ВВЕДЕНИЕ

Математической статистикой называется наука, занимающаяся методами обработки экспериментальных данных (ЭД), полученных в результате наблюдений над случайными явлениями.

Перед любой наукой ставятся следующие задачи:

- Описание явлений;
- Анализ и прогноз;
- Выборка оптимальных решений.

Применительно к математической статистике пример задачи первого типа: пусть имеется статистический материал, представляющий собой случайные числа. Требуется его упростить, представить в виде таблиц и графиков, обеспечивающих наглядность и информативность представленного материала.

Пример задачи второго типа: оценка (хотя бы приблизительная) характеристик случайных величин, например, математического ожидания, дисперсии и т.д. Какова точность полученных оценок.

Одной из характерных задач третьего типа является задача проверки правдоподобия гипотез, которая формулируется следующим образом: можно ли предполагать, что имеющаяся совокупность случайных чисел не противоречит некоторой гипотезе (например о виде распределения, наличия корреляционной зависимости и т.д.).

В курсе рассматриваются задачи всех трех типов: способы описания результатов опыта, способы обработки опытных данных и оценки по ним характеристик случайного явления, способы выбора разумных решений.

1. ЗАКОН БОЛЬШИХ ЧИСЕЛ

Пусть проводится опыт Е, в котором нас интересует признак X, или СВ X. При однократном проведении Е нельзя заранее сказать, какое значение примет X. Но при n-кратном повторении «среднее» значение величины X (среднее арифметическое) теряет случайный характер и становится близким к некоторой константе.

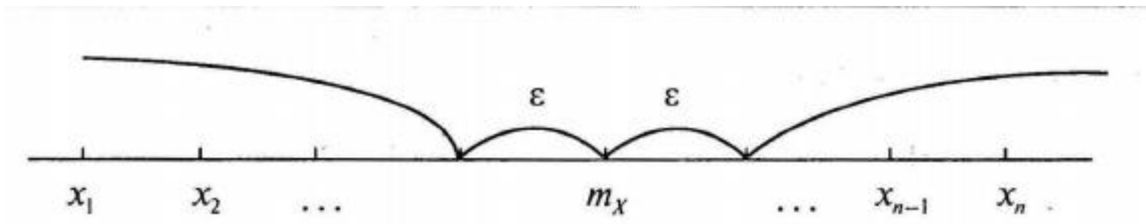
Закон больших чисел – совокупность теорем, определяющих условия стремления средних арифметических значений случайных величин к некоторой константе, при увеличении числа опытов до бесконечности ($n \to \infty$).

1.2. НЕРАВЕНСТВА ЧЕБЫШЕВА

<u>Теорема</u>. *Для любой случайной величины X с m_x, D_x выполняется следующее неравенство* $P(|X - m_x| \geq \varepsilon) \leq \dfrac{D_x}{\varepsilon^2}$ *где $\varepsilon>0$.*

<u>Доказательство:</u>

1. Пусть величина X – ДСВ. Изобразим значения X и M_x в виде точек на числовой оси Ox



Вычислим вероятность того, что при некотором $\varepsilon \geq 0$ величина X отклонится от своего математического ожидания не меньше чем на ε:

$$P(|X - m_x| \geq \varepsilon).$$

Это событие заключается в том, что точка X не попадет на отрезок $[m_x-\varepsilon, m_x+\varepsilon]$, т.е.

$$P(|X - m_x| \geq \varepsilon) = P(X \not\subset [m_x - \varepsilon, m_x + \varepsilon) = \sum_{|x_i - m_x| \geq \varepsilon} p_i \text{ --}$$

для тех значений x, которые лежат вне отрезка $[m_x-\varepsilon, m_x+\varepsilon]$.

Рассмотрим дисперсию с.в. X:

$$D_x = M[(X - m_x)^2] = \sum_{i=1}^{n} |x_i - m_x|^2 p_i.$$

Т.к. все слагаемые – положительные числа, то если убрать слагаемые, соответствующую отрезку $[m_x-\varepsilon, m_x+\varepsilon]$, то можно записать:

$$D_x \geq \sum_{|x_i - m_x| \geq \varepsilon}^{n} |x_i - m_x|^2 p_i,$$

т.к. $|X - m_x| \geq \varepsilon$, то неравенство можно усилить

$$D_x \geq \sum_{|x_i - m_x| \geq \varepsilon}^{n} \varepsilon^2 p_i = \varepsilon^2 \sum_{|x_i - m_x| \geq \varepsilon}^{n} p_i$$

