

Лекция 2

Видорка. Статистики.
Эмпирические функции распределения.

Таким X -СВ с ф.р. $F(x) = P\{X \leq x\}$,
 $x \in (-\infty, +\infty)$.

Определение 1. Собокупность $\{X_k, k=1, 2, \dots, n\}$ независимых СВ, имеющих одинаковое ф.р. $F_{X_k}(x) = F(x)$, называемое однородной видоркой одинаковой, соответствующей ф.р. $F(x)$ и обозначаемой Z_n :

$$Z_n = \{X_1, X_2, \dots, X_n\}.$$

СВ X_k ($k=1, 2, \dots, n$) называемые k -и элеменами видорки.

Определение 2. Реализации видорки Z_n называемые неслучайные вектором $Z_n = \{x_1, \dots, x_n\}$, компонентами которого является реализации СВ X_1, \dots, X_n .

Таким $Z_n = \{x_1, \dots, x_n\}$ — вектор компонентами которого является упорядочение по возрастанию числа x_1, \dots, x_n , т.е. $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$.

Определение 3. СВ $X_{(k)}$, реализации которых для каждого реализации Z_n является число $x_{(k)}$, называемое k -и непрерывной статистикой, $k=1, \dots, n$. При этом СВ $X_{(1)}$ и $X_{(n)}$ называются экспрессивными статистиками.

Если $\varphi(\dots)$ — некоторое заданное
функции от n переменных, то итог
суммируем величина Y , определяемая
как $Y = \varphi(X_1, X_2, \dots, X_n)$, называемая
статистикой.

Пример статистик:

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k \quad - \text{видеоформное среднее}$$

$$\bar{S}_n^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X}_n)^2 \quad - \text{видеоформленное дисперсионное}$$

$$\bar{Y}_r(n) = \frac{1}{n} \sum_{k=1}^n (X_k)^r, \quad r=1, 2, \dots \quad - \text{видеоформленные}\newline \text{каскадные}\newline \text{моменты}\newline r-\text{го порядка}$$

$$\bar{M}_r(n) = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X}_n)^r, \quad r=1, 2, \dots \quad - \text{видеоформленные}\newline \text{центрированные}\newline \text{моменты}\newline r-\text{го порядка}$$

Замечание, что $\bar{X}_n = \bar{Y}_1(n)$, $\bar{S}_n^2 = \bar{M}_2(n)$.

Теорема (без док-ва). Пусть функция —
ение $F(x)$ непрерывно, что означает вычисление свя-
зующее теоретические моменты этого
представления X_k видоформки Z_n :

$$m_x = M\{X_k\}, \quad Y_r = M\{(X_k)^r\}, \quad M_r = M\{(X_k - m_x)^r\}, \quad r=1, 2, \dots$$

. Понятие

$$\bar{Y}_r(n) \xrightarrow[n \rightarrow \infty]{n.H.} Y_r, \quad \bar{M}_r(n) \xrightarrow[n \rightarrow \infty]{n.H.} M_r.$$

Введение функции распределения:

$$I(A) = \begin{cases} 1, & \text{если событие } A \text{ произошло} \\ 0, & \text{если } A \text{ не произошло} \end{cases}$$

Для каждого $x \in (-\infty, +\infty)$ определяем
случайную величину $F_n^*(x)$ и определяем

$$F_n^*(x) = \frac{1}{n} \sum_{i=1}^n I(X_i < x).$$

Здесь $\sum_{i=1}^n I(X_i < x)$ есть случайное число
элементов выборки, меньших x . Эту
функцию будем называть эмпириче-
ской функцией распределения выборки.
Каждое реализующее выборки дадут
свою реализацию случайной функции
 $F_n^*(x)$ (все этих реализаций ~~есть~~
существует в лекции 1)

Непрерывное утверждение, что при
фиксированном x существует
вероятность

$$\sum_{i=1}^n I(X_i < x)$$

имеет следующее значение
с наращиванием n и $p = P\{X_i < x\} = F(x)$,
так что

$$P\left\{F_n^*(x) = \frac{k}{n}\right\} = C_n^k F(x)^k (1-F(x))^{n-k},$$

где $k = 0, 1, \dots, n$.

