Теория вероятностей и математическая статистика
Лектор А.А. Лобузов

Семестр 6 **Лекция 4**

Эмпирическая функция распределения

Пусть $\mathbf{X} = (X_1, X_2, ..., X_N)$ − случайная выборка объёма N, $\mathbf{x} = (x_1, x_2, ..., x_N)$ − выборка (реализация).

При упорядочении $\mathbf{x}$ получаем числовой вариационный ряд $x_{(1)} \le x_{(2)} \le ... \le x_{(N-1)} \le x_{(N)}$.

Случайная величина $X_{(k)}$, принимающая значение $x_{(k)}$, называется k-ой порядковой статистикой.

$$F_N(x, \mathbf{x}) = \frac{1}{N} \sum_{k=1}^{N} I_{(-\infty, x]}(x_k) = \frac{1}{N} \sum_{k=1}^{N} I_{(x_k \le x]}$$

$F_N(x, \mathbf{x}) = F_N(x, x_1, x_2, ..., x_N)$ − зависит от выборки.

$$F_N(x, \mathbf{x}) = \begin{cases} 0, x < x_{(1)}; \\ \frac{k}{N}, x_{(k)} \le x < x_{(k+1)}, 1 \le k \le N; \\ 1, x \ge x_{(N)}. \end{cases}$$

$$F_N(x, x_1, x_2, \ldots, x_N) = \sum_{x_k \le x} \frac{1}{N} = \begin{cases} 0, & x < x_{(1)}, \\ \frac{1}{N}, