① Сходимость по р.-ю

$$X_n \Rightarrow X$$

$$X_n \xrightarrow{L} X$$

$$\begin{tikzcd}
\boxed{X_n \xrightarrow{\text{п.н.}} X} \arrow[dr] & & \\
 & \boxed{X_n \xrightarrow{P} X} \arrow[r] & \boxed{X_n \xrightarrow{d} X} \\
\boxed{X_n \xrightarrow{L^2} X} \arrow[ur] & &
\end{tikzcd}$$

слабая сх.
сх. по р-ю

Дэвид Хантер
Asymptotic tools $\boxed{\text{Глава 2}}$

<u>Опр.</u> Послед. сл. вел. $X_1, X_2, X_3, \ldots\ldots$ сх. по р.-ю к сл. вел. X, если при $n \to \infty$ послед.

$F_{X_1}(x), F_{X_2}(x), \ldots$ сх. к $F_X(x)$ во всех точках x, где $F_X(x)$ непрерывна

<u>Упражнение</u>

$$F_{X_n}(x) = \begin{cases} 1 - \left(1-\frac{1}{n}\right)^{nx}, & x \geq 0 \\ 0, & x < 0 \end{cases} \qquad \left(1-\frac{1}{n}\right)^{nx} \to e^{-x}$$

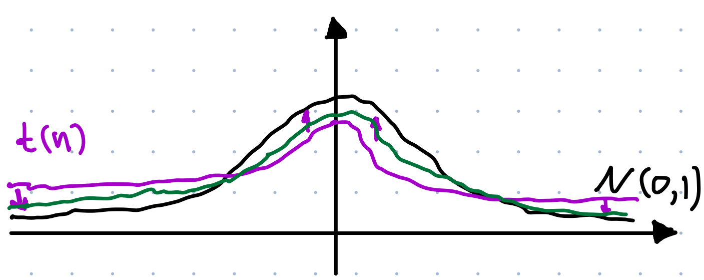
$$X_n \xrightarrow{d} ?$$

$$\lim_{n\to\infty} F_{X_n}(x) = 1 - e^{-x}, \quad x \geq 0$$

$$X_n \xrightarrow{d} Exp(1)$$

<u>Другой пример:</u>

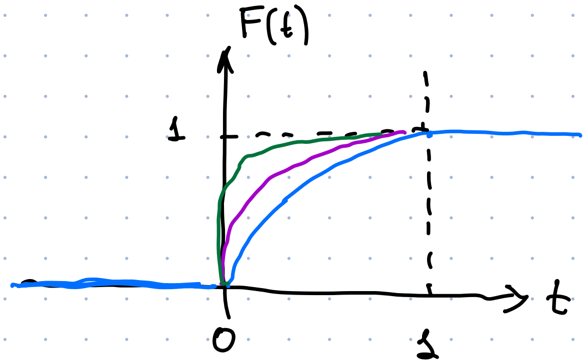
$$t(n) \xrightarrow[n\to\infty]{d} N(0,1)$$



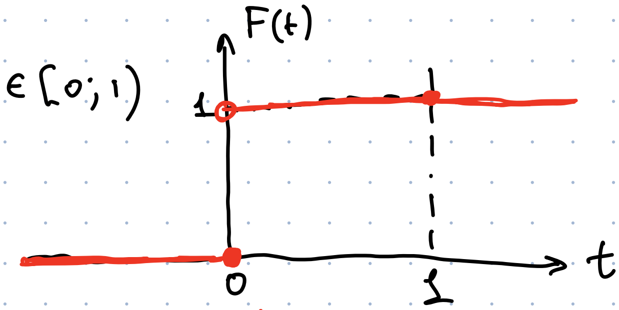
Упражнение

$X_n \sim U[0;1]$

$y_n = \min(X_1, X_2, \dots, X_n)$



$$F_{Y_n}(y) = \begin{cases} 0, & y < 0 \\ 1, & y \geq 1 \\ 1-(1-y)^n, & y \in [0;1) \end{cases}$$



