Московский государственный технический университет
имени Н. Э. Баумана

Факультет: Информатика и системы управления

Кафедра: Программное обеспечение ЭВМ и информационные технологии

Математическая статистика
Лекции

Москва, 2015 г.

Содержание

1 Предельные теоремы теории вероятностей

1.1 Неравенства Чебышева

Теорема 1.1 (*первое неравенство господина Чебышева*).

$$\left.\begin{array}{l} X \geq 0; \\ \exists MX. \end{array}\right\} \Rightarrow \forall \varepsilon > 0,\ \mathsf{P}\{X \geq \varepsilon\} \leq \frac{MX}{\varepsilon} \tag{1}$$

- X — случайная величина;
- $\mathsf{P}\{X \leq 0\} = 0$ так как $X \geq 0$.

Доказательство. Для непрерывной случайное величины X и зная, что при $X \geq 0 \Rightarrow f(x) = 0,\ x < 0$

$$MX = \int_{-\infty}^{+\infty} x\, f(x)\, dx = \int_{0}^{+\infty} x\, f(x)\, dx = \underbrace{\int_{0}^{\varepsilon} x\, f(x)\, dx}_{\geq 0} + \int_{\varepsilon}^{+\infty} x\, f(x)\, dx$$

учитывая $x \geq \varepsilon$

$$\underbrace{\int_{0}^{\varepsilon} x\, f(x)\, dx}_{\geq 0} + \int_{\varepsilon}^{+\infty} x\, f(x)\, dx \geq \int_{\varepsilon}^{+\infty} x\, f(x)\, dx \geq \varepsilon \cdot \int_{\varepsilon}^{+\infty} f(x)\, dx$$

где

$$\varepsilon \cdot \int_{\varepsilon}^{+\infty} f(x)\, dx = \varepsilon \cdot \mathsf{P}\{X \geq \varepsilon\}$$

таким образом

$$MX \geq \varepsilon \cdot \mathsf{P}\{X \geq \varepsilon\} \Rightarrow \mathsf{P}\{X \geq \varepsilon\} \leq \frac{MX}{\varepsilon}$$

□

Теорема 1.2 (*второе неравенство лорда Чебышева*).

$$\exists MX, \exists DX \Rightarrow \forall \varepsilon > 0,\ \mathsf{P}\big\{|X - MX| \geq \varepsilon\big\} \leq \frac{DX}{\varepsilon^2} \tag{2}$$

- X — случайная величина.

Доказательство. Выпишем дисперсию

$$DX = M\big[(X - MX)^2\big]$$

Рассмотрим случайную величину $Y = (X - MX)^2$, где $Y \geq 0$. Тогда из *первого неравенства Чебышева* следует, что $\forall \delta \geq 0$, $MY \geq \delta\, \mathsf{P}\{Y \geq \delta\}$, где получается, что $\delta = \varepsilon^2$.

$$\Big[DX = M\big[(X - MX)^2\big]\Big] \geq \Big[\varepsilon^2 \cdot \mathsf{P}\big\{(X - MX)^2 \geq \varepsilon^2\big\} = \varepsilon^2 \cdot \mathsf{P}\big\{|X - MX| \geq \varepsilon\big\}\Big]$$

таким образом

$$DX \geq \varepsilon^2 \cdot \mathsf{P}\big\{|X - MX| \geq \varepsilon\big\} \Rightarrow \mathsf{P}\big\{|X - MX| \geq \varepsilon\big\} \leq \frac{DX}{\varepsilon^2}$$

□

Пример 1.1. Предельно допустимое давление в пневмосистеме ракеты равна 200 (Па). После проверки большого количество ракет было получено среднее значение давления 150 (Па). Оценить вероятность того, что давление в пневмосистеме очередной ракеты будет больше 200 (Па), если по результатам проверки ракет было получено среднеквадратичное отклонение 5 (Па).

Решение. Имеем следующее:

- случайная величина X — давление в пневмосистеме;
- $X \geq 0$;
- $MX = 150$ (Па);
- $DX = 25$ (Па);

Решим поставленную задачу с помощью *первого неравенства Чебышева*

$$\left[\mathsf{P}\{X \geq \varepsilon\} = \mathsf{P}\{X \geq 200\}\right] \leq \left[\frac{MX}{\varepsilon} = \frac{150}{200} = \frac{3}{4} = 0.75\right]$$
$$\mathsf{P}\{X \geq 200\} \leq 0.75$$

Поскольку нам известна дисперсия почему бы не воспользоваться *вторым неравенством Чебышева*? Действуем. Для начало рассмотрим вероятность следующего события

$$\mathsf{P}\{X \geq \varepsilon\} = \mathsf{P}\{X \geq 200\} = \mathsf{P}\{X - \underbrace{150}_{MX} \geq \underbrace{50}_{\varepsilon}\}$$

