Пример 0

Докажите, что если случайная величина ξ имеет плотность $\gamma_{\alpha,\lambda}(x)$, то $\eta = \alpha\xi$ имеет плотность $\gamma_{1,\lambda}(x)$. Таким образом, параметр α – несущественный (масштабный)

$$\gamma_{\alpha,\lambda}(x) = \begin{cases} \dfrac{\alpha^{\lambda}}{\Gamma(\lambda)} x^{\lambda-1} e^{-\alpha x}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

$$\eta = \alpha\xi \Rightarrow y = \alpha x \Rightarrow x = \frac{1}{\alpha} y = g^{-1}(y) \Rightarrow \left(g^{-1}(y)\right)' = \frac{1}{\alpha}$$

$$p_{\eta}(y) = \begin{cases} \dfrac{\alpha^{\lambda}}{\Gamma(\lambda)} \left(\dfrac{1}{\alpha} y\right)^{\lambda-1} e^{-\alpha \frac{1}{\alpha} y} \cdot \dfrac{1}{\alpha}, & y > 0 \\ 0, & x \le 0 \end{cases} = \begin{cases} \dfrac{1}{\Gamma(\lambda)} y^{\lambda-1} e^{-y} \cdot, & y > 0 \\ 0, & x \le 0 \end{cases} = \gamma_{1,\lambda}$$

Сходимость почти наверное (с вероятностью 1).

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

Определение 2. $\xi_n \xrightarrow{\text{п.н.}} \xi$, если при любом $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\omega\colon \sup_{m\ge n} |\xi_m - \xi| > \varepsilon\right) = 0.$$

Сходимость по вероятности.

Определение 3. $\xi_n \xrightarrow{P} \xi$, если при любом $\varepsilon > 0$

$$\lim_{n\to\infty} P(\omega\colon |\xi_n - \xi| > \varepsilon) = 0$$

Утверждение 2. $\xi_n \xrightarrow{\text{п.н.}} \xi \Rightarrow \xi_n \xrightarrow{P} \xi$.

Утверждение 3. Если последовательность $\{\xi_n\}$ монотонно возрастает (убывает) и $\xi_n \xrightarrow{P} \xi$, то $\xi_n \xrightarrow{\text{п.н.}} \xi$.

Сходимость по распределению (слабая).

Определение 4. $\xi_n \xrightarrow{d} \xi$, или $F_{\xi_n}(x) \Rightarrow F_\xi(x)$, если для любой непрерывной ограниченной функции $g(x)$

$$\int_{-\infty}^{+\infty} g(x)d\,F_{\xi_n}(x) \underset{n\to\infty}{\longrightarrow} \int_{-\infty}^{+\infty} g(x)d\,F_\xi(x)$$

Определение 5. $F_{\xi_n}(x) \Rightarrow F_\xi(x)$, если $F_{\xi_n}(x) \to F_\xi(x)$ в каждой точке непрерывности $F_\xi(x\ )$.

Следовательно, $P(\xi_n < x) \underset{n\to\infty}{\longrightarrow} P(\ \xi < x)$.

Факт без доказательства. Определение 4 и определение 5 эквивалентны.

Утверждение 5. $\xi_n \xrightarrow{P} \xi \Longrightarrow \xi_n \xrightarrow{d} \xi$.

Сходимость в среднем порядка r.

Определение 6. $\xi_n \xrightarrow{r} \xi$, если $M|\xi_n - \ \xi|^r \underset{n\to\infty}{\longrightarrow} 0$.

Утверждение 7. $\xi_n \xrightarrow{r} \xi \Longrightarrow \xi_n \xrightarrow{P} \xi$.

