1.33. Неравенство Чебышева

$\underline{T|}$ Неравенство Чебышева

Пусть случайная величина X имеет второй начальный момент MX^2, тогда: $\forall \varepsilon > 0 \;\; P(|X| \geq \varepsilon) \leq \dfrac{MX^2}{\varepsilon^2}$ – *неравенство Чебышева* (∗)

$\underline{Док|}$ (X – непрерывная случайная величина)

$$MX^2 = \int_{-\infty}^{\infty} x^2 f_X(x)dx = \int_{-\infty}^{-\varepsilon} + \int_{-\varepsilon}^{\varepsilon} + \int_{\varepsilon}^{\infty} \geq \int_{-\infty}^{-\varepsilon} + \int_{\varepsilon}^{\infty} = \int_{|x|\geq\varepsilon} x^2 f_X(x)dx \geq \varepsilon^2 \int_{|x|\geq\varepsilon} f_X(x)dx =$$

$$= \varepsilon^2 P(|X| \geq \varepsilon) \Rightarrow P(|X| \geq \varepsilon) \leq \frac{MX^2}{\varepsilon^2}.$$

(Для дискретной случайной величины доказательство аналогично).

Если случайная величина X имеет конечные математическое ожидание MX и дисперсию $DX = \sigma^2$, то, заменяя в неравенстве (∗) X на (X – MX), получим другую форму *неравенства Чебышева*:

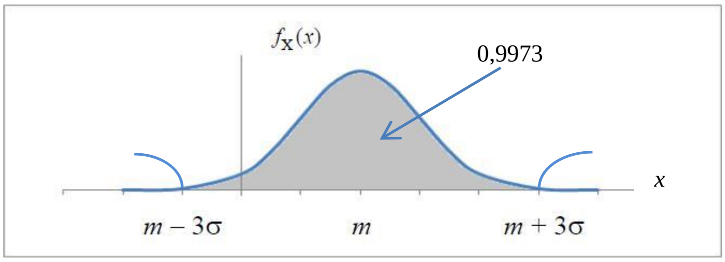
$$\forall \varepsilon > 0 \;\; P(|X - MX| \geq \varepsilon) \leq \frac{DX}{\varepsilon^2}.$$

Следствие. При $\varepsilon = 3\sigma$ из последнего неравенства получаем, что *для любой случайной величины* X, *имеющей математическое ожидание* MX *и дисперсию* $DX = \sigma^2$ имеет место "3-σ *оценка*" вероятности абсолютного отклонения случайной величины от ее математического ожидания на величину, *не меньшую* 3σ:

$$P(|X - MX| \geq 3\sigma) \leq 1/9.$$

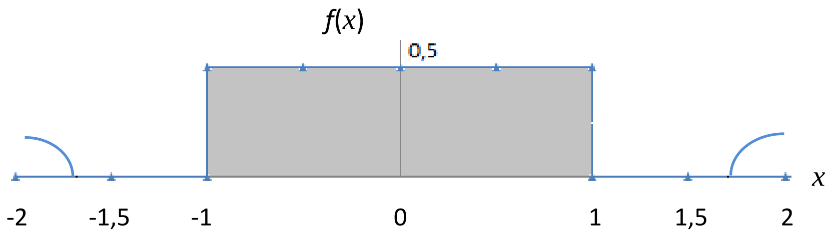
Эта оценка справедлива для *любого* распределения (при условии ∃MX, ∃DX), однако она является довольно грубой. Так, например,
(1) для нормального распределения имеем:

$$P(|X - MX| \geq 3\sigma) = 1 - 0{,}9973 = 0{,}0027 \leq 1/9$$



(2) для равномерного распределения: $P(|X - MX| \geq 3\sigma) = 0 \leq 1/9$

Равномерное распределение на [–1;1]:



$$3\sigma_X = 3\frac{2}{2\sqrt{3}} = \sqrt{3} \approx 1{,}7; \quad P(|X| \geq \sqrt{3}) = 0\,.$$

1.34. Закон больших чисел (теорема Чебышева)

Определение независимости *двух* случайных величин – см. п. 1.25.

def| определение независимости n ($n \geq 2$) случайных величин:

(а) непрерывные случайные величины $X_1, \dots, X_n$ – независимы, если $\forall(x_1, \dots, x_n)$ события $(X_1 < x_1), \dots, (X_n < x_n)$ независимы в совокупности;

(б) дискретные случайные величины $X_1, \dots, X_n$ – независимы, если $\forall(x_1, \dots, x_n)$ события $(X_1 = x_1), \dots, (X_n = x_n)$ независимы в совокупности.