$$\Rightarrow D_x \geq \varepsilon^2 P(|X - m_x| \geq \varepsilon) \Rightarrow$$

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

2. Для НСВ:

$$P(|X - m_x| \geq \varepsilon) = \int_{|x_i - m_x| \geq \varepsilon} f(x)dx$$

$$D_x = \int_{-\infty}^{+\infty} (x - m_x)^2 f(x)dx =$$

$$\int_{-\infty}^{+\infty} |x - m_x|^2 f(x)dx \geq$$

$$\int_{|x - m_x| \geq \varepsilon} |x - m_x|^2 f(x)dx$$

- это интегрирование по внешней части отрезка [m_x-ε, m_x+ε].

Применяя неравенство $|x - m_x| \geq \varepsilon$ и подставляя его под знак интеграла, получаем

$$D_x \geq \int_{|x - m_x| \geq \varepsilon} |x - m_x|^2 f(x)dx \geq \varepsilon^2 \int_{|x - m_x| \geq \varepsilon} f(x)dx = \varepsilon^2 P(|x_i - m_x| \geq \varepsilon).$$

Откуда и вытекает неравенство Чебышева для НСВ.

<u>Следствие.</u> $P(|X - m_x| \leq \varepsilon) \leq 1 - \frac{D_x}{\varepsilon^2}$ - это 2-е неравенство Чебышева.

<u>Доказательство:</u> События $|X - m_x| \leq \varepsilon$ и $|X - m_x| \geq \varepsilon$ - противоположны $\Rightarrow P(|X - m_x| \leq \varepsilon) + P(|X - m_x| \geq \varepsilon) = 1$.

1. ***<u>Лемма:</u>*** Пусть X –СВ, ε>0 – любое число. Тогда

$$P(|X| \geq \varepsilon) < \frac{M(X^2)}{\varepsilon^2}$$

<u>Доказательство:</u>

$$P(|X| \geq \varepsilon) = \int_{|x| \geq \varepsilon} f(x)dx = \int_{|x| \geq \varepsilon} \frac{x^2}{\varepsilon^2} \cdot \frac{\varepsilon^2}{x^2} f(x)dx \leq \int_{|x| \geq \varepsilon} \frac{x^2}{\varepsilon^2} f(x)dx$$

$$= \frac{1}{\varepsilon^2} \int_{|x| \geq \varepsilon} x^2 f(x)dx \leq$$

$$\leq \frac{1}{\varepsilon^2} \int_{-\infty}^{+\infty} x^2 f(x)dx = \frac{1}{\varepsilon^2} M(X^2),$$

Т.к.$\frac{\varepsilon^2}{x^2} \leq 1$.

<u>Следствие: (правило трех сигм для произвольного распределения):</u>

Полагаем в неравенстве Чебышева $\alpha = 3\sigma_x$, имеем

$$P(|X - m_x| \geq 3\sigma_x) \leq \frac{D_x}{9\sigma_x^2} = \frac{1}{9}.$$

Т.е. вероятность того, что отклонение св от ее МО выйдет за пределы трех СКО, не больше 1/9.

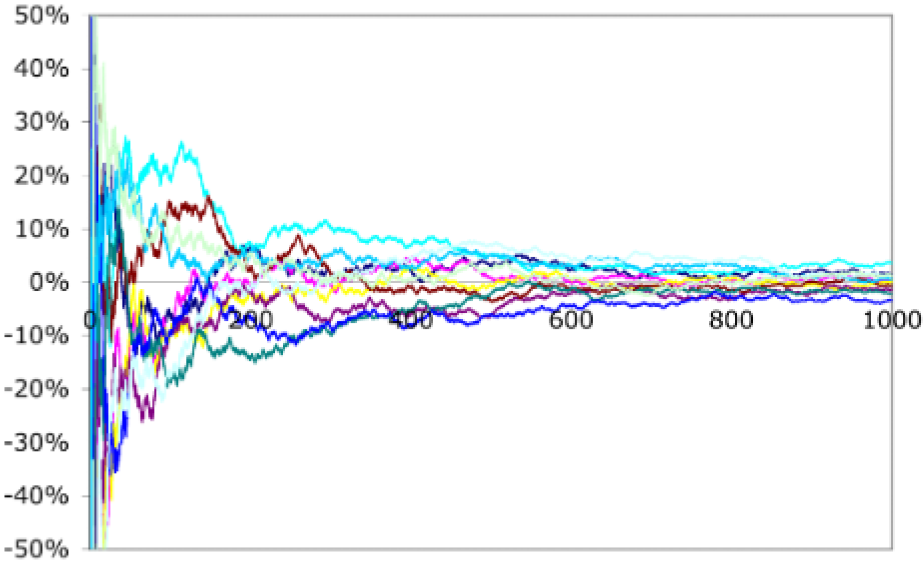
Неравенство Чебышева дает только верхнюю границ вероятности данного отклонения. Выше этой границы - значение не может быть **ни при никаком** распределении.