Центральная предельная теорема

Если случайные величины $\xi_1, \xi_2, \ldots$ - независимы, одинаково распределены и имеют конечные $M\xi_i = a,\ D\xi_i = \sigma^2 > 0$, то

$$\lim_{n\to\infty} P\left(\frac{\xi_1 + \xi_2 + \ldots + \xi_n - na}{\sigma\sqrt{n}} < x\right) = \Phi(x)$$

-используется сходимость по распределению

Пример 1

Докажите, используя центральную предельную теорему и задачу 1, что

$$\frac{\sqrt{\lambda}}{\alpha}\gamma_{\alpha,\lambda}\left(x\frac{\sqrt{\lambda}}{\alpha} + \frac{\lambda}{\alpha}\right) \underset{\lambda\to\infty}{\to} \frac{1}{\sqrt{2\pi}}e^{-x^2/2} \ .$$

Решение. Рассмотрим независимые случайные величины $\xi_1, \xi_2, \ldots, \xi_n, \ldots$, распределенные по закону $\gamma_{\alpha,\lambda}(x)$. Математическое ожидание и дисперсия гамма распределения были выведены ранее

$$M\xi_i = \frac{\lambda}{\alpha}, D\xi_i = \frac{\lambda}{\alpha^2}.$$

По центральной предельной теореме случайная величина

$\zeta_n = \frac{\alpha}{\sqrt{n\lambda}}\left(\xi_1+\xi_2+\ldots+\xi_n-\frac{n\lambda}{\alpha}\right)=\frac{\alpha}{\sqrt{n\lambda}}\left(\xi_1+\xi_2+\ldots+\xi_n\right)-\sqrt{n\lambda}$ в пределе имеет стандартное нормальное распределение.

С другой стороны, согласно задаче 1, распределение суммы $\xi_1+\xi_2+\ldots+\xi_n$ будет $\gamma_{\alpha,\lambda+\lambda+\ldots\lambda}=\gamma_{\alpha,n\lambda}$

Пересчитаем плотность ζ_n по формуле $p_{a\xi+b}(x)=\frac{1}{a}p_\xi\left(\frac{x-b}{a}\right)$.

Получаем:

$$p_{\zeta_n}(x)=\frac{\sqrt{n\lambda}}{\alpha}\gamma_{\alpha,n\lambda}\left(\frac{x+\sqrt{n\lambda}}{\alpha}\sqrt{n\lambda}\right)=\frac{\sqrt{n\lambda}}{\alpha}\gamma_{\alpha,n\lambda}\left(x\frac{\sqrt{n\lambda}}{\alpha}+\frac{n\lambda}{\alpha}\right)$$

Осталось заменить в этой формуле переменную $n\lambda\to\infty$ на $\lambda\to\infty$ и получаем, что согласно центральной предельной теореме. Случайная величина $p_{\zeta_n}(x)$ распределена по стандартному нормальному закону.

Пример 2

Имеется 100 квадратов, сторона которых может принимать значения равномерно на $[0.9;1.1]$. С какой вероятностью суммарная площадь всех квадратов будет в пределах от 99 до 101?

Решение. Рассмотрим независимые случайные величины $\xi_1,\xi_2,\ldots,\xi_{100}$, распределенные по закону $R[0.9;1.1]$. Тогда $\sum_{k=1}^{100}\xi_k^2$ имеет смысл суммарной площади.

Воспользуемся ЦПТ:

$$M\xi_k^2=\frac{(b-a)^2}{12}+\left(\frac{a+b}{2}\right)^2=\frac{301}{300};$$

$$M\left(\xi_k^2\right)^2=\int_a^b\frac{x^4}{b-a}dx=\frac{1}{5}\left(a^4+a^3b+a^2b^2+ab^3+b^4\right)$$

$$D\xi_k^2=M\left(\xi_k^2\right)^2-\left(M\xi_k^2\right)^2=\frac{a^4+a^3b+a^2b^2+ab^3+b^4}{5}-\left(\frac{(b-a)^2}{12}+\left(\frac{a+b}{2}\right)^2\right)^2=$$

$$=\frac{(b-a)^2\left(4a^2+7ab+4b^2\right)}{45}=0.01334(2)$$

$$P\left(99 \le \sum_{k=1}^{100} \xi_k^2 \le 101\right) = P\left(\frac{99-301/3}{\sqrt{1.334(2)}} \le \frac{\sum_{k=1}^{100} \xi_k^2 - 301/3}{\sqrt{1.334(2)}} \le \frac{101-301/3}{\sqrt{1.334(2)}}\right) =$$

$$= \Phi\left(\frac{101-301/3}{\sqrt{1.334(2)}}\right) - \Phi\left(\frac{99-301/3}{\sqrt{1.334(2)}}\right) \approx 0.594$$

Пример 3

Урожайность куста картофеля равна 0 кг с вероятностью 0,1, 1 кг с вероятностью 0,2, 1,5 кг с вероятностью 0,2, 2 кг с вероятностью 0,3 и 2,5 кг с вероятностью 0,2. Какое наименьшее число клубней надо посадить, чтобы с вероятностью не менее 0,975 урожай был не менее 1 тонны?