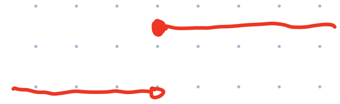
$y < 0 \qquad F_{Y_n}(y) = 0 \xrightarrow[n\to\infty]{} 0$

$y \geq 1 \qquad = 1 \to 1$

$y = 0 \qquad = 1-(1-0)^n = 0 \to 0$

$y \in (0;1) \qquad = 1-(1-y)^n \to 1$

не ф. р.-я



ф. р.-я

$F_X(x) = \mathbb{P}(X \leq x)$

Y	0
$\mathbb{P}(\dots)$	1

$$F(y) = \begin{cases} 1, & y \geq 0 \\ 0, & y < 0 \end{cases}$$

$Y_n \xrightarrow{d} 0$

Т. (Слуцкого)

$X_n \xrightarrow{P} a \neq 0$

$Y_n \xrightarrow{d} Y$

$\Rightarrow$

$Y_n + X_n \xrightarrow{d} Y + a$

$Y_n \cdot X_n \xrightarrow{d} a \cdot Y$

$\dfrac{Y_n}{X_n} \xrightarrow{d} \dfrac{Y}{a}$

② ЦПТ

Бывают как и ЗБЧ очень разными

Т. (Прокопий Петрович Ляпунов)

Пусть $X_1, X_2, \dots, X_n, \dots$ попарно нез. и один. распр. и $\mathbb{E}(X_i^2) < \infty$ тогда при $n \to \infty$

$$\overline{X}_n = \frac{X_1 + \dots + X_n}{n} \xrightarrow{d} N\left(\mathbb{E}(X_1), \frac{Var(X_1)}{n}\right)$$

$$S_n = X_1 + \dots + X_n \xrightarrow{d} N\left(n \cdot \mathbb{E}(X_1), n \cdot Var(X_1)\right)$$

$$Var(n\bar{X}) = n^2 \cdot Var(\bar{X}) = n\, Var(X_1)$$

$$\sqrt{n} \cdot (\overline{X}_n - \mathbb{E}(X_1)) \xrightarrow{d} N(0, Var(X_1))$$

$$Var(\sqrt{n}\, \overline{X}_n) = n \cdot Var(\overline{X}_n) = Var(X_1)$$

$$\sqrt{n} \cdot \frac{\overline{X}_n - \mathbb{E}(X_1)}{\sqrt{Var(X_1)}} \xrightarrow{d} N(0; 1)$$

$$\boxed{\begin{gathered} Y_n \xrightarrow{d} N(0,1) \\ Y_n \overset{asy}{\sim} N(0,1) \end{gathered}}$$

Упражнение (овербукинг)

продают билетов больше чем мест

10% купивших билет не приходит на рейс

300 мест

а) Продали 330 билетов

$\mathbb{P}$(всем хватит места)

$$X_i = \begin{cases} 0, & 1-p=0.1 \text{ не пришёл} \\ 1, & p=0.9 \text{ пришёл} \end{cases}$$

$$\mathbb{E}(X_i) = 0.9 \qquad Var(X_i) = 0.09$$

$$S_{330} = X_1 + \ldots + X_{330} \overset{asy}{\sim} N(297, 29.7)$$

$$\mathbb{P}(S_{330} \leq 300) = \mathbb{P}\left(\frac{S_{330} - 297}{\sqrt{29.7}} \leq \frac{300-297}{\sqrt{29.7}}\right) =$$

$$= \mathbb{P}(Z \leq 0.55) = F_{N(0,1)}(0.55) \approx 0.709$$

$Z \sim N(0,1)$ → В древности считали по таблицам

→ from scipy.stats import norm

norm.cdf(0.55)

8) Сколько билетов надо продать, чтобы переполнение было раз в 10 рейсов.

