrt{2\pi}} \times \underbrace{\int_a^b e^{-\frac{y^2}{2}}\,\mathrm{d}y}_{\text{този интеграл не се интересува от това с какви случайни величини сме стартирали. Той зависи само от } a \text{ и } b.}.$$

$$\mathbb{P}(Z_n < a) \xrightarrow[n\to\infty]{} \Phi(a) = \mathbb{P}(Z < a); \qquad \mathbb{P}(Z_n \geq a) \xrightarrow[n\to\infty]{} \overline{\Phi}(a) = 1 - \Phi(a)$$

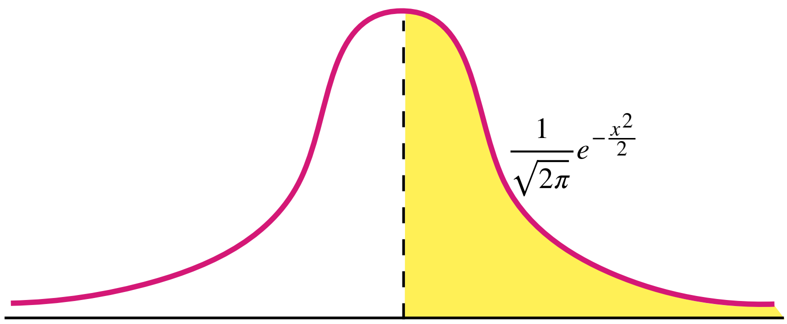
$$\mathbb{P}(Z_n < \mu) \xrightarrow[n\to\infty]{} \Phi(\mu).$$

Т.е. ЦГТ дава инструмент, с който може да приближаваме вероятности.

$\oplus$ $\quad \mu = 0$, $\sigma^2 = 1$, тогава $\dfrac{S_n}{\sqrt{n}} \xrightarrow[n\to\infty]{d} Z \in \mathcal{N}(0,1)$, където $S_n = \sum_{j=1}^{n} X_j$.

$$\mathbb{P}\left(S_n \in (a\sqrt{n}, b\sqrt{n})\right) \sim \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{y^2}{2}}\,\mathrm{d}y.$$

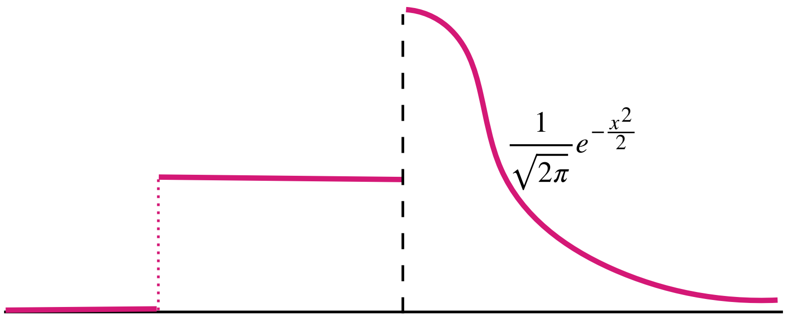
$$\mathbb{P}(S_n > 0) = \mathbb{P}\left(\frac{S_n}{\sqrt{n}} > 0\right) \sim \mathbb{P}(Z > 0) = \frac{1}{\sqrt{2\pi}}\int_0^\infty e^{-\frac{y^2}{2}}\,\mathrm{d}y = \frac{1}{2}.$$



⊕ Хвърляме зарче $6\,000\,000$ пъти. Каква е вероятността измежду тези $6\,000\,000$ пъти да сме хвърлили повече от $1\,000\,000$ пъти 6-ца? Решение: (виж последния пример от зад. 10 от файла SEM HW).

