Предельные теоремы ТВ

Рассмотрим ряд утверждений и теорем, устанавливающих связь между теоретическими и экспериментальными характеристиками СВ при большом числе испытаний.

Предельные теоремы условно разделим на 2 группы:

1. Закон больших чисел (ЗБЧ) – устанавливает устойчивость средних значений: при большом числе испытаний средний результат может быть предсказан с достаточной точностью.

2. Центральная предельная теорема (ЦПТ) – устанавливает условия, при которых закон распределения суммы большого числа СВ неограниченно приближается к нормальному.

Рассмотрим неравенство Чебышева. Используется для:

1) грубой оценки вероятностей событий, связанных со СВ, распределение которых не известны;
2) док-ва теорем ЗБЧ.

Неравенство Чебышева

Теорема1. Если СВ X имеет м.о. $MX=a$ и дисперсию DX, то для любого $\varepsilon>0$ справедливо нер-во Чебышева $$P\{|X - MX| \geqslant \varepsilon\} \leqslant \frac{DX}{\varepsilon^2}$$

Д-во.

Докажем неравенство для непрерывной с. в. X с плотностью $f(x)$. Вероятность $P\{|X-a| \geqslant \varepsilon\}$ есть вероятность попадания с. в. X в область, лежащую вне промежутка $[a-\varepsilon, a+\varepsilon]$.

$$P\{|X-a| \geqslant \varepsilon\} = \int_{-\infty}^{a-\varepsilon} f(x)\,dx + \int_{a+\varepsilon}^{+\infty} f(x)\,dx = \int_{|x-a|\geqslant\varepsilon} f(x)\,dx = \int_{|x-a|\geqslant\varepsilon} 1\cdot f(x)\,dx \leqslant \int_{|x-a|\geqslant\varepsilon} \frac{(x-a)^2}{\varepsilon^2} f(x)\,dx,$$

так как область интегрирования $|x-a| \geqslant \varepsilon$ можно записать в виде $(x-a)^2 \geqslant \varepsilon^2$, откуда следует $1 \leqslant \dfrac{(x-a)^2}{\varepsilon^2}$. Имеем

$$P\{|X-a| \geqslant \varepsilon\} \leqslant \frac{1}{\varepsilon^2} \int_{|x-a|\geqslant\varepsilon} (x-a)^2 f(x)\,dx \leqslant \frac{1}{\varepsilon^2} \int_{-\infty}^{\infty} (x-a)^2 f(x)\,dx,$$

$$P\{|X-a| \geqslant \varepsilon\} \leqslant \frac{1}{\varepsilon^2} \int_{-\infty}^{\infty} (x-a)^2 f(x)\,dx = \frac{1}{\varepsilon^2} DX,$$ т. е. $$P\{|X-a| \geqslant \varepsilon\} \leqslant \frac{DX}{\varepsilon^2}.$$

Аналогично док-ся и для ДСВ. Его можно записать и в виде

$$P\{|X - MX| < \varepsilon\} \geqslant 1 - \frac{DX}{\varepsilon^2}.$$

Нер-во Чебышева справедливо для любых СВ. Например, для СВ $X=m$, имеющей *биномиальное распределение* с м.о. $MX=a=np$ и дисперсией $DX=npq$

$$P\{|m-np|<\varepsilon\} \geqslant 1-\frac{npq}{\varepsilon^2}$$

для *частости* $\frac{m}{n}$ события в n независимых испытаниях, в каждом из которых оно может произойти с вероятностью $p = M\left(\frac{m}{n}\right) = a$, дисперсия которых $D\left(\frac{m}{n}\right) = \frac{pq}{n}$, неравенство

$$P\left\{\left|\frac{m}{n}-p\right|<\varepsilon\right\} \geqslant 1-\frac{pq}{n\varepsilon^2}.$$

Оценку вероятности попадания СВ X в $[\varepsilon, \infty)$ дает нер-во Маркова.

Теорема2. Для любой неотрицательной СВ X, имеющей м.о. MX и $\varepsilon>0$, справедливо нер-во

$$P\{X \geqslant \varepsilon\} \leqslant \frac{MX}{\varepsilon}$$

Д-во.

$$P\{X \geqslant \varepsilon\} = \int_{\varepsilon}^{\infty} f(x)\,dx \leqslant \int_{\varepsilon}^{\infty} \frac{x}{\varepsilon} f(x)\,dx = \frac{1}{\varepsilon}\int_{\varepsilon}^{\infty} x f(x)\,dx \leqslant \frac{1}{\varepsilon}\int_{0}^{\infty} x f(x)\,dx = \frac{MX}{\varepsilon}.$$

Нер-во также можно записать и в виде $P\{X < \varepsilon\} \geqslant 1-\frac{MX}{\varepsilon}.$

Пример Оценить с помощью неравенства Чебышева вероятность того, что отклонение с. в. X от своего м. о. будет меньше трех с. к. о., т. е. меньше $3\sigma_x$.