Решение. Рассмотрим независимые случайные величины $\xi_1, \xi_2, \ldots, \xi_n$ - урожайность куста картофеля, они распределены по закону:

$P(\xi_k = 0) = 0.1$, $P(\xi_k = 1) = 0.2$, $P(\xi_k = 1.5) = 0.2$, $P(\xi_k = 2) = 0.3$, $P(\xi_k = 2.5) = 0.2$

x_i	0	1	1.5	2	2.5
$P(\xi_k = x_i)$	0.1	0.2	0.2	0.3	0.2

Тогда нужно найти такое количество кустов картофеля n, чтобы с вероятностью 0.975 урожайность была больше или равна 1000 кг, то есть $P\left(\sum_{k=1}^{n} \xi_k \geq 1000\right) \geq 0.975$.

Воспользуемся ЦПТ: $M\xi_k = 1.6,\ M\xi_k^2 = 3.1,\ D\xi_k = 0.54$

$$P\left(\sum_{k=1}^{n} \xi_k \geq 1000\right) = P\left(\frac{\sum_{k=1}^{n} \xi_k - n\, M\xi_k}{\sqrt{nD\xi_k}} \geq \frac{1000 - nM\xi_k}{\sqrt{nD\xi_k}}\right) \approx$$

$$= 1 - \Phi\left(\frac{1000 - 1.6n}{\sqrt{0.54n}}\right) \geq 0.975 \to \Phi\left(\frac{1000 - 1.6n}{\sqrt{0.54n}}\right) \leq 0.025 \to \frac{1000 - 1.6n}{\sqrt{0.54n}} \leq -1.96 \to$$

$$1.6n - 1.96\sqrt{0.54n} - 1000 \geq 0 \to n > 648.$$

Пример 4

Вы экономист в государственной службе статистики. Ваша задача — оценить средемесячный доход домохозяйства μ в крупном регионе с $N = 1.5$ млн домохозяйств.

Проблема: Провести сплошной опрос всех N домохозяйств дорого и долго. Вместо этого вы случайным образом отбираете $n = 1000$ домохозяйств и вычисляете выборочное среднее $\bar{X}$. Вопрос: насколько $\bar{X}$ близко к истинному μ?

Решение: Применяем Центральную Предельную Теорему (ЦПТ).

ЦПТ утверждает, что если $X_1, X_2, \ldots, X_n$ — независимые, одинаково распределенные случайные величины с конечным математическим ожиданием μ и дисперсией σ^2, то при достаточно большом n распределение их выборочного среднего стремится к нормальному:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{d} \mathcal{N}(\mu, \frac{\sigma^2}{n})$$

Или в стандартизированной форме:

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \xrightarrow{d} \mathcal{N}(0,1)$$

Распределение доходов X_i в генеральной совокупности не является нормальным. Оно асимметрично (скошено вправо). Однако, согласно ЦПТ, распределение выборочного среднего $\bar{X}$ будет приблизительно нормальным.

Результаты выборки: Вы провели опрос и получили:

Выборочное среднее: $\bar{x} = 60000$ руб.

Известное из предыдущих исследований стандартное отклонение по совокупности: $\sigma = 25000$ руб.

Объем выборки: $n = 1000$.

Стандартная ошибка: Рассчитывается по формуле:

$$\text{SE} = \frac{\sigma}{\sqrt{n}} = \frac{25000}{\sqrt{1000}} \approx \frac{25000}{31.62} \approx 791.$$

Это стандартное отклонение распределения выборочных средних.