Доказательство следующего теоремы 2, показывающее, что $F_n^*(x)$ с убыванием n сближается в конечной мере x с теоремической ф.р. $F(x)$.

Теорема 2 Для любого $x \in (-\infty; +\infty)$ и любого $\epsilon > 0$

$$\lim_{n \rightarrow \infty} P\{|F_n^*(x) - F(x)| < \epsilon\} = 1, \text{ т.е. } F_n^*(x) \xrightarrow{P} F(x).$$

Док-во. Рассмотрим систему независимых случайных величин

$$\eta_i = \begin{cases} 1, & \text{если } X_i < x \\ 0, & \text{если } X_i \geq x \end{cases}, \quad i = 1, 2, \dots, n.$$

Каждая из η_i имеет один и тот же закон распределения

$$\begin{array}{c|cc} \eta_i & 1 & 0 \\ \hline P & p & 1-p \end{array}, \quad M\eta_i = p, \quad D\eta_i = p(1-p)$$

т.е. $p = P\{X_i < x\} = F(x)$. Так как это выполнение все условие ЗБЧ в форме Чебышёва (см. теорему Чебышёва в курсе ТВ) и

$$\lim_{n \rightarrow \infty} P\left\{\left|\frac{\eta_1 + \dots + \eta_n}{n} - p\right| < \epsilon\right\} = 1.$$

Но $p = F(x)$, $\frac{\eta_1 + \dots + \eta_n}{n} = F_n^*(x)$.

Теорема 2 доказана.

Замечание. Описание (без доказательства), что имеет место и более сильный предустановлен:

$$F_n^*(x) \xrightarrow{n \cdot n} F(x).$$

Эти свойства говорят о некоторой сходимости $F_n^*(x)$ к $F(x)$ при $n \rightarrow \infty$. Оказывается, существует и более сильное свойство, определяющее фактическую сходимость $F_n^*(x)$ к $F(x)$.

Теорема 3 (Гимбентко) (без док-ва)
Сходимость исследований СВ

$$Y_n = \sup_{-\infty < x < +\infty} |F_n^*(x) - F(x)|, \quad n=1,2,\dots$$

При этом

$$Y_n \xrightarrow[n \rightarrow \infty]{n \cdot n} 0.$$

Таким образом, теоремы 2, 3 говорят о том, что $F_n^*(x)$ при больших значениях n является хардом приближением для $F(x)$.

В заключение, получим ещё один предустановлен, позволяющий в вероятностном смысле отбросить некоторые приближение для $F(x)$ с помощью $F_n^*(x)$. Применение ЧПП (см. курс ТВ) к выведенной при доказательстве теоремы 2 исследований не-затрудняющихся СВ η_1, η_2, \dots по ЧПП

- 6 -

если $V_n = \gamma_1 + \dots + \gamma_n$, $\tilde{V}_n = \frac{V_n - M\{V_n\}}{\sqrt{D\{V_n\}}}$,

то $\lim_{n \rightarrow \infty} P\{\alpha \leq \tilde{V}_n \leq \beta\} = \int_{\alpha}^{\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$

В нашем случае

$$V_n = \gamma_1 + \dots + \gamma_n = n F_n^*(x),$$

$$M\{V_n\} = np = nF(x),$$

$$D\{V_n\} = n \cdot p(1-p) = nF(x)(1-F(x)),$$

так что

$$\lim_{n \rightarrow \infty} P\left\{\alpha \leq \frac{n F_n^*(x) - n F(x)}{\sqrt{n F(x)(1-F(x))}} \leq \beta\right\} = \int_{\alpha}^{\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

В частности,

$$\lim_{n \rightarrow \infty} P\left\{\frac{\sqrt{n}|F_n^*(x) - F(x)|}{\sqrt{F(x)(1-F(x))}} \leq \varepsilon\right\} = 2\Phi(\varepsilon),$$

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