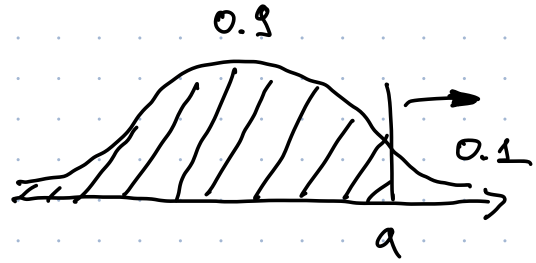
$$S_n \overset{asy}{\sim} N(0.9\cdot n;\ 0.09\cdot n)$$

$$\mathbb{P}(S_n > 300) \leq 0.1$$

$$\mathbb{P}\left(\underbrace{\frac{S_n - 0.9n}{\sqrt{0.09n}}}_{Z} > \underbrace{\frac{300 - 0.9n}{\sqrt{0.09n}}}_{a}\right) \leq 0.1$$

$$1 - \mathbb{P}(Z \leq a) \leq 0.1$$

$$\mathbb{P}(Z \leq a) \geq 0.9$$

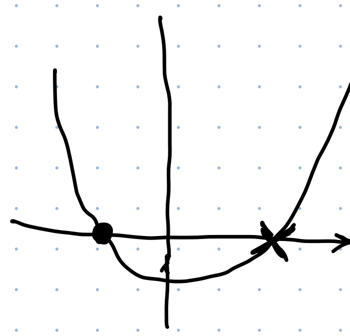


norm.ppf(0.9) $\quad a = 1.28$

$$\frac{300 - 0.9n}{\sqrt{0.09n}} \geq 1.28$$

$$300 - 0.9n \geq 1.28 \cdot 0.3 \cdot \sqrt{n}$$

$x^2 \qquad x$



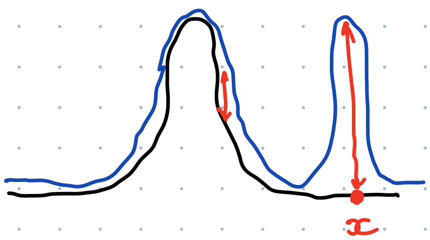
$$n \leq 325.5$$

$$n^* = 325$$

А когда наступает асимптотика?

III. (Нер-во Берри-Эссена)

$$\sup_x \left| F_n(x) - F_{N(0,1)}(x) \right| \leq \frac{C_0 \cdot \mathbb{E}\left[|X_1 - \mu|^3\right]}{\sigma^3 \cdot \sqrt{n}}$$



$\mathbb{E}(X_1) = \mu$

$Var(X_1) = \sigma^2$

$0.4 < C_0 < 0.78$ старая

< 0.4784 новая (2008)

Ирина Шевцова

≤ 0.1

ЦПТ: 0.7

Реальность: 0.7 ± 0.1

Альтернатива: Big Data

③ Зачем всё это, зачем?

$x_1, \dots, x_n$ выборка

F_θ модель

$\hat{\theta}$ точечные оценки

меняем выборку $\Rightarrow$ меняется $\hat{\theta}$

А насколько сильно? $\Rightarrow$ интервальные оценки

$$\mathbb{P}\left(\theta \in [\hat\theta_L ; \hat\theta_R]\right) = 1-\alpha$$

$$\hat\theta \longrightarrow f_{\hat\theta}(t) \longrightarrow$$

д.и. для $\hat\theta$
гипотезы
д.и. для $\hat y_{n+1}$

→ метод моментов
→ метод max правд.

→ ЦПТ
→ Дельта-метод
→ χ^2_n, $t(n)$, F_{nk}
.....
бутстрап

Упражнение

$X_1, \dots, X_n \sim iid\ U[0; a]$ $\quad \hat a = 2\bar X$ $\quad$ ЗБЧ

$$\bar X_n \overset{asy}{\sim} N\left(\frac{a}{2}; \frac{a^2}{12n}\right)$$

$$\mathbb{E}(\bar X_n) = \mathbb{E}\left(\frac{X_1 + \dots + X_n}{n}\right) = \mathbb{E}(X_1) = \frac{a}{2}$$