Полагая $\varepsilon = 3\sigma_x$ в формуле $P\{|X - MX| < \varepsilon\} \geqslant 1 - \frac{DX}{\varepsilon^2}$

$$P\{|X - MX| < 3\sigma_x\} \geqslant 1 - \frac{\sigma_x^2}{(3\sigma_x)^2} = 1 - \frac{1}{9} = \frac{8}{9} \approx 0{,}8889.$$

Теорема Чебышева

Случайные величины $X_1, X_2, \ldots, X_n, \ldots$ *сходятся по вероятности* к величине A (случайной или неслучайной), если для любого $\varepsilon > 0$ вероятность события $\{|X_n - A| < \varepsilon\}$ при $n \to \infty$ стремится к единице, т. е.

$$\lim_{n\to\infty} P\{|X_n - A| < \varepsilon\} = 1 \qquad X_n \xrightarrow[n\to\infty]{P} A.$$

Теорема 3 (ЗБЧ в форме П. Л. Чебышева, 1886 г.). Если случайные величины $X_1, X_2, \ldots, X_n, \ldots$ независимы и существует такое число $C > 0$, что $DX_i \leqslant C$, $i = 1, 2, \ldots$, то для любого $\varepsilon > 0$

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} MX_i\right| < \varepsilon\right\} = 1, \qquad \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{P} \frac{1}{n}\sum_{i=1}^{n} MX_i.$$

Д-во.

Так как $DX_i \leqslant C$, $i = 1, 2, \ldots$, то $D\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} D\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} DX_i =$

$$= \frac{1}{n^2}(DX_1 + DX_2 + \ldots + DX_n) \leqslant \frac{1}{n^2}(C + C + \ldots + C) = \frac{1}{n^2}Cn = \frac{C}{n}.$$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} MX_i\right| < \varepsilon\right\} \geqslant 1 - \frac{D\left(\frac{1}{n}\sum\limits_{i=1}^{n} X_i\right)}{\varepsilon^2} \geqslant 1 - \frac{C}{n\varepsilon^2}.$$

Переходя к пределу при $n \to \infty$ и учитывая, что вероятность любого события не превышает 1, получаем

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} MX_i\right| < \varepsilon\right\} = 1.$$

Следствие. Если с. в. $X_1, X_2, \ldots, X_n, \ldots$ независимы и одинаково распределены, $MX_i = a$, $DX_i = \sigma^2$, то для любого $\varepsilon > 0$

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - a\right| < \varepsilon\right\} = 1 \qquad \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{P} a.$$

Д-во.

$$\frac{1}{n}\sum_{i=1}^{n} MX_i = \frac{1}{n}(MX_1 + MX_2 + \ldots + MX_n) = \frac{1}{n}(a + a + \ldots + a) = \frac{1}{n}\cdot na = a,$$

а дисперсии с. в. X_i равны числу σ^2, т. е. ограничены, то, применив ЗБЧ

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} MX_i\right| < \varepsilon\right\} = 1,$$

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - a\right| < \varepsilon\right\} = 1$$

Пример Глубина моря измеряется прибором, не имеющим систематической ошибки. Среднее квадратическое отклонение измерений не превосходит 15 м. Сколько нужно сделать независимых измерений, чтобы с вероятностью, не меньшей 0,9, можно было утверждать, что среднее арифметическое этих измерений отличается от a (глубины моря) по модулю меньше, чем на 5 м?

Обозначим через X_i результаты n независимых измерений глубины моря. Нужно найти число n, которое удовлетворяет неравенству

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} MX_i\right| < \varepsilon\right\} \geqslant 1 - \frac{C}{n\varepsilon^2},$$

где $MX_i = a$, *что означает отсутствие при измерениях систематической ошибки* (т. е. измерения производятся с одинаковой точностью). По условию $\varepsilon = 5$, $C = 225$, так как $\sigma = \sqrt{DX} = 15$м. Отсюда

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} -a\right| < 5\right\} \geqslant 1 - \frac{225}{25n} \geqslant 0{,}9,$$

$$0{,}1 \geqslant \frac{9}{n},\ n \geqslant 90.$$

Измерение нужно проводить не менее 90 раз.

Теорема Бернулли

Теорема 4 (ЗБЧ в форме Я. Бернулли, 1713 г.). Если вероятность появления события A в одном испытании равна p, число наступления этого события при n независимых испытаниях равно n_A, то для любого числа $\varepsilon > 0$ имеет место равенство

$$\lim_{n\to\infty} P\left\{\left|\frac{n_A}{n} - p\right| < \varepsilon\right\} = 1,$$

т. е. относительная частота $P^*(A)$ события A сходится по вероятности к вероятности p события A: $P^*(A) \xrightarrow[n\to\infty]{P} P(A)$.

Д-во.