Построение доверительного интервала: Так как $\bar{X}$ распределено нормально, мы можем построить 95%-ый доверительный интервал для истинного среднего μ:

$$\bar{X} \pm z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

где $z_{1-\alpha/2}$ — квантиль стандартного нормального распределения. Для 95% доверия ($\alpha = 0.05$): $z_{0.975} \approx 1.96$.

Подставляем значения:

$60000 \pm 1.96 \cdot 791$

60000 ± 1550

Итоговый вывод:

$P(58450 \le \mu \le 61550) \approx 0.95$

Мы можем быть на 95% уверены, что истинный средний доход домохозяйств в регионе лежит в интервале от 58 450 до 61 550 рублей.

ЦПТ является фундаментальным инструментом в экономике и эконометрике, так как позволяет:

Использовать выборочные данные для оценки параметров генеральной совокупности.

Строить доверительные интервалы и проверять статистические гипотезы (t-тесты, ANOVA), даже если исходное распределение не является нормальным.

Теорема 1 (неравенства Чебышева). Для любого $x > 0$ имеют место неравенства:

$$P\left(|\xi| \ge x\right) \le \frac{M|\xi|}{x}, \qquad P\left(|\xi - M\xi| \ge x\right) \le \frac{D\xi}{x^2}.$$

Доказательство.

$$|\xi| = |\xi| \cdot I_{(|\xi|\ge x)} + |\xi| \cdot I_{(|\xi|<x)} \ge |\xi| \cdot I_{(|\xi|\ge x)} \ge x \cdot I_{(|\xi|\ge x)},$$

$$M|\xi| \ge M\left[x \cdot I_{(|\xi|\ge x)}\right] = xM\left[I_{(|\xi|\ge x)}\right] = xP\left(|\xi| \ge x\right)$$

Второе неравенство получается из первого подстановкой $\left(\xi - M\xi\right)^2$ вместо ξ:

$$P\left(|\xi - M\xi| \ge x\right) = P\left(\left(\xi - M\xi\right)^2 \ge x^2\right) \le \frac{M\left(\xi - M\xi\right)^2}{x^2} = \frac{D\xi}{x^2}.$$

Следствие. $P\left(|\xi - M\xi| < x\right) \ge 1 - \frac{D\xi}{x^2}$

Теорема 2 (закон больших чисел Чебышева). Пусть $\xi_1, \xi_2, \ldots$ - независимые случайные величины и существует такая константа $c > 0$, что все $D\xi_i \le c$, $i = 1, 2, \ldots$

Тогда при любом $\varepsilon > 0$

$$\lim_{N\to\infty} P\left(\left|\frac{\xi_1 + \xi_2 + \ldots + \xi_N}{N} - \frac{M\xi_1 + M\xi_2 + \ldots + M\xi_N}{N}\right| < \varepsilon\right) = 1.$$

Теорема 4 (Ляпунов). Если случайные величины $\xi_1, \xi_2, \ldots$ - независимы, имеют конечные моменты $M\xi_k = a_k$, $D\xi_k = b_k^2$, $M|\xi_k - a_k|^3 = c_k^3$, пусть $A_n = \sum_{k=1}^{n} a_k$, $B_n^2 = \sum_{k=1}^{n} b_k^2$, $C_n^3 = \sum_{k=1}^{n} c_k^3$, причем $\frac{C_n}{B_n} \to 0$,то

$$\lim_{n\to\infty} P\left(\frac{\xi_1 + \xi_2 + \ldots + \xi_n - A_n}{B_n} < x\right) = \Phi(x)$$

-используется сходимость по распределению: $\frac{\xi_1 + \xi_2 + \ldots + \xi_n - A_n}{B_n} \xrightarrow{d} \mathcal{N}(0,1)$

Теорема 5 (теорема Хинчина). Пусть $\xi_1, \xi_2, \ldots$ - одинаково распределенные случайные величины с конечным математическим ожиданием a.