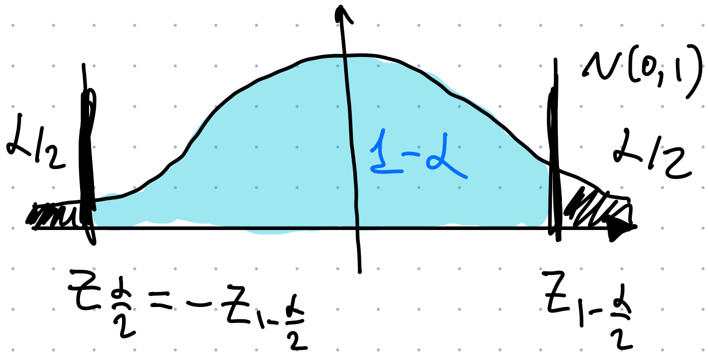
$$Var(\bar X_n) = \dots = \frac{Var(X_1)}{n} = \frac{a^2}{12n}$$

$$\hat a = 2\bar X_n \overset{asy}{\sim} N\left(a; \frac{a^2}{3n}\right) \qquad \frac{\hat a - a}{\sigma/\sqrt{n}} \overset{asy}{\sim} N(0,1)$$

$$\hat a \overset{asy}{\sim} N\left(a; \frac{\hat a^2}{3n}\right) \qquad \frac{\hat a - a}{\hat\sigma/\sqrt{n}} \overset{asy}{\sim} N(0,1)$$

$\sigma^2 = \frac{a^2}{3}$
$\hat\sigma^2 = \frac{\hat a^2}{3}$

$$\frac{\hat a - a}{\hat\sigma/\sqrt{n}} \cdot \frac{\sigma/\sqrt{n}}{\sigma/\sqrt{n}} = \underbrace{\frac{\hat a - a}{\sigma/\sqrt{n}}}_{\to N(0,1)} \cdot \frac{\sigma/\sqrt{n}}{\hat\sigma/\sqrt{n}} \xrightarrow{P} 1$$



$$\mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \le \frac{a-\hat{a}}{\hat{\sigma}/\sqrt{n}} \le z_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

$$\mathbb{P}\left(a - z_{1-\frac{\alpha}{2}} \cdot \frac{\hat{\sigma}}{\sqrt{n}} \le \hat{a} \le a + z_{1-\frac{\alpha}{2}} \cdot \frac{\hat{\sigma}}{\sqrt{n}}\right) = 1-\alpha$$

сл. вел.

предиктивный интервал

$= \frac{\hat{a}^2}{3} \quad \frac{a}{3}$

$$\mathbb{P}\left(\underbrace{\hat{a} - z_{1-\frac{\alpha}{2}} \cdot \frac{\hat{\sigma}}{\sqrt{n}}}_{\hat{a}_L} \le a \le \underbrace{\hat{a} + z_{1-\frac{\alpha}{2}} \cdot \frac{\hat{\sigma}}{\sqrt{n}}}_{\hat{a}_R}\right) = 1-\alpha$$

доверительный интервал

Док-во ЦПТ

→ обычно доказывают через хар. ф.-ии

$$\varphi_X(t) = \mathbb{E}[e^{itX}]$$

→ хочется без!

$$Z \sim \mathcal{N}(0;1) \quad f_Z(t) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}}, \quad t \in \mathbb{R}$$

Th. CLT (Lindeberg – Lévy)

$$X_1, \ldots, X_n, \ldots \sim \text{iid}(\underset{\mathbb{E}(X_i)}{\underset{\parallel}{0}} \; \underset{Var(X_i)}{\underset{\parallel}{1}})$$

$$\Rightarrow \lim_{n\to\infty} \mathbb{P}\left(\frac{X_1 + \ldots + X_n}{\sqrt{n}} \leq t\right) = \mathbb{P}(Z \leq t) \quad \forall t \in \mathbb{R}$$

$$\sqrt{n} \cdot \bar{X} \overset{asy}{\sim} \mathcal{N}(0;1)$$

лемма

Пусть $S_1, S_2, \ldots, S_n, \ldots$ – сл. вел.