Остаётся построить вероятность, которая будет удовлетворять форме *второго неравенства Чебышева* (т. е. сделать модуль).

$$\mathsf{P}\{X - 150 \geq 50\} \leq \mathsf{P}\{X - 150 \geq 50\} + \mathsf{P}\{X - 150 \leq -50\}$$

Так как *события* $\{X - 150 \geq 50\}$ и $\{X - 150 \leq -50\}$ *несовместные*, то по *формуле сложения вероятностей несовместных событий* получаем

$$\mathsf{P}\{X - 150 \geq 50\} + \mathsf{P}\{X - 150 \leq -50\} =$$
$$= \mathsf{P}\big\{\{X - 150 \geq 50\} + \{X - 150 \leq -50\}\big\} = \mathsf{P}\big\{|X - 150| \geq 50\big\}$$

Таким образом применяем *второе неравенство Чебышева*

$$\left[\mathsf{P}\big\{|X - MX| \geq \varepsilon\big\} = \mathsf{P}\big\{|X - 150| \geq 50\big\}\right] \leq \left[\frac{DX}{\varepsilon^2} = \frac{25}{50^2} = \left(\frac{5}{50}\right)^2 = 0.01\right]$$
$$\mathsf{P}\big\{|X - 150| \geq 50\big\} \leq 0.01$$

Ответ:

- с использованием *первого неравенства Чебышева* $\mathsf{P} \leq 0.75$;
- с использованием *второго неравенства Чебышева* $\mathsf{P} \leq 0.01$.

Замечание. *Второе неравенство Чебышева* даёт более точную оценку, так как используется информация о дисперсии случайной величины.

Замечание. Использование *первого неравенства Чебышева* при $\varepsilon < MX$ и *второго неравенства Чебышева* при $\varepsilon < \sqrt{DX}$ даёт тривиальную оценку: $\mathsf{P} \leq 1$.

1.2 Сходимость последовательности случайных величин

Будем считать, что $X_1, \dots, X_n, \dots$ — *последовательность случайных величин*, заданных на одном вероятностном пространстве.

Определение 1.1. Последовательность случайных величин $X_1, \dots, X_n, \dots$ *сходится по вероятности* к случайной величине Z, если $\forall \varepsilon > 0,\ \mathsf{P}\big\{|X_n - Z| \geq \varepsilon\big\} \xrightarrow[n\to\infty]{} 0$
Обозначение:

$$X_n \xrightarrow[n\to\infty]{\mathsf{P}} Z$$

Определение 1.2. Последовательность случайных величин $X_1, \dots, X_n, \dots$ *слабо сходится* к случайной величине Z, если $\forall x \in \Re$ где F_z непрерывна в точке x, числовая последовательность $F_{X_1}(x), \dots, F_{X_n}(x), \dots$ сходится к $F_Z(x)$. Обозначение:

$$F_{X_n}(x) \underset{n\to\infty}{\Longrightarrow} F_Z(x)$$

Замечание. Данные виды сходимости *неэквивалентны.*

1.3 Закон больших чисел (ЗБЧ)

Будем считать, что

- $X_1, \dots, X_n, \dots$ — *последовательность случайных величин*;
- $\exists MX_i = m_i$, где $i = \overline{1, \infty}$.

Определение 1.3. Последовательность случайных величин $X_1, \dots, X_n, \dots$ удовлетворяет *закону больших чисел*, если

$$\forall \varepsilon > 0,\ \mathsf{P}\left\{\left|\frac{1}{n} \cdot \sum_{i=1}^{n} X_i - \frac{1}{n} \cdot \sum_{i=1}^{n} m_i\right| \geq \varepsilon\right\} \xrightarrow[n\to\infty]{} 0$$

Замечание. Выполнение *закона больших чисел* для последовательности $X_1, \dots, X_n, \dots$ означает, что при достаточно больших n величина

$$Y_n = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i - \frac{1}{n} \cdot \sum_{i=1}^{n} m_i$$

практически теряет случайный характер.