Замечание:

n-мерной случайной величиной (n-мерным случайным вектором) называется система n одномерных случайных величин $X_1, \dots, X_n$, такая, что для любых вещественных $x_1, \dots, x_n$ определена вероятность произведения n событий (функция совместного распределения):

$$\forall(x_1, \dots, x_n) \quad P((X_1 < x_1), \dots, (X_n < x_n)) = F_{X_1,\dots,X_n}(x_1,\dots,x_n).$$

Определение независимости n случайных величин $X_1, \dots, X_n$ ($n \geq 2$) может быть сформулировано так: случайные величины $X_1, \dots, X_n$ независимы, если для любых чисел $x_1, \dots, x_n$ справедливо равенство:

$$\forall(x_1, \dots, x_n) \quad F_{X_1,\dots,X_n}(x_1,\dots,x_n) = F_{X_1}(x_1) \cdot F_{X_2}(x_2) \cdot \ldots \cdot F_{X_n}(x_n).$$

(*функция совместного распределения n-мерной случайной величины* ($X_1, \dots, X_n$) *равна произведению функций распределения компонент* $X_1, \dots, X_n$)

def| *определение последовательности независимых случайных величин*

Последовательность $\{X_n\}$ случайных величин называется *последовательностью взаимно независимых* (независимых в совокупности) случайных величин, если $\forall n$ $X_1, \dots, X_n$ – – независимые случайные величины.

Т| Закон больших чисел (теорема Чебышева)

Пусть $\{X_i\}$ – последовательность независимых случайных величин и пусть каждый член последовательности X_i имеет конечное математическое ожидание $MX_i = m_i$ и дисперсию $DX_i = D_i$, причем дисперсии ограничены в совокупности ($\exists B\ 0 < B < +\infty\ \forall i\ D_i \le B < +\infty$), тогда:

$$\forall \varepsilon > 0\ P(|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} m_i| \ge \varepsilon) \underset{n\to+\infty}{\to} 0. \qquad \text{(ЗБЧ)}$$

Док| Заметим: $M(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} MX_i = \frac{1}{n}\sum_{i=1}^{n} m_i$,

$D(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} DX_i = \frac{1}{n^2}\sum_{i=1}^{n} D_i \le \frac{nB}{n^2} = \frac{B}{n}$ и запишем

неравенство Чебышева (см. п. 1.33.) для случайной величины $\frac{1}{n}\sum_{i=1}^{n} MX_i$:

$$\forall \varepsilon > 0 \quad 0 \le P(|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} m_i| \ge \varepsilon) \le \frac{D(\frac{1}{n}\sum_{i=1}^{n} X_i)}{\varepsilon^2} \le \frac{B}{n\varepsilon^2}.$$

По теореме о сжатой функции при $n \to \infty$ получаем:

$$\lim_{n\to\infty} P(|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} m_i| \ge \varepsilon) = 0.$$

Замечание 1.

Смысл закона больших чисел состоит в том, что абсолютная величина отклонения среднего арифметического n случайных величин $\frac{1}{n}\sum_{i=1}^{n} X_i$ от среднего арифметического их математических ожиданий $\frac{1}{n}\sum_{i=1}^{n} m_i$ на любое фиксированное $\varepsilon > 0$ *с увеличением n становится все менее вероятным.*

Замечание 2.

Доказанная теорема справедлива и для последовательности $\{X_i\}$ *попарно независимых,* а также для *попарно некоррелированных* случайных величин (поскольку при этом $D(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} D(X_i)$, см. п.1.27.).

1.35. Сходимость по вероятности; два следствия закона больших чисел

def Последовательность случайных величин $\{S_n\}$ называют *сходящейся по вероятности* к случайной (или неслучайной) величине S, если

$$\forall\varepsilon > 0 \ \mathrm{P}(|S_n - S| \geq \varepsilon) \underset{n\to+\infty}{\to} 0$$

или, равносильно, $\forall\varepsilon > 0 \ \mathrm{P}(|S_n - S| < \varepsilon) \underset{n\to+\infty}{\to} 1.$

Замечание

Сходимость по вероятности $\{S_n\}$ к S означает, что *вероятность* события $(|S_n - S| \geq \varepsilon)$, состоящего в том, что S_n *оказывается вне* ε-*окрестности* S, с ростом n становится сколь угодно близкой к нулю (иначе: *вероятность* события $(|S_n - S| < \varepsilon)$, состоящего в том, что S_n *находится в* ε-*окрестности* S, с ростом n становится сколь угодно близкой к единице). Для сходимости по вероятности принято обозначение: $S_n \underset{n\to+\infty}{\xrightarrow{P}} S$.