Если $\lim\limits_{n\to\infty} \mathbb{E}[h(S_n)] = \mathbb{E}[h(Z)]$

$h: \mathbb{R} \to \mathbb{R}$

$h \; h' \; h'' \; h'''$ – ограничены

тогда $\lim\limits_{n\to\infty} \mathbb{P}(S_n \le t) = \mathbb{P}(Z \le t) \ \forall t \in \mathbb{R}$

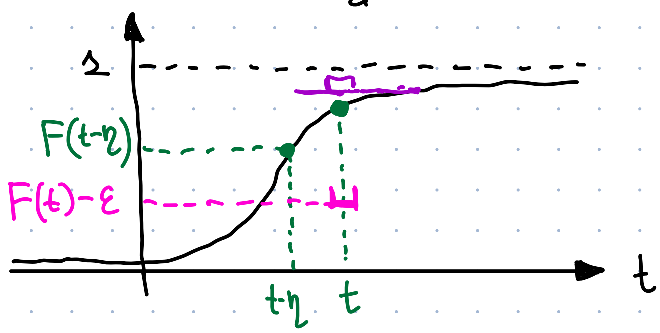
$$\boxed{\begin{array}{ccccc} a_n & \le & c_n & \le & b_n \\ \downarrow & & \downarrow & & \downarrow \\ A & & A & & A \end{array}}$$

<u>Док-во:</u>

$t \in \mathbb{R}$ $\quad \varepsilon > 0$ – заданы

Z – непр. $\quad F_Z(t)$ $\qquad \forall \varepsilon \ \exists \eta:$

картинка слева верна

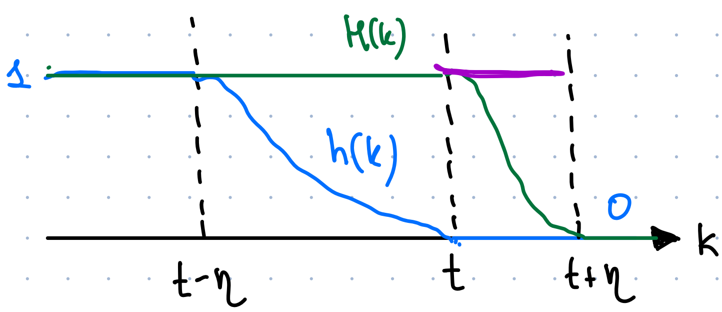


$$F_Z(t) - \varepsilon < F_Z(t-\eta)$$

$$\mathbb{P}(Z \le t) - \varepsilon < \mathbb{P}(Z \le t-\eta)$$

$$\mathbb{P}(Z \le t+\eta) < \mathbb{P}(Z \le t) + \varepsilon$$

$h, H: \mathbb{R} \to [0;1]$ $\qquad$ smooth transition function



$h(k) = 1 \quad k \le t-\eta$

$h(k) = 0 \quad k \ge t$

$H(k) = 1 \quad k \le t$

$H(k) = 0 \quad k \ge t+\eta$

а) $\lim\limits_{n\to\infty} \mathbb{E}(h(S_n)) = \mathbb{E}[h(\underset{S_N}{Z})] \geq \mathbb{P}(Z \leq t-\eta) > \mathbb{P}(Z \leq t) - \varepsilon$

$$1 \cdot \mathbb{P}(Z \leq t-\eta) + \int\limits_{t-\eta}^{t} h(z)\cdot f_Z(z)\,dz + 0 \cdot \mathbb{P}(Z \geq t)$$

б) $\lim\limits_{n\to\infty} \mathbb{E}(H(S_n)) = \mathbb{E}[H(Z)] \leq \mathbb{P}(Z \leq t+\eta) < \mathbb{P}(Z \leq t) + \varepsilon$

$\varepsilon \quad \exists N$

$$\mathbb{P}(Z \leq t) - \varepsilon < \mathbb{E}[h(S_N)] \leq \mathbb{P}(S_n \leq t) \leq$$

$$\leq \mathbb{E}[H(S_N)] < \mathbb{P}(Z \leq t) + \varepsilon$$

$$\mathbb{E}[h(S_N)] = \mathbb{P}(S_n \leq t-\eta) + \int\limits_{t-\eta}^{t} h(s) f_{S_N}(s)\,ds \leq \mathbb{P}(S_n \leq t)$$

$$\mathbb{P}(Z \leq t) - \varepsilon \leq \mathbb{P}(S_N \leq t) \leq \mathbb{P}(Z \leq t) + \varepsilon$$