Тогда при любом $\varepsilon > 0$

$$\lim_{N\to\infty} P\left(\left|\frac{\xi_1+\xi_2+\ldots+\xi_N}{N}-a\right|<\varepsilon\right)=1.$$

-используется сходимость по вероятности: $\dfrac{\xi_1+\xi_2+\ldots+\xi_N}{N}\xrightarrow{P} a$

Теорема 6 (теорема Маркова). Пусть $\xi_1,\xi_2,\ldots$ - последовательность случайных величин, удовлетворяющая условию $\lim_{n\to\infty}\dfrac{1}{n^2}D\left[\sum_{k=1}^{n}\xi_k\right]=0$. Тогда при любом $\varepsilon>0$

$$\lim_{N\to\infty} P\left(\left|\frac{\xi_1+\xi_2+\ldots+\xi_N}{N}-\frac{M\xi_1+M\xi_2+\ldots+M\xi_N}{N}\right|<\varepsilon\right)=1.$$

- используется сходимость по вероятности:

$$\frac{\xi_1+\xi_2+\ldots+\xi_N}{N}-\frac{M\xi_1+M\xi_2+\ldots+M\xi_N}{N}\xrightarrow{P}0$$

Теорема 7(необходимое и достаточное условие). Пусть $\xi_1,\xi_2,\ldots$ - последовательность случайных величин с $M\xi_k=a_k$. Тогда закон больших чисел имеет место тогда и только тогда, когда $\lim_{N\to\infty} M\dfrac{\left(\frac{1}{n}\sum_{k=1}^{n}(\xi_k-a_k)\right)^2}{1+\left(\frac{1}{n}\sum_{k=1}^{n}(\xi_k-a_k)\right)^2}=0$. Тогда при любом $\varepsilon>0$

$$\lim_{N\to\infty} P\left(\left|\frac{\xi_1+\xi_2+\ldots+\xi_N}{N}-\frac{M\xi_1+M\xi_2+\ldots+M\xi_N}{N}\right|<\varepsilon\right)=1.$$

-используется сходимость по вероятности:

$$\frac{\xi_1+\xi_2+\ldots+\xi_N}{N}-\frac{M\xi_1+M\xi_2+\ldots+M\xi_N}{N}\xrightarrow{P}0$$

Теорема 8. (усиленный закон больших чисел Колмогорова) Пусть $\xi_1,\xi_2,\ldots$ - независимые и одинаково распределенные случайные величины. Для того, чтобы

$\dfrac{\xi_1+\xi_2+\ldots+\xi_n}{n}\xrightarrow{п.н.} a$ необходимо и достаточно, чтобы существовало конечное $M\xi_n=a$.

- используется сходимость почти наверное: $\frac{\xi_1+\xi_2+\ldots+\xi_n}{n} \xrightarrow{п.н.} a$.

Пример 4

В примере 2 одинаково распределены.