T *Следствие* 1 *из закона больших чисел.*

Пусть $\{\mathrm{X}_i\}$ – последовательность *независимых одинаково распределенных случайных величин*, имеющих математическое ожидание $\mathrm{MX}_i = m$ и дисперсию $\mathrm{DX}_i = \sigma^2$ (m и σ^2 – одни и те же для всех X_i), $\bar{\mathrm{X}} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{X}_i$ – среднее арифметическое $\mathrm{X}_1,\ldots,\mathrm{X}_n$, тогда

среднее арифметическое независимых одинаково распределенных случайных величин сходится по вероятности к их математическому ожиданию:

$$\bar{\mathrm{X}} \underset{n\to+\infty}{\xrightarrow{P}} m.$$

Док Подставляя $\frac{1}{n}\sum_{i=1}^{n} m_i = m$ и $\frac{1}{n}\sum_{i=1}^{n}\mathrm{X}_i = \bar{\mathrm{X}}$ в (ЗБЧ) см. п.1.34., получаем:

$$\forall\varepsilon > 0 \ \mathrm{P}(|\bar{\mathrm{X}} - m| \geq \varepsilon) \underset{n\to+\infty}{\to} 0 \ \Leftrightarrow \ \bar{\mathrm{X}} \underset{n\to+\infty}{\xrightarrow{P}} m.$$

$\underline{T}|$ *Следствие* 2 *из закона больших чисел* (теорема Бернулли).

Пусть случайная величина X – число наступлений события A (успеха) при проведении n независимых испытаний по схеме Бернулли, $p = \mathrm{P}(A)$ – вероятность успеха в каждом испытании, $\frac{X}{n}$ – относительная частота числа успехов в n испытаниях, тогда *относительная частота* $\frac{X}{n}$ *числа успехов в* n *испытаниях по схеме Бернулли по вероятности сходится к вероятности успеха* p *в каждом испытании*:

$$\frac{X}{n} \xrightarrow[n\to+\infty]{P} p.$$

$\underline{Док}|$ $X = \sum_{i=1}^{n} X_i$ – сумма n индикаторов успеха X_i ($i = 1,\dots,n$),

$MX_i = p$; $DX_i = p(1-p)$; $\frac{X}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ – среднее арифметическое

$X_1,\dots, X_n$, поэтому, в соответствии со следствием 1, имеем:

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to+\infty]{P} p \quad \text{или} \quad \frac{X}{n} \xrightarrow[n\to+\infty]{P} p.$$

1.36. Центральная предельная теорема

Предварительные замечания.

Рассмотрим последовательность $\{X_i\}$ независимых случайных величин ($\forall n$ $X_1, \ldots, X_n$ – независимы в совокупности) с математическими ожиданиями $\{m_i\}$ и дисперсиями $\{D_i\}$.

При любом фиксированном n имеем:

$$M(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} MX_i = \sum_{i=1}^{n} m_i,$$

$$D(\sum_{i=1}^{n} X_i) = \left[\text{в силу независимости } X_i\right] = \sum_{i=1}^{n} DX_i = \sum_{i=1}^{n} D_i.$$

Случайная величина $\overset{\circ}{Y}_n = \dfrac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} m_i}{\sqrt{\sum_{i=1}^{n} D_i}}$ – центрированная и нормированная, ее математическое ожидание $M\overset{\circ}{Y}_n = 0$, а дисперсия $D\overset{\circ}{Y}_n = 1$.

Термин *центральная предельная теорема* (ЦПТ) вообще означает ряд теорем, утверждающих (при каких-либо условиях) справедливость предельного равенства вида:

$$\lim_{n\to\infty} P(\overset{\circ}{Y}_n < x) = \lim_{n\to\infty} P\left(\frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} m_i}{\sqrt{\sum_{i=1}^{n} D_i}} < x\right) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt \quad (*).$$

Если указанное предельное равенство $(*)$ имеет место, то при достаточно больших значениях n на его основе строится приближение:

$$P\left(\frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} m_i}{\sqrt{\sum_{i=1}^{n} D_i}} < x\right) \approx \Phi(x).$$

Считается, что такое приближение становится приемлемым уже при $n > 30$.