1.3. СХОДИМОСТЬ ПО ВЕРОЯТНОСТИ

Последовательность случайных величин X_n ***сходится по вероятности к величине a***, $X_n \xrightarrow[n\to\infty]{} a$ если при увеличении n вероятность того, что X_n и а будут сколь угодно близки, неограниченно приближается к единице:

$$p(|X_n - a| < \varepsilon) > 1 - \delta$$

где ε, δ - произвольно малые положительные числа.



1.4.ТЕОРЕМЫ ЧЕБЫШЕВА

Одна из важнейших, но наиболее важных форм закона больших чисел – **<u>теорема</u>** Чебышева – она устанавливает связь между средним арифметическим наблюденных значений СВ и ее МО.

1.4.1.Первая теорема Чебышева.

<u>Теорема.</u> *При достаточно большом числе независимых опытов среднее арифметическое значений случайной величины сходится по вероятности к ее математическому ожиданию:*

$$\frac{\sum_{i=1}^{n} X_i}{n} \xrightarrow[n\to\infty]{} m_x$$

***Доказательство*:** Рассмотрим величину Y равную

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

Определим числовые характеристики Y_n m_y и D_Y.

$$m_y = M[Y_n] = \frac{1}{n}\sum_{i=1}^{n} M[X_i] = \frac{1}{n} n m_x = m_x, \qquad D_y = \frac{1}{n^2}\sum_{i=1}^{n} D[X_i] = \frac{D_x}{n}$$

Запишем неравенство Чебышева для величины Y_n

$$p\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - m_x\right| \geq \varepsilon\right) < \frac{D_y}{\varepsilon^2} = \frac{D_x}{\varepsilon^2 n}$$

Как бы ни было мало число ε, можно взять n таким большим, чтобы выполнялось неравенство $\frac{D_x}{\varepsilon^2 n} \leq \delta$, где δ - сколь угодно малое число. Тогда

$$p\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - m_x\right| \geq \varepsilon\right) < \delta.$$

Переходя к противоположному событию:

$$p\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - m_x\right| \leq \varepsilon\right) < 1 - \delta$$

Т.е. вероятность может быть сколь угодно близкой к 1.

1.4.2. Вторая теорема Чебышева:

Теорема. Если $X_1.....X_n$ – последовательность попарно независимых СВ с МО $m_{x1}....m_{xn}$ и дисперсиями $D_{x1}..D_{xn}$ ограничены одним и тем же числом $D_{xi}<L$ $(i=1..n)$, $L=const$, тогда для любого $\varepsilon, \delta>0$ – бесконечно малых

$$P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - \frac{\sum_{i=1}^{n} m_{X_i}}{n}\right| < \varepsilon\right) > 1 - \delta \text{ или } \lim_{n\to\infty} P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - \frac{\sum_{i=1}^{n} m_{X_i}}{n}\right| < \varepsilon\right) = 1$$

***Доказательство*:** Рассмотрим СВ

$$Y = \frac{\sum_{i=1}^{n} X_i}{n}, \quad m_y = \frac{\sum_{i=1}^{n} m_{X_i}}{n}, \quad D_y = \frac{\sum_{i=1}^{n} D_{X_i}}{n^2}.$$

Применим к Y неравенство Чебышева:

$$P(|Y - m_Y| \leq \varepsilon) \geq 1 - \frac{D_x}{\varepsilon^2} \text{ или } P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - \frac{\sum_{i=1}^{n} m_{X_i}}{n}\right| \leq \varepsilon\right) \geq 1 - \frac{\sum_{i=1}^{n} D_{X_i}}{n^2 \varepsilon^2}$$

Заменим:

$$P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n}-\frac{\sum_{i=1}^{n} m_{X_i}}{n}\right|\leq\varepsilon\right)\geq 1-\frac{L}{n^2\varepsilon^2}$$

Как бы ни было мало число ε, можно взять n таким большим, чтобы выполнялось неравенство $\frac{D_x}{\varepsilon^2 n}\leq\delta$, где δ - сколь угодно малое.