$\mu = 0,\, \sigma^2 = 1,\, S_n = \sum_{j=1}^{n} X_j$

$\mathbb{P}(S_n > 0) \sim \frac{1}{2},\, f_{X_1}(x) = \begin{cases} \frac{1}{2}e^{-x}, & x > 0 \\ \frac{1}{4}, & x \in (-2, 0) \\ 0, & \text{иначе} \end{cases}$



$$\mathbb{E}X_1 = \int_{-2}^{0} x\frac{1}{4}\,\mathrm{d}\,x + \int_{0}^{\infty} x\frac{1}{2}e^{-x}\,\mathrm{d}\,x = \frac{1}{4}\frac{x^2}{2}\bigg|_{-2}^{0} - \frac{1}{2}\int_{0}^{\infty} x\,\mathrm{d}\,e^{-x} =$$

$$= -\frac{1}{4}\times\frac{4}{2} - \underbrace{\frac{1}{2}xe^{-x}\bigg|_{0}^{\infty}}_{=0} + \frac{1}{2}\int_{0}^{\infty} e^{-x}\,\mathrm{d}\,x = -\frac{1}{2} + \frac{1}{2}\left(-e^{-x}\right)\bigg|_{0}^{\infty} =$$

$$= -\frac{1}{2} + \frac{1}{2} = 0.$$

Т.е. дори и за случайни величини, които са много далеч от симетрия, ако ги сумираме всички от тях, то вероятността да видим нещо положително е $\sim \frac{1}{2}$.

$$X_i = \begin{cases} 1, & \frac{1}{2} \\ -1, & \frac{1}{2} \end{cases}, \quad S_n = \mathscr{D}_n - \Lambda_n = \sum_{j=1}^{n} X_j; \ \mathbb{P}\left(\frac{S_n}{\sqrt{n}} > a\right) \sim \overline{\Phi}(a),$$

$$\mathbb{P}\left(\mathscr{D}_n > \Lambda_n + a\sqrt{n}\right) = \mathbb{P}\left(\mathscr{D}_n > \frac{n}{2} + \frac{a}{2}\sqrt{n}\right) \sim \overline{\Phi}(a).$$

Функция на моментите (ФМ)

Дефиниция. Нека X е случайна величина. Ако $\mathbb{E}e^{tX}$ съществува за $t \in (-\varepsilon, \varepsilon)$ и някое $\varepsilon > 0$, то $M_X(t) = \mathbb{E}e^{tX}$ за $t \in (-\varepsilon, \varepsilon)$ се нарича функция на моментите.

$\oplus$ $\mathbb{E}e^{tX} = \sum_i e^{tx_i}\mathbb{P}(X = x_i)$ – винаги съществува за $\forall t$, ако стойностите са краен брой. Но ако не са, тази сума можеда не е сумируема и да отива към ∞.

Но ако вчемем $x_i = j$ и $\mathbb{P}(X = j) = \dfrac{1}{j^2}$, то няма да може да направим сумировката за $t > 0$.

Ако имаме непрекъсната случайна величина, функцията на моментите се изчислява чрез интеграла $\mathbb{E}e^{tX} = \displaystyle\int_{-\infty}^{\infty} e^{tx}f_X(x)\mathrm{d}\,x$, който може да съществува само за някаква част от t, но е важно да съществува за $t \in (-\varepsilon, \varepsilon)$, за да може да го наречем функция на моментите.

$$X \sim Unif(0, 1).$$

$$M_X(t) = \mathbb{E}e^{tX} = \int_0^1 e^{tx} \times f_X(x)\mathrm{d}\,x = \int_0^1 e^{tx}\frac{1}{1-0}\,\mathrm{d}\,x = \frac{1}{t}\int_0^1 e^{tx}\,\mathrm{d}\,tx = \frac{1}{t}e^{tx}\Big|_0^1 = \frac{e^t - 1}{t}$$

е добре дефинирано за всяко $t \in \mathbb{R}$.

Винаги когато нашата случайна величина може да приема краен брой стойности или възможните стойности са в някакъв компакт или някакво крайно множество (в случая са от 0 до 1), то винаги функцията на моментите е добре дефиниране за всяко t.