$\varepsilon = 0$

$$\mathbb{P}(S_N \leq t) \xrightarrow{n\to\infty} \mathbb{P}(Z \leq t)$$

$\blacksquare$

Перейдём к теореме

Док-во:

$Y_1, Y_2, \dots \sim iid\, N(0,1)$ нез. от $X_1, X_2, X_3 \dots$

$$S_{n,i} = \frac{X_1 + X_2 + \dots + X_{i-1} + Y_{i+1} + \dots + Y_n}{\sqrt{n}} \quad \text{для } i = 1, \dots, n$$

$$Z_{n,i} = \frac{X_1 + X_2 + \dots + X_i + Y_{i+1} + \dots + Y_n}{\sqrt{n}} \quad \text{для } i = 0, \dots, n$$

$Z_{n,0} \sim N(0,1)$

$Z_{n,n}, Z_{n,n-1}, Z_{n,n-2} \dots Z_{n0}$ сх. к $N(0,1)$

На каждом шаге один X_i превращается в Y_i

$Z_{n,i} - S_{n,i} = \frac{X_i}{\sqrt{n}}$ $\quad$ $S_{n,i}$ и X_i независимые

$Z_{n,0} \sim N(0,1)$

$$\mathbb{E}[h(Z_{n,n})] \xrightarrow{n\to\infty} \mathbb{E}[h(Z_{n,0})]$$

по лемме $\quad Z_{n,n} \xrightarrow{d} N(0,1)$

$Z_{n,i} \quad Z_{n,i-1}$ $\quad$ $h \quad h, h', h'', h'''$ ограничены

убедимся что они похожи

S_{ni} – тупо сделать шаг между $Z_{n,i}$ $Z_{n,i-1}$

Ряд Тэйлора:

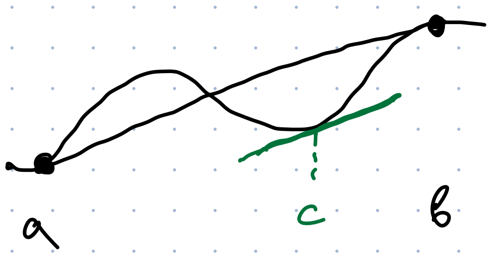
$$f(t) = f(t_0) + f'(t_0)\cdot(t-t_0) + \frac{f''(t_0)}{2!}(t-t_0)^2 + \frac{f'''(t_0)}{3!}(t-t_0)^3 + \dots$$

$+ R(\dots)$ остаточные члены — форма Лагранжа, форма Пеано, ...

Т. Лагранжа

Если f дифф. на $(a;b)$, непр. на $[a;b]$,

тогда $\exists\, c \in (a;b)$: $f(b) - f(a) = f'(c)\cdot(b-a)$



$$h(Z_{n,i}) = h(S_{ni}) + \frac{h'(S_{ni})\cdot \overbrace{[Z_{ni} - S_{ni}]}^{\frac{X_i}{\sqrt{n}}}}{1} + \frac{h''(S_{ni})\cdot X_i^2}{2h} +$$

S_{ni}

$$+ \frac{[h''(C_{ni}) - h''(S_{ni})]\cdot X_i^2}{2h}$$

$= R_{ni}$ — Остаточный член в форме кого-то

C_{ni} лежит между S_{ni} и Z_{ni}

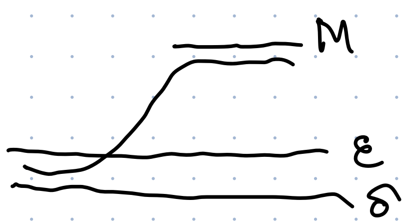
$$h(Z_{n,i}) - h(S_{ni}) - \frac{h'(S_{ni})\, X_i}{\sqrt{n}} - \frac{h''(S_{ni})\cdot X_i^2}{2n} = R_{ni}$$

$$\mathbb{E}\left[h(Z_{ni})\right] - \mathbb{E}\left[h(S_{ni})\right] - 0 - \mathbb{E}\left[\frac{h''(S_{ni})}{2n}\right] = \mathbb{E}(R_{ni})$$

$$\mathbb{E}(|R_{ni}|) = \mathbb{E}(|R_{ni}|\cdot I_A) + \mathbb{E}(|R_{ni}|\cdot I_{\bar{A}})$$