Т.е., взяв предел при $n\to\infty$ от обеих частей и $\lim\limits_{n\to\infty}\frac{L}{n\varepsilon^2}=0$ получаем:

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

(так как вероятность не может быть больше 1).

<u>Пример1.1.</u> Производится большое число n независимых опытов, в каждом из которых некоторая случайная величина имеет равномерное распределение на участке [1,2]. Рассматривается среднее арифметическое наблюденных значений случайной величины *X*. На основании Закона больших чисел выяснить, к какому числу *a* будет приближаться величина *Y* при $n\to\infty$. Оценить максимальную практически возможную ошибку равенства $Y\approx a$.

<u>Решение.</u> $a=M[Y]=M\left[\frac{1}{n}\sum_{i-1}^{n}X_i\right]=\frac{1}{n}M\left[\sum_{i=1}^{n}X_i\right]=\frac{1}{n}\sum_{i=1}^{m}M[X_i]=\frac{1}{n}\cdot n\cdot M[X]=M[X]=1{,}5.$

$D[Y]=D\left[\frac{1}{n}\sum_{i=1}^{n}X_i\right]=\frac{1}{n^2}D\left[\sum_{i=1}^{n}X_i\right]=\frac{1}{n^2}\sum_{i-1}^{n}D[X_i]=\frac{1}{n^2}\cdot n\cdot D[X]=\frac{1}{n}\cdot\frac{1}{12}=\frac{1}{12n}.$

$$\sigma_y=\frac{1}{2\sqrt{3n}}.$$

Максимальное практически возможное значение ошибки

$$3\sigma_y=\frac{3}{2\sqrt{3n}}$$

1.5. ТЕОРЕМА БЕРНУЛЛИ

Пусть произведены n независимых опытов, в каждом из которых возможно событие А с вероятностью p. Тогда относительная частота появления события А в n опытах стремится по вероятности к вероятности появления А в одном опыте.

<u>Теорема Бернулли:</u> *При неограниченном увеличении числа опытов частота события А сходится по вероятности к его вероятности p*:
$p^*(A) \xrightarrow[n\to\infty]{} p(A)$,

где $p^*(A)=\dfrac{m}{n}$ -- частота события А в n опытах, где m-число опытов в которых произошло событие А, n-число проведенных опытов или

$$\lim_{n\to\infty} P\left(\left|\frac{m}{n}-p\right| \le \varepsilon\right)=1 \text{ или } \frac{m}{n}\approx p .$$

Событие X_i – число появлений события А в i-м опыте. СВ X может принимать только два значения: X=1 (событие наступило) и X=0 (событие не наступило). Пусть СВ X_i – индикатор события А в i-м опыте. Числовые характеристики x_i: $m_i = p$ $D_i = pq$.

Они независимы, следовательно, можем применить теорему Чебышева:

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

$$\Rightarrow \lim_{n\to\infty} P\left(\left|\frac{X_1+X_2+.....+X}{n}-p\right|<\varepsilon\right)=1$$

Дробь $\dfrac{X_1+X_2+.....+X}{n}=\dfrac{m}{n}$ равна относительной частоте появлений события А в испытаниях $\Rightarrow$ получаем

$$\lim_{n\to\infty} P\left(\left|\frac{m}{n}-p\right| \le \varepsilon\right)=1.$$

Пояснения: В теореме речь идет лишь о вероятности того, что при достаточно большом числе испытаний относительная частота будет как угодно мало отличаться от постоянной вероятности появления события в каждом испытании.

Теорема Бернулли объясняет, почему относительная частота при достаточно большом числе испытаний обладает свойством устойчивости и оправдывает статистическое определение вероятности: в качестве <u>*статистической вероятности события*</u> принимают относительную частоту или число, близкое к ней.

1.6. ТЕОРЕМА ПУАССОНА

<u>Теорема Пуассона.</u> Частота события в **n** повторных испытаниях в каждом из которыхх оно может наступить соответствует с вероятностями $p_1, p_2, \dots p_n$ при $n\to\infty$ сходится <u>по вероятности</u> к средней арифметической вероятности наступления события в отдельных испытаниях

$$\lim_{n\to\infty} \quad P\left(\left|\frac{m}{n}-\frac{1}{n}\sum_{i=1}^{n} p\right| \le E\right)=1$$

1.6. ЦЕНТРАЛЬНАЯ ПРЕДЕЛЬНАЯ ТЕОРЕМА

Данная теорема определяет условия, при которых возникает СВ с нормальным законом распределения – т.е. *закон распределения суммы большого числа СВ* $X_1 + X_2 + \ldots + X_n$ *близок к нормальному.*

Эта теорема впервые была сформулирована русским математиком Ляпуновым А.М. (1857-1918).

