Лекция 10

Неравенство Чебышёва. Закон больших чисел. Центральная предельная теорема.

Пусть Y — СВ такая, что $\exists\, M\{|Y|^k\}$, где $k>0$. Тогда $\forall \varepsilon>0$ имеет место неравенство Чебышёва:

$$P\{|Y|\geq\varepsilon\}\leq\frac{M\{|Y|^k\}}{\varepsilon^k}. \qquad (1)$$

Докажем (1). Пусть для определенности Y — непрерывная СВ с плотностью $f(y)$. Тогда

$$M\{|Y|^k\}=\int_{-\infty}^{\infty}|y|^k f(y)dy\geq\int_{-\infty}^{-\varepsilon}|y|^k f(y)dy+\int_{\varepsilon}^{+\infty}|y|^k f(y)dy\geq$$

$$\geq\int_{-\infty}^{-\varepsilon}\varepsilon^k f(y)dy+\int_{\varepsilon}^{+\infty}\varepsilon^k f(y)dy=$$

$$=\varepsilon^k\left(\int_{-\infty}^{-\varepsilon}f(y)dy+\int_{\varepsilon}^{+\infty}f(y)dy\right)=\varepsilon^k P\{|Y|\geq\varepsilon\},$$

откуда и следует неравенство (1).

Пусть X — некоторая СВ, $m_x=M\{X\}$. Положим $Y=X-m_x$, $k=2$ и запишем (1):

$$P\{|X-m_x|\geq\varepsilon\}\leq\frac{M\{(X-m_x)^2\}}{\varepsilon^2}=\frac{DX}{\varepsilon^2}$$

или

$$P\{|X-m_x|<\varepsilon\}\geq 1-\frac{DX}{\varepsilon^2},$$

где DX — дисперсия СВ X. Пусть $\varepsilon=3\sigma_x$, где $\sigma_x=\sqrt{DX}$. Тогда получим

$$P\{|X-m_x|<3\sigma_x\}\geq 1-\frac{DX}{(3\sigma_x)^2},$$

или

$$P\{m_x-3\sigma_x<X<m_x+3\sigma_x\}\geq\frac{8}{9}.$$

Последнее неравенство называется „правилом трёх сигм":

вероятность того, что значение случайной величины с математическим ожиданием m_x и среднеквадратичным отклонением σ_x отличается от m_x менее, чем на $3\sigma_x$, не меньше $8/9$.

<u>Теорема 1</u> (теорема Чебышёва) Пусть имеется бесконечная последовательность X_1, $X_2, \dots$ независимых СВ таких, что

$$MX_1 = MX_2 = \dots = m, \quad DX_1 \le C, \quad DX_2 \le C, \dots,$$

где $C>0$ и m – постоянные. Тогда $\forall \varepsilon > 0$

$$\lim_{n\to\infty} P\left\{\left|\frac{X_1+X_2+\dots+X_n}{n} - m\right| < \varepsilon\right\} = 1.$$

Доказательство. Пусть $S_n = \dfrac{X_1+X_2+\dots+X_n}{n}$.
Напишем неравенство (1) для СВ $Y \equiv S_n - M\{S_n\}$ при $k=2$. Получим

$$P\{|S_n - M\{S_n\}| \ge \varepsilon\} \le \frac{D\{S_n\}}{\varepsilon^2},$$

или

$$P\{|S_n - M\{S_n\}| < \varepsilon\} \ge 1 - \frac{D\{S_n\}}{\varepsilon^2}. \quad (2)$$

Но

$$M\{S_n\} = m, \quad D\{S_n\} = \frac{1}{n^2}\sum_{i=1}^{n} D\{X_i\} \le \frac{nC}{n^2} = \frac{C}{n},$$

так что из (2) получаем

$$P\left\{\left|\frac{X_1+X_2+\dots+X_n}{n} - m\right| < \varepsilon\right\} \ge 1 - \frac{C}{n\varepsilon^2}.$$

Переходя в этом неравенстве к пределу при $n\to\infty$, получаем требуемое.

Физический смысл теоремы:

Пусть неизвестная величина m многократно измеряется некоторым прибором. Пусть

$$x_i = m + \varepsilon_i$$

— показание прибора при i-м измерении, ε_i — ошибка прибора. Значение x_i можно рассматривать как реализацию случайной величины

$$X_i = m + \mathcal{E}, \qquad (3)$$

где $\mathcal{E}$ — некоторая случайная величина с нулевым математическим ожиданием $M\mathcal{E} = 0$, если прибор не имеет систематической ошибки. С серией из n измерений свяжем случайные величины $X_1, X_2, \ldots, X_n$, так что реализация СВ X_i даёт показание прибора x_i при i-м измерении. Все случайные величины $X_1, X_2, \ldots, X_n$ определяются формулой (3), имеют один и тот же закон распределения и одно и то же математическое ожидание m. Значит, для СВ $X_1, X_2, \ldots, X_n$ выполняется теорема 1. Следовательно, среднее арифметическое

$$\frac{x_1 + x_2 + \ldots + x_n}{n}$$

серии из n измерений как реализация случайной величины $\frac{1}{n}(X_1 + X_2 + \ldots + X_n)$ будет отличаться от неизвестного значения m не более, чем на ε, с вероятностью, сколь угодно близкой к 1 при достаточно большом значении n.