Т| Центральная предельная теорема (ЦПТ) *для независимых одинаково распределенных случайных величин* (*без доказательства*)

Пусть $\{X_i\}$ – последовательность независимых *одинаково распределенных* случайных величин, имеющих математическое ожидание m и дисперсию $\sigma^2 \neq 0$, тогда:

$$\forall x \ \lim_{n\to\infty} P\Big(\frac{\sum_{i=1}^{n} X_i - nm}{\sigma\sqrt{n}} < x\Big) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt .$$

Заметим: $M(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} MX_i = nm$; $D(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} DX_i = n\sigma^2$.

Символически утверждение ЦПТ записывают так:

$$\sum_{i=1}^{n} X_i \underset{n\to\infty}{\sim} N(nm; \sigma\sqrt{n}) \quad \text{или} \quad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \underset{n\to\infty}{\sim} N(m; \frac{\sigma}{\sqrt{n}})$$

– среднее арифметическое *любых независимых одинаково распределенных случайных величин* (имеющих математическое ожидание и дисперсию) *асимптотически распределено по нормальному закону.*

Две последние *символические* записи удобны для применения ЦПТ в “*допредельной*” форме – при *конечных достаточно больших значениях* n.

1.37. Предельная теорема Муавра-Лапласа

Т Пусть X – число успехов при проведении n независимых испытаний (по схеме Бернулли) с вероятностью p успеха в каждом испытании (MX= np, DX = npq), при этом

$\forall n$ случайная величина $\dfrac{\mathrm{X}-\mathrm{MX}}{\sqrt{\mathrm{DX}}} = \dfrac{\mathrm{X}-np}{\sqrt{npq}}$ – центрированная и нормированная, тогда, согласно ЦПТ, справедливо утверждение:

$$\forall x \ \lim_{n\to\infty} \mathrm{P}\left(\frac{\mathrm{X}-np}{\sqrt{npq}}<x\right) = \Phi(x).$$

Иными словами, теорема утверждает, что случайная величина X – число успехов в схеме Бернулли, распределена асимптотически нормально:

$$\mathrm{X} \underset{n\to+\infty}{\sim} N(np;\sqrt{npq}).$$

Док Пусть X_i – индикатор появления успеха в i–м испытании:

X_i	0	1
P	q	p

$\mathrm{MX}_i = p$, $\mathrm{DX}_i = pq$, тогда $\mathrm{X} = \sum_{i=1}^{n} \mathrm{X}_i$ – сумма n одинаково распределенных независимых случайных величин $\mathrm{X}_1, \ldots, \mathrm{X}_n$ удовлетворяет требованиям центральной предельной теоремы, откуда и следует утверждение теоремы Муавра-Лапласа.

Замечание 1.

Случайная величина X (число успехов в схеме Бернулли) подчиняется биномиальному распределению:

$$\mathrm{P}(\mathrm{X}=k) = b_k\,(n;p) = C_n^k p^k (1-p)^{n-k} \quad (k=0,1,\ldots,n).$$

Применяя теорему Муавра-Лапласа в “допредельной” форме (при достаточно больших n), получаем *интегральную приближенную формулу Муавра-Лапласа* для подсчета сумм биномиальных вероятностей:

$$\sum_{k=0}^{k_m} b_k(n;p) = \mathrm{P}(0 \le \mathrm{X} \le k_m) \approx \Phi\left(\frac{k_m - np}{\sqrt{npq}}\right) = \frac{1}{2} + \Phi_0\left(\frac{k_m - np}{\sqrt{npq}}\right)$$

Обоснование этой формулы: поскольку $\mathrm{X} \underset{n\to +\infty}{\sim} N(np;\sqrt{npq})$, то

$$\mathrm{P}(0\le \mathrm{X} \le k_m) = \mathrm{P}(-\infty < \mathrm{X} \le k_m) \approx \Phi\left(\frac{k_m - np}{\sqrt{npq}}\right) - \Phi(-\infty) = \Phi\left(\frac{k_m - np}{\sqrt{npq}}\right).$$

Полезна также приближенная формула:

$$\mathrm{P}(k_1 \le \mathrm{X} \le k_2) = \sum_{k=k_1}^{k=k_2} b_k(n;p) \approx \Phi\left(\frac{k_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1 - np}{\sqrt{npq}}\right)$$

Это приближение применяют при значениях p не слишком близких к нулю, величинах k_1 и k_2 – порядка несколько десятков и $\frac{k_1 - np}{\sqrt{npq}}$ и $\frac{k_2 - np}{\sqrt{npq}}$ – порядка нескольких единиц (обычно в пределах от 0 до 5).