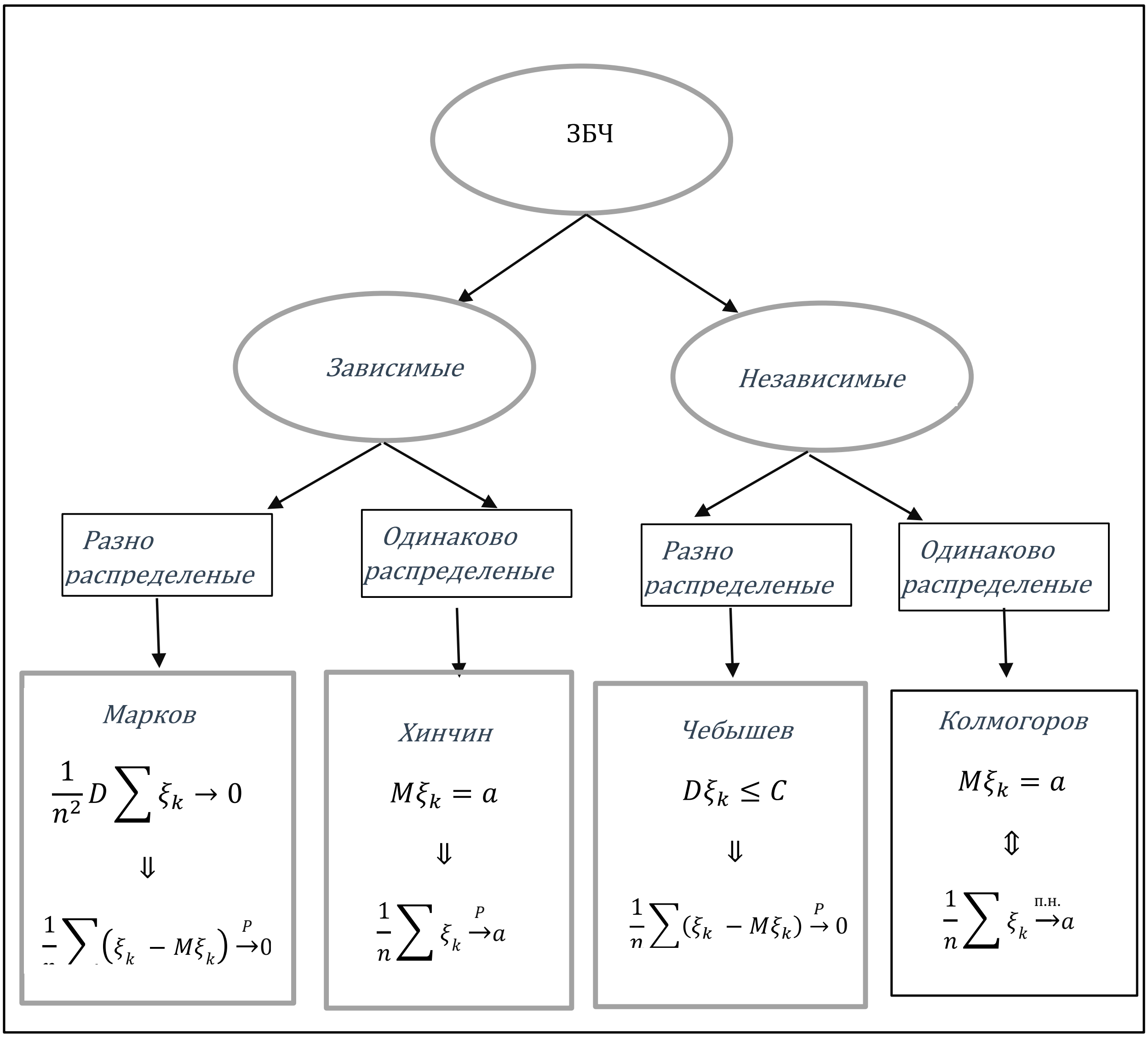
Пример 5

Последовательность независимых и одинаково распределенных случайных величин $\xi_1, \xi_2, \dots, \xi_n, \dots$ задана рядом распределения $P(\xi_i = k) = \dfrac{1}{k^3 \varsigma(3)}$, $k = 1, 2, \dots,$

где $\varsigma(3) = \sum_{k=1}^{\infty} \dfrac{1}{k^3} = 1,20256 -$ значение функции Римана при аргументе 3. Проверьте, применим ли к этой последовательности закон больших чисел. Проверим ограничены ли дисперсии. Найдем математическое ожидание

$$M\xi_i = 1 \cdot \frac{1}{1^3 \varsigma(3)} + 2 \cdot \frac{1}{2^3 \varsigma(3)} + \cdots = \sum_{k=1}^{\infty} \frac{1}{k^2 \varsigma(3)} = \frac{1}{\varsigma(3)} \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6\varsigma(3)}$$

Математическое ожидание существует, следовательно применим закон больших чисел Колмогорова



ЗБЧ
Зависимые
Независимые
Разно распределеные
Одинаково распределеные
Разно распределеные
Одинаково распределеные
Марков
$\frac{1}{n^2} D \sum \xi_k \to 0$
$\Downarrow$
$\frac{1}{n} \sum (\xi_k - M\xi_k) \xrightarrow{P} 0$
Хинчин
$M\xi_k = a$
$\Downarrow$
$\frac{1}{n} \sum \xi_k \xrightarrow{P} a$
Чебышев
$D\xi_k \leq C$
$\Downarrow$
$\frac{1}{n} \sum (\xi_k - M\xi_k) \xrightarrow{P} 0$
Колмогоров
$M\xi_k = a$
$\Updownarrow$
$\frac{1}{n} \sum \xi_k \xrightarrow{\text{п.н.}} a$