<u>Теорема 2</u> (теорема Бернулли) Пусть k — число успехов в серии из n независимых испытаний с вероятностью успеха p в одном испытании. Тогда $\forall \varepsilon > 0$

$$\lim_{n\to\infty} P\left\{\left|\frac{k}{n} - p\right| < \varepsilon\right\} = 1. \qquad (4)$$

Доказательство. Пусть СВ X_i — число успехов в i-м испытании, т.е. закон распределения СВ X_i, $i=1,2,\dots,n$, имеет вид

X_i	1	0
P	p	$1-p$

Для последовательности СВ $X_1, X_2, \dots$ выполнены условия теоремы 1, ~~так что~~ причем $MX_1 = MX_2 = \dots = p$, так что

$$\lim_{n\to\infty} P\left\{\left|\frac{X_1+X_2+\dots+X_n}{n} - p\right| < \varepsilon\right\} = 1,$$

что совпадает с (4), ибо $X_1+X_2+\dots+X_n = k$.

Комментарий. Теорему Бернулли можно проинтерпретировать как факт, подтверждающий непротиворечивость двух подходов к определению вероятности: аксиоматического Колмогоровского подхода и экспериментального статистического подхода, согласно которому вероятность $P\{A\}$ события A определяется как предел $\lim\limits_{n\to\infty}(k(n)/n)$, где n — общее число испытаний, $k(n)$ — число испытаний, приведших к событию A.

Теорема 3 (центральная предельная теорема в форме Ляпунова)

Пусть последовательность $X_1, X_2, \ldots$ независимых СВ удовлетворяет условию Ляпунова:

$$\lim_{n\to\infty} \frac{\sum\limits_{i=1}^{n} h_i}{\left(\sum\limits_{i=1}^{n} d_i\right)^{3/2}} = 0, \qquad (5)$$

где $d_i = D\{X_i\}$, $h_i = M\{|X_i - MX_i|^3\}$. Тогда

$$\lim_{n\to\infty} P\{\alpha \le \tilde{X}_n \le \beta\} = \int\limits_{\alpha}^{\beta} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx,$$

где

$$\tilde{X}_n = \frac{\hat{X}_n - M\{\hat{X}_n\}}{\sqrt{D\{\hat{X}_n\}}}, \quad \hat{X}_n = X_1 + X_2 + \ldots + X_n.$$

Замечание. ~~Если~~ Условие (5) выполняется, если все СВ $X_1, X_2, \ldots$ имеют один и тот же закон распределения.

Действительно, в этом случае $d_i = d \ \forall i = 1, 2, \ldots$ и $h_i = h \ \forall i = 1, 2, \ldots$. Поэтому

$$\lim_{n\to\infty} \frac{\sum\limits_{i=1}^{n} h_i}{\left(\sum\limits_{i=1}^{n} d_i\right)^{3/2}} = \lim_{n\to\infty} \frac{h \cdot n}{(d \cdot n)^{3/2}} = 0.$$

В качестве примера использования теоремы 3 покажем, как из нее м.б. получена интегральная теорема Муавра-Лапласа.

Пусть k – число успехов в серии из n независимых испытаний с вероятностью успеха p в одном испытании.

Тогда

$$k = X_1 + X_2 + \dots + X_n,$$

где $X_1, X_2, \dots, X_n$ – независимые СВ, имеющие одинаковый з-н распределения

X_i	1	0
P	p	1-p

Поскольку $M\{k\} = np$, $D\{k\} = npq$, где $q = 1-p$, то по теореме 3

$$\lim_{n\to\infty} P\left\{\alpha \le \frac{k-np}{\sqrt{npq}} \le \beta\right\} = \int_\alpha^\beta \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \qquad (6')$$

$$= \Phi(\beta) - \Phi(\alpha),$$

где $\Phi(z) = \int_0^z \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$ — функция Лапласа

Равенство (6') можно переписать в виде (разрешив двойное неравенство под знаком вероятности относительно k):

$$\lim_{n\to\infty} P\left\{np + \alpha\sqrt{npq} \le k \le np + \beta\sqrt{npq}\right\} = \Phi(\beta) - \Phi(\alpha),$$

а это и есть утверждение интегральной теоремы Муавра-Лапласа.