Теорема 1.3 (*Закон больших чисел в форме Чебышева* или *достаточное условие выполнимости для последовательности случайных величин*)**.** Последовательность случайных величин $X_1, \dots, X_n, \dots$ удовлетворяет *закону больших чисел* тогда и только тогда, когда выполняются следующие условия:

- случайные величины $X_1, \dots, X_n, \dots$ — независимы;
- $\exists MX_i = m_i$, $\exists DX_i = \sigma_i^2$, $i = 1, 2, \dots$;
- Дисперсия случайных величин ограничена в совокупности т. е.

$$\exists C > 0 : \sigma_i^2 \leq C, \quad i = 1, 2, \dots$$

Доказательство.

$$\overline{X}_n = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i \quad n \in N$$

$$M\overline{X}_n = \frac{1}{n} \cdot \sum_{i=1}^{n} m_i$$

$$D\overline{X}_n = \frac{1}{n^2} \cdot D\left(\sum_{i=1}^{n} X_i\right) = \{X_i \text{ независимы}\} = \frac{1}{n^2} \cdot \sum_{i=1}^{n} DX_i = \frac{1}{n^2} \cdot \sum_{i=1}^{n} \sigma_i^2$$

Используем *второе неравенство Чебышева*

$$\mathsf{P}\big\{|\overline{X}_n - M\overline{X}_n| \geq \varepsilon\big\} \leq \frac{D\overline{X}_n}{\varepsilon^2}$$

В нашем случае

$$\mathsf{P}\left\{\left|\frac{1}{n} \cdot \sum_{i=1}^{n} X_i - \frac{1}{n} \cdot \sum_{i=1}^{n} m_i\right| \geq \varepsilon\right\} \leq \frac{1}{\varepsilon^2} \cdot \frac{1}{n^2} \cdot \sum_{i=1}^{n} \sigma_i^2 \leq \frac{1}{\varepsilon^2} \cdot \frac{1}{n^2} \cdot \sum_{i=1}^{n} C = \frac{1}{\varepsilon^2} \cdot \frac{1}{n^2} \cdot n\,C = \frac{C}{\varepsilon^2\, n}$$

таким образом

$$0 \leq \mathsf{P}\left\{\left|\frac{1}{n} \cdot \sum_{i=1}^{n} X_i - \frac{1}{n} \cdot \sum_{i=1}^{n} m_i\right| \geq \varepsilon\right\} \leq \frac{C}{\varepsilon^2\, n}$$

По *теореме о двух милиционерах*

$$\mathsf{P}\left\{\left|\frac{1}{n} \cdot \sum_{i=1}^{n} X_i - \frac{1}{n} \cdot \sum_{i=1}^{n} m_i\right| > \varepsilon\right\} \xrightarrow[n\to\infty]{} 0$$

□

Следствие. Пусть

- выполняется теорема о ЗБЧ в форме Чебышева
- X_i одинаково распределены т. е. $MX_i = m$, $DX_i = \sigma^2$, $i \in \aleph$

Тогда

$$\mathsf{P}\left\{\left|\frac{1}{n} \cdot \sum_{i=1}^{n} X_i - m\right| > \varepsilon\right\} \xrightarrow[n\to\infty]{} 0 \tag{3}$$

Следствие (Теорема Бернулли, ЗБЧ в форме Бернулли)**.** Пусть

- проводится серия испытаний по *схеме Бернулли*
  - с вероятностью успеха p;
  - с вероятностью неудачи $q = 1 - p$;
- наблюденная частота успеха $r_n = \{\text{число успехов в первых n испытаниях}\}/n$.

Тогда

$$r_n \xrightarrow[n\to\infty]{\mathsf{P}} p \tag{4}$$

Доказательство. Рассмотрим случайную величину

$$X_i = \begin{cases} 1, & \text{если в } i\text{-ом испытании успех;} \\ 0, & \text{иначе.} \end{cases}$$

тогда $MX_i = p$, $DX_i = p\,q$. Поскольку $\exists MX_i$, $\exists DX_i$ и X_i — независимы по определению *схемы Бернулли* $\Rightarrow$ выполняются все условия предыдущего следствия

$$\mathsf{P}\left\{\left|\underbrace{\frac{1}{n}\cdot\sum_{i=1}^{n} X_i}_{r_n} - p\right| > \varepsilon\right\} \xrightarrow[n\to\infty]{} 0$$

Таким образом

$$r_n \xrightarrow[n\to\infty]{\mathsf{P}} p$$

□

1.4 Центральная предельная теорема (ЦПТ)

Пусть имеется:

- последовательность независимых случайных величин $X_1, \ldots, X_n, \ldots$;
- X_i одинаково распределены;
- $MX_i = m$, $DX_i = \sigma^2$, $i \in \aleph$

Составим последовательность:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad n \in \aleph; \qquad M\overline{X}_n = m; \qquad D\overline{X}_n = \frac{1}{n^2}\sum_{i=1}^{n} DX_i = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{\sigma^2}{n}.$$

Составим случайную величину $Y = \dfrac{\overline{X}_n - M\overline{X}_n}{\sqrt{D\overline{X}_n}} = \dfrac{1/n\cdot\sum_{i=1}^{n} X_i - m}{\sigma/\sqrt{n}}$. Тогда $MY_n = 0$, $DY_n = 1$.