Дефиниция. X е случайна величина. Тогава:

a) $\mathbb{E}X^k$ се нарича момент от ред $k \geq 1$;
b) $\mathbb{E}|X|^k$ Се нарича абсолютен момент от ред k;
c) $\mathbb{E}(X - \mathbb{E}X)^k$ се нарича централен момент от ред k; $\quad DX = \mathbb{E}(X - \mathbb{E}X)^2$
d) $\mathbb{E}|X - \mathbb{E}X|^k$ се нарича абсолютен централен момент от ред k.

Свойства на M_X. Ще допускаме, че $M_X(t)$ е добре дефинирана за $t \in (-\varepsilon, \varepsilon)$

a) $M_X(0) = \mathbb{E}e^{0.X} = \mathbb{E}1 = 1$;

b) $\left.\dfrac{\partial}{\partial t^k}M_X(t)\right|_{t=0} = \mathbb{E}X^k$, за $\forall k \geq 1$;

c) $M_X(t) = \mathbb{E}e^{tX} \overset{\text{ред на Тейлър}}{=} \mathbb{E}\sum_{k=0}^{\infty}\frac{t^k}{k!}X^k = \sum_{k=0}^{\infty}\frac{t^k}{k!}\mathbb{E}X^k;$

d) Ако $M_{X_n}(t) \xrightarrow[n\to\infty]{} M_X(t)$, за $t \in (-\varepsilon, \varepsilon)$, където $(X_n)_n$ е редица от случайни величини, то $X_n \xrightarrow[n\to\infty]{d} X$;

e) $X \overset{d}{=} Y \Leftrightarrow M_X = M_Y$;

f) Ако $X \perp\!\!\!\perp Y$ и M_X, M_Y са добре дефинирани за $t \in (-\varepsilon, \varepsilon)$, то $M_{X+Y}(t) = \mathbb{E}e^{t(X+Y)} = \mathbb{E}e^{tX}e^{tY} = M_X(t)M_Y(t) \overset{X\perp\!\!\!\perp Y}{=} \mathbb{E}e^{tX}\mathbb{E}e^{tY}$.

Нека например X, Y са непрекъснати с плътности f_X, f_Y. Как да докажем, че $M_X(t) = M_Y(t)$.

1. $f_{X+Y}(x) = \int_{-\infty}^{\infty} f_X(x-y)f_Y(y)\mathrm{d}\,y;$

2. $\int_{-\infty}^{\infty} e^{tX}f_{X+Y}(x)\mathrm{d}\,x = \int_{-\infty}^{\infty} e^{tX}f_X(x)\mathrm{d}\,x \times \int_{-\infty}^{\infty} e^{tX}f_Y(x)\mathrm{d}\,x;$

g) Ако $Y = aX + b$, то $M_Y(t) = e^{bt}M_X(at)$, за всяко t, такова че $M_X(at)$ е добре дефинирано.

Ако M_X е добре дефинирано за $t \in (-\varepsilon, \varepsilon)$, то $M_X(at)$ е добре дефинирано за $-\varepsilon < at < \varepsilon$ и следователно $M_Y(t)$ е добре дефинирано за $-\frac{\varepsilon}{|a|} < t < \frac{\varepsilon}{|a|}$.

$M_Y(t) = \mathbb{E}e^{t(aX+b)} = \mathbb{E}e^{bt}e^{taX} = e^{bt}\mathbb{E}e^{atX} = e^{bt}M_X(at).$

Твърдение. $X \in \mathcal{N}(\mu, \sigma^2)$, то за $\forall t \in \mathbb{R}$, то $M_X(t) = e^{\mu t}e^{\frac{\sigma^2 t^2}{2}}$.