Одна из простейших форм – относится к случаю одинаково распределенных слагаемых.

Теорема. *Если $X_1 \ldots X_n$-случайные независимые величины имеющие одно т тоже распределение с математическим ожиданием m и дисперсией σ^2, то при увеличении n закон распределения суммы*

$$Y = \sum_{i=1}^{n} X_i$$

неограниченно приближается к нормальному.

Теорема Ляпунова. *Пусть $X_1, \ldots, X_n$ – независимые случайные величины с математическими ожиданиями $m_1, \ldots, m_n$ и дисперсиями $D_1, \ldots, D_n$, причем при $n \to \infty$*

$$\lim_{n\to\infty}\left[\sum_{i=1}^{n}\left|\overset{0}{X}\right|^3 \Big/ \left(\sum_{i=1}^{n} D_i\right)^{3/2}\right] = 0.$$

При наличии данных условий закон распределения

$$Y_n = \sum_{i=1}^{n} X_n$$

неограниченно стремится к нормальному при $n \to \infty$

Например, теоремы Муавра–Лапласа – частный случай ЦПТ. Если производится m независимых опытов, в каждом из которых событие А появляется с вероятностью p, то справедливо соотношение:

$$P\left(\alpha < \frac{Y - np}{\sqrt{npq}} < \beta\right) = \Phi(\beta) - \Phi(\alpha)$$

Упрощенный вариант – Если СВ есть сумма большого числа независимых СВ, влияние которых на всю сумму мало, то Х имеет закон распределения, близкий к нормальному.

Пример1.2. Требуется произвести 60 выплат. Размер выплат случаен, но средняя выплата равна 50, а средне квадратичное отклонение равно 20.

1. Сколько должно быть денег в кассе, чтобы с вероятностью 0,95 хватило всем?

2. Сколько денег с вероятностью 0,95 останется в кассе, если первоначально было 3500.

Решение. Суммарная выплата $Y = \sum_{i=1}^{60} X_i$. На основании центральной предельной теореме для одинаково распределенных слагаемых Y имеет приблизительно нормальное распределение с параметрами

$$M[Y] = nM[X] = 50 \cdot 60 = 3000; \quad \sigma_y = \sqrt{n}\sigma_x = \sqrt{60} \cdot 20 \cong 154{,}8.$$

Необходимый запас определяем с использованием функции Лапласа:

$$\Phi\left(\frac{L - M[Y]}{\sigma_y}\right) + 0{,}5 = 0{,}95; \quad \frac{L - M[Y]}{\sigma_y} = 1{,}65; \quad L = 3255{,}6.$$

Остается

3500-3255,6=244.4.

1.7. ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ

1.7.1. Локальная теорема Муавра-Лапласа.

Теорема Муавра-Лапласа устанавливает условия, при которых биномиальную случайную величину можно приближённо рассматривать как нормальную.

Пусть $X \sim B(n, p)$. При $n \to \infty$

$$P_n(m) = C_n^m p^m q^{n-m} \approx \frac{1}{\sqrt{npq}} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{-x^2}{2}} = \frac{1}{\sqrt{npq}} \cdot \varphi(x)$$

Где $\varphi(x)$ – функция Гаусса, а

$$x = \frac{m - np}{\sqrt{npq}}$$

Ошибка приближения зависит от того, достаточно ли велико n, не слишком ли близко p к 0 или к 1 и каково интересующее нас значение m. Эта ошибка в настоящее время хорошо изучена и оценена; при необходимости всю нужную информацию можно найти в литературе.

1.7.2. Интегральная теорема Муавра-Лапласа.

Пусть $X \sim B(n, p)$. Тогда при $n \to \infty$ и любых фиксированных a и b, $a \leq b$:

$$\lim_{n \to \infty} P\left\{a \leq \frac{m - np}{\sqrt{npq}} \leq b\right\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{y^2}{2}} dy$$

Теорема Муавра-Лапласа позволяет уточнить связь относительной частоты и вероятности.