Введем с. в. $X_1, X_2, \ldots, X_n$ следующим образом: $X_i = 1$, если в i-м испытании появилось событие A, а если не появилось, то $X_i = 0$. Тогда

$n_A = \sum_{i=1}^{n} X_i.$

X_i	0	1
P	$1-p$	p,

$MX_i = p \quad DX_i = pq.$

с. в. X_i независимы, их дисперсии ограничены одним и тем же числом

$$p(1-p) = p - p^2 = \frac{1}{4} - \left(p - \frac{1}{2}\right)^2 \leqslant \frac{1}{4}$$

Поэтому к этим с. в. можно применить теорему Чебышева

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} MX_i\right| < \varepsilon\right\} = 1. \quad \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{n_A}{n}, \quad \frac{1}{n}\sum_{i=1}^{n} MX_i = \frac{1}{n}\cdot np = p.$$

Следовательно, $\lim_{n\to\infty} P\left\{\left|\frac{n_A}{n} - p\right| < \varepsilon\right\} = 1.$

Неравенство Чебышева для случайных величин $n_A = \frac{1}{n}\sum_{i=1}^{n} X_i$ принимает вид

$$P\left\{\left|\frac{n_A}{n} - p\right| < \varepsilon\right\} \geqslant 1 - \frac{pq}{n\varepsilon^2}.$$

Обобщением теоремы Бернулли на случай, когда вероятности p_i появления события A в каждом из n испытаний различны, является *теорема Пуассона*:

$$P^*(A) \xrightarrow[n\to\infty]{P} \frac{1}{n}\sum_{i=1}^{n} p_i,$$

где p_i — вероятность события A в i-м испытании.

Пример Вероятность наличия опечатки на одной странице рукописи равна 0,2. Оценить вероятность того, что в рукописи, содержащей 400 страниц, частость появления опечатки отличается от соответствующей вероятности по модулю меньше, чем 0,05.

$$p = 0{,}2,\ q = 0{,}8,\ \ n = 400,\ \varepsilon = 0{,}05.$$

$$P\left\{\left|\frac{n_A}{n} - 0{,}2\right| < 0{,}05\right\} \geqslant 1 - \frac{0{,}2 \cdot 0{,}8}{400 \cdot 0{,}05^2} = 0{,}84,$$

т. е. $p \geqslant 0{,}84$.

Центральная предельная теорема

Теорема 5. Пусть с. в. $X_1, X_2, \ldots, X_n$ независимы, одинаково распределены, имеют конечные математическое ожидание $MX_i = a$ и дисперсию $DX_i = \sigma^2$, $i = \overline{1,n}$. Тогда функция распределения центрированной и нормированной суммы этих случайных величин стремится при $n \to \infty$ к функции распределения стандартной нормальной случайной величины:

$$Z_n = \frac{\sum\limits_{i=1}^{n} X_i - M\left(\sum\limits_{i=1}^{n} X_i\right)}{\sqrt{D\left(\sum\limits_{i=1}^{n} X_i\right)}} = \frac{\sum\limits_{i=1}^{n} X_i - na}{\sigma\sqrt{n}},$$

$$F_{Z_n}(x) = P\{Z_n < x\} \xrightarrow[n\to\infty]{} \Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt.$$

при достаточно большом n сумма Z_n приближенно распределена по нормальному закону: $Z_n \sim N(0,1)$. Это означает, что сумма $S_n = X_1 + X_2 + \ldots + X_n$ приближенно распределена по нормальному закону: $S_n \sim N(na, \sqrt{n}\sigma)$. Говорят, что при $n \to \infty$ с. в. $\sum\limits_{i=1}^{n} X_i$ *асимптотически нормальна.*

Пример Независимые с. в. X_i распределены равномерно на отрезке $[0,1]$. Найти закон распределения с. в.

$$Y = \sum_{i=1}^{100} X_i,$$

а также вероятность того, что $55 < Y < 70$.

Условия ЦПТ соблюдаются, поэтому с. в. Y имеет приближенно плотность распределения

$$f_Y(y) \approx \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{(y-m_y)^2}{2\sigma_y^2}}.$$

По известным формулам для м. о. и дисперсии в случае равномерного распределения находим: $MX_i = \frac{0+1}{2} = \frac{1}{2}$, $DX_i = \frac{(1-0)^2}{12} = \frac{1}{12}$, $\sigma X_i = \frac{1}{\sqrt{12}} = \frac{1}{2\sqrt{3}}$. Тогда

$$m_y = M\left(\sum_{i=1}^{100} X_i\right) = \sum_{i=1}^{100} MX_i = 100 \cdot \frac{1}{2} = 50,$$

$$\sigma_y^2 = D\left(\sum_{i=1}^{100} X_i\right) = \sum_{i=1}^{100} DX_i = 100 \cdot \frac{1}{12} = \frac{25}{3},$$

$$f_Y(y) \approx \frac{3}{5\sqrt{6\pi}} e^{-\frac{3(y-50)^2}{50}}.$$

$$P\{55 < Y < 70\} = \Phi\left(\frac{70-50}{\frac{5\sqrt{3}}{3}}\right) - \Phi\left(\frac{55-50}{\frac{5\sqrt{3}}{3}}\right) = \Phi(6{,}9282) - \Phi(1{,}73) \approx 0{,}04$$