Задача

Проведено n независимых испытаний по схеме Бернулли.

Вычислить приближенно вероятность отклонения относительной частоты $\frac{\mathrm{X}}{n}$ числа успехов от вероятности p на величину, не превосходящую по абсолютной величине $\varepsilon > 0$: $\mathrm{P}(|\frac{\mathrm{X}}{n} - p| \le \varepsilon) = ?$

Величины n, p и ε заданы.

Указание: использовать приближение, основанное на теореме Муавра-Лапласа.

Решение

Напомним: $\mathrm{Y} \sim N(m;\sigma)$, $\mathrm{P}(|\mathrm{Y} - m| \le \sigma t) = 2\Phi_0(t)$.

При достаточно больших n имеем приближенно:

$$\mathrm{X} \underset{n\to +\infty}{\sim} N(np;\sqrt{npq}) \Leftrightarrow \frac{\mathrm{X}}{n} \underset{n\to +\infty}{\sim} N\left(p;\frac{\sqrt{pq}}{\sqrt{n}}\right) \Rightarrow$$

$$\Rightarrow \mathrm{P}\left(\left|\frac{\mathrm{X}}{n} - p\right| \le t\frac{\sqrt{pq}}{\sqrt{n}}\right) \approx 2\,\Phi_0(t).$$

Положим $t\frac{\sqrt{pq}}{\sqrt{n}}$ = ε, откуда $t = \frac{\varepsilon\sqrt{n}}{\sqrt{pq}}$; в итоге получаем:

$$P(|\frac{X}{n} - p| \le \varepsilon) \approx 2\,\Phi_o(\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}).$$

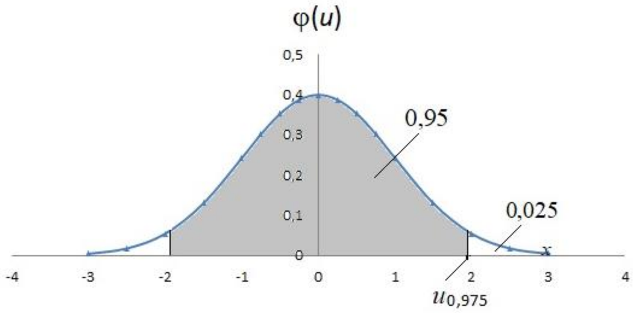
Задача

Найти наименьшее значение числа испытаний n в схеме Бернулли, при котором выполняется условие:

$$P(|\frac{X}{n} - p| \le \varepsilon) = 0{,}95 \text{ при } \varepsilon = 10^{-2} \text{ и } p = 10^{-3}.$$

Указание: воспользоваться приближением $P(|\frac{X}{n} - p| \le \varepsilon) \approx 2\,\Phi_o(\frac{\varepsilon\sqrt{n}}{\sqrt{pq}})$.

Решение:



Из $P(|\frac{X}{n} - p| \le \varepsilon) \approx 2\,\Phi_o(\frac{\varepsilon\sqrt{n}}{\sqrt{pq}})$ и $P(|\frac{X}{n} - p| \le \varepsilon) = 0{,}95$ следует:

$$2\,\Phi_o(\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}) = 0{,}95 \Rightarrow [2\,\Phi_o(u) = 0{,}95 \Rightarrow u = u_{0{,}975} = 1{,}96\,] \Rightarrow$$

$$\Rightarrow \frac{\varepsilon\sqrt{n}}{\sqrt{pq}} = 1{,}96 \Rightarrow \sqrt{n} = \frac{\sqrt{p(1-p)}}{\varepsilon} \cdot 1{,}96.$$

При $\varepsilon = 10^{-2}$, $p = 10^{-3}$ имеем: $\sqrt{n} = \frac{\sqrt{10^{-3}(1-10^{-3})}}{10^{-2}}\,1{,}96 = 6{,}194964 \Rightarrow$

$\Rightarrow n = 39$ (наименьшее значение n).