Теорема 1.4. Пусть выполнены условия приведённые выше, тогда последовательность Y_n *слабо сходится* к случайной величине $Z \sim N(0,1)$ т. е.

$$\forall x \in \Re, \quad F_{Y_n}(x) \xrightarrow[n\to\infty]{} \Phi(x), \quad \text{где}$$

- $\Phi(x) = \dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-\infty}^{x} e^{-t^2/2}\,dt$

Замечание. В этом случае говорят, что *случайная величина* Y_n имеет *асимптотически стандартное нормально распределение.*

Замечание. Для последовательности X_i удовлетворяющей *центральной предельной теореме* также выполнены условия *закона больших чисел в форме Чебышева.* Поэтому из закона больших чисел $\Rightarrow$ $\overline{X}_n \xrightarrow[n\to\infty]{} m$. *Центральная предельная теорема* уточняет характер этой сходимости.

Замечание. *Центральная предельная теорема* иллюстрирует особую роль *нормального распределения.* Все естественные процессы, протекание которых обусловлено многочисленными случайными факторами (независимыми) “в среднем” имеют нормальное распределение.

Теорема 1.5 (ЦПТ Муавра-Лапласа)**.** Пусть

- проводится $n \gg 1$ испытаний по *схеме Бернулли* с вероятностью успеха p ($q = 1-p$ — вероятность неудачи);
- k — число успехов из n испытаний.

Тогда $\mathsf{P}\{k_1 \le k \le k_2\} \approx \Phi_0(x_2) - \Phi(x_1)$, где

- $x_i = \frac{k_i - np}{\sqrt{npq}}$, $i = \overline{1,2}$;
- $\Phi_0(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2}\,dt$.

Доказательство. Рассмотрим $X_i = \begin{cases} 1, \text{в } i\text{-ом испытании успех} \\ 0, \text{иначе} \end{cases}$ при этом:

- X_i, $i \in \aleph$ — независимы
- $MX_i = p$, $DX_i = p\,q$
- $\sum_{i=1}^n X_i$ — число успехов в серии из n испытаний

Тогда имеем

$$\mathsf{P}\{k_1 \le k \le k_2\} = \mathsf{P}\{k_1 \le \sum_{i=1}^n X_i \le k_2\} =$$

применяем *центральную предельную теорему* к $x_1, \ldots, x_n$ (проверить)

$$= \mathsf{P}\left\{\frac{1}{n}k_1 - m \le \frac{1}{n}\sum_{i=1}^n X_i - m \le \frac{1}{n}k_2 - m\right\} =$$

$$= \mathsf{P}\left\{\frac{1/n\ k_1 - m}{\sqrt{p\,q}/\sqrt{n}} \le \frac{1/n\ \sum_{i=1}^n X_i - m}{\sqrt{p\,q}/\sqrt{n}} \le \frac{1/n\ k_2 - m}{\sqrt{p\,q}/\sqrt{n}}\right\} =$$

$\dfrac{1/n\ \sum_{i=1}^n X_i - m}{\sqrt{p\,q}/\sqrt{n}} = Y_n \sim N(0,1)$ — приближённо, так как $n \gg 1$

$$= \mathsf{P}\{x_1 \le Y_n \le n_2\} \approx \Phi_0(x_2) - \Phi_0(x_1)$$

$$x_i = \frac{1/n\ k_i - m}{\sqrt{p\,q}/\sqrt{n}} = \frac{k_i - mp}{\sqrt{npq}}$$

□

Пример 1.2. В эксперименте Пиреосса о подбрасывании монеты из 24000 бросков герб выпал 12012 раз. Какова вероятность того, что при повторном испытании отклонение относительной частоты успеха окажется таким же или больше.

Решение. Используем *схему Бернулли*

- $n = 24000$, $p = q = 1/2$

- $A = \{$отклонение окажется не меньше$\}$
- $\overline{A} = \{$отклонение окажется меньше$\}$

Тогда имеем

$$\mathsf{P}(A) = 1 - \mathsf{P}(\overline{A}) = 1 - \mathsf{P}\{12000 - 12 < k < 12000 + 12\} = 1 - \mathsf{P}\{11988 < k < 12012\} \approx$$

Муавра-Лапласа

$$\approx 1 - \left[\Phi_0(x_2) - \Phi_0(x_1)\right] = 1 - 2\Phi_0(0.155) \approx -0.877$$

$$x_1 = \frac{11988 - 12000 \cdot 1/2}{\sqrt{24000 \cdot 1/2 \cdot 1/2}} \approx -0.155; \qquad\qquad x_2 \approx 0.155$$