Доказателство. $X = \mu + \sigma Z$, където $Z \in \mathcal{N}(0, 1)$.

$$M_Z(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{tX}e^{-\frac{x^2}{2}}\,\mathrm{d}\,x = \frac{e^{\frac{t^2}{2}}}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}(x-t)^2}\,\mathrm{d}\,x \overset{x-t=y}{=} e^{\frac{t^2}{2}}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}}\,\mathrm{d}\,y = e^{\frac{t^2}{2}}$$

$M_X(t) = e^{\mu t}M_Z(\sigma t) = e^{\mu t}e^{\frac{\mu^2 t^2}{2}}, \forall t \in \mathbb{R}.$

Доказателство (ЦГТ). $\left(X_i\right)_{i=1}^{\infty}$ е редица от независими, еднаквоpазпределени случайни величини с $\mathbb{E}X_1 = \mu$, $\mathbb{D}X_1 = \sigma^2$.

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow[n\to\infty]{d} Z.$$

X_1 има функция на моментите. $M_{X_1}(t)$ е добре дефинирана за $t \in (-\varepsilon, \varepsilon)$

$\dfrac{S_n - n\mu}{\sigma\sqrt{n}} = \dfrac{1}{\sqrt{n}} \sum_{j=1}^{n} \dfrac{X_j - \mu}{\sigma} = \dfrac{1}{\sqrt{n}} \sum_{j=1}^{n} Y_j = \dfrac{V_n}{\sqrt{n}} =: W_n$, където сме положили

$Y_j = \dfrac{X_j - \mu}{\sigma}$, $\forall j \geq 1$ и $(Y_j)_{j=1}^{\infty}$ са независими с еднакво разпределение сл. вел.

$\mathbb{E}Y_1 = \dfrac{\mathbb{E}X_1 - \mu}{\sigma} = 0;\ \mathbb{D}Y_1 = \dfrac{\mathbb{D}X_1}{\sigma^2} = 1.$

$M_{Y_1}(t) = e^{-\frac{\mu}{\sigma}t} M_{X_1}\left(\dfrac{t}{\sigma}\right),\qquad \varepsilon < \dfrac{t}{\sigma} < t\varepsilon \Rightarrow -\sigma\varepsilon < t < \sigma\varepsilon$

M_{Y_1} е добре дефинирана за $|t| < \sigma\varepsilon$.

Нека фиксираме t. Ще докажем, че $M_{W_t} \xrightarrow[n\to\infty]{} M_Z(t) = e^{\frac{t^2}{2}}$.

$$M_{W_n}(t) = \mathbb{E}e^{\frac{t}{\sqrt{n}}V_n} \overset{\text{деф.}}{=} \mathbb{E}e^{\frac{t}{\sqrt{n}}\sum_{j=1}^{n} Y_j} \overset{\text{незав.}}{=} \prod_{j=1}^{n} M_{Y_j}\left(\frac{t}{\sqrt{n}}\right) \overset{Y_j \overset{d}{=} Y_1,\ \forall j}{=} \left[M_{Y_1}\left(\frac{t}{\sqrt{n}}\right)\right]^n.$$

Ако $\left|\dfrac{t}{\sqrt{n}}\right| < \sigma\varepsilon$, то $M_{W_n}(t)$ е добре дефинирано.

$$M_{W_n}(t) = \left[M_{Y_1}\left(\frac{t}{\sqrt{n}}\right)\right],\ M_{Y_1}\left(\frac{t}{\sqrt{n}}\right) = \mathbb{E}e^{\frac{t}{\sqrt{n}}Y_1}.$$

$$e^{\frac{t}{\sqrt{n}}Y_1} = 1 + \frac{t}{\sqrt{n}}Y_1 + \frac{t^2}{2n}Y_1^2 + \frac{\theta(Y_1)t^3Y_1^3}{3!n^{\frac{3}{2}}};\quad |\theta(Y_1)| \leq 1;$$

$$M_{Y_1}(t) = \mathbb{E}e^{\frac{t}{\sqrt{n}}Y_1} = 1 + \frac{t^2}{2n} + \frac{\mathbb{E}\theta(Y_1)t^3Y_1^3}{3!n^{\frac{3}{2}}};$$

$$M_{Y_1}\left(\frac{t}{\sqrt{n}}\right) = 1 + \frac{t^2}{2n} + \frac{t^3}{3!n^{\frac{3}{2}}}\mathbb{E}\theta(Y_1)Y_1^3;$$

$$|\theta(Y_1)Y_1^3| \le |Y_1|^3 \Rightarrow |\mathbb{E}\left(\theta(Y_1)Y_1^3\right)| \le \mathbb{E}|Y_1|^3 = \rho_3.$$

$$M_{W_a}(t) = \left[M_{Y_1}\left(\frac{t}{\sqrt{n}}\right)\right]^n = \left[1 + \frac{t^2}{2n} + \frac{t^2}{2n}\left(\frac{\rho_3}{\sqrt{n}}\right)\right]^n =$$

$$= \left[1 + \frac{t^2}{2n}\left(1 + \underbrace{\frac{\rho_3}{\sqrt{n}}}_{\to 0}\right)\right]^n \sim \left(1 + \frac{t^2}{2n}\right)^n \xrightarrow[n\to\infty]{} e^{\frac{t^2}{2}}$$

$$\Rightarrow \lim_{n\to\infty} M_{W_n}(t) = e^{\frac{t^2}{2}} = M_Z(t)$$

Свойство $\Rightarrow W_n \xrightarrow[n\to\infty]{d} Z;$ $\qquad Z_n = \dfrac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow[n\to\infty]{d} Z.$

$$\mathbb{P}\left(Z_n \in (a,b)\right) \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{y^2}{2}}\,\mathrm{d}y = \Phi(b) - \Phi(a).$$

$\oplus$ $\qquad X_i = \begin{cases} 1, & p \\ 0, & 1-p=q \end{cases}, \qquad X_i \in Ber(p)$

$\dfrac{S_n}{n} \xrightarrow[n\to\infty]{\text{п.с.}} p, \quad \mathbb{E}n = \left|\dfrac{S_n}{n} - p\right|$

$$\mathbb{P}\left(|\mathbb{E}n > \varepsilon|\right) = \mathbb{P}\left(\underbrace{\left|\frac{S_n - np}{\sqrt{n}\sqrt{pq}}\right|}_{n\to\infty} > \frac{\sqrt{n}\varepsilon}{\sqrt{pq}}\right) = \mathbb{P}\left(|Z| > \frac{\sqrt{n}\varepsilon}{\sqrt{pq}}\right) = 2\overline{\Phi}\left(\frac{\sqrt{n}\varepsilon}{\sqrt{pq}}\right) \leq$$

$$\leq 2\overline{\Phi}\left(\frac{\sqrt{n}\varepsilon}{\frac{1}{2}}\right) = 2\overline{\Phi}(2\sqrt{n}\varepsilon) = 2\int_{2\sqrt{n}\varepsilon}^{+\infty} e^{-\frac{y^2}{2}}\,\mathrm{d}\,y\,.$$

Тоерема (Берн-Есеен). Нека $(X_i)_{i=1}^{\infty}$ са независими и еднакво разпределени случайни величини с $\mathbb{E}X_1 = \mu$, $DX_1 = \sigma^2$ и $\mathbb{E}\,|X_1 - \mathbb{E}X_1|^3 = \rho_3$.

Тогава $\operatorname{Sup}_{x\in\mathbb{R}} \left| \mathbb{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) - \underbrace{\Phi(x)}_{=\mathbb{P}(Z<x)} \right| \leq 0{,}4748\frac{\rho^3}{\sigma^{\frac{3}{2}}\sqrt{n}}$.

Следствие. $X \in Bin(n, p)$, то $\lim_{n\to\infty} \mathbb{P}\left(\frac{X - np}{\sqrt{npq}} \leq x\right) = \Phi(x)$

Доказателство. $X = \sum_{j=1}^{n} X_j,\ X_j \in Ber(p)$. Тогава прилагаме ЦГТ с $\mu = p$ и $\sigma^2 = pq = p(1-p)$.