$\varepsilon > 0$ $\quad h'''$ ограничена

можно найти δ: $\quad \delta\cdot|h'''(t)| \le \varepsilon \quad \forall t$



$$h''(x) - h''(y) = h'''(c)\cdot(x-y)$$

$$|x-y| < \delta$$

$$c_{ni} - S_{ni} \qquad |h''(x) - h''(y)| \le \varepsilon$$

$$X_i = Z_{ni} - S_{ni} \qquad |X_i| < \delta$$

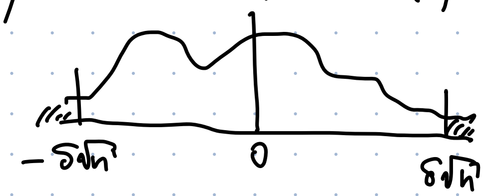
$$\frac{\overbrace{\left[h''(c_{ni}) - h''(S_{ni})\right]}^{\le \varepsilon} X_i^2}{2n} \qquad |c_{ni} - S_{ni}| < \delta$$

$$I_A = I_{|X_i| \le \underbrace{\delta\cdot\sqrt{n}}_{\delta_1}}$$

$$\mathbb{E}(|R_{ni}|\cdot I_A) \le \frac{\varepsilon}{2n}\cdot \mathbb{E}(X_i^2\cdot I_A) \le \frac{\varepsilon}{2n}\mathbb{E}(X_i^2) = \frac{\varepsilon}{2n}$$

$$\mathbb{E}(|R_{ni}|\cdot I_{\bar A}) \le \mathbb{E}\left(\frac{2M}{2n}\cdot X_i^2 I_{\bar A}\right) = \frac{M}{n}\cdot \mathbb{E}(X_i^2\cdot I_{\bar A}) \le \frac{\varepsilon}{2n}$$

$\bar A = \{|X_i| > \delta\sqrt{n}\}$



$\forall t \in \mathbb{R} \quad |h''(t)| \le M$ ф. ограничена

$$\mathbb{E}(|R_{ni}|) \le \frac{\varepsilon}{2n} + \frac{\varepsilon}{2n} = \frac{\varepsilon}{n} \quad \forall i = 1, \dots, n$$

n - большое

$$\left|\mathbb{E}[h(Z_{ni})] - \mathbb{E}[h(S_{ni})] - \mathbb{E}\left[\frac{h''(S_{ni})}{2n}\right]\right| \le \frac{\varepsilon}{n} \quad (*)$$

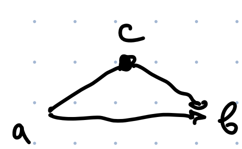
Для $Z_{n,i-1}$ всё то же самое

Роль X_i будет играть Y_i

$$\left|\mathbb{E}[h(Z_{n,i-1})] - \mathbb{E}[h(S_{ni})] - \mathbb{E}\left[\frac{h''(S_{ni})}{2n}\right]\right| \le \frac{\varepsilon}{n} \quad (**)$$

Сложим * и **

$|b-a| \le |c-a| + |b-c|$



$$\left| \mathbb{E}[h(Z_{ni})] - \mathbb{E}[h(Z_{n,i-1})] \right| \leq \frac{2\varepsilon}{n} \quad \Big| \begin{array}{l} \forall i = 1, \dots, n \\ n\text{ - большие} \end{array}$$

Сложим $i = 1, \dots, n$

$$\left| \mathbb{E}[h(Z_{nn})] - \mathbb{E}[h(Z_{n0})] \right| \leq 2\varepsilon$$

$\forall \varepsilon > 0$ и нашли большое n

$$\mathbb{E}[h(Z_{nn})] \longrightarrow \mathbb{E}[h(Z_{n0})]$$

по лемме $Z_{nn} \xrightarrow{d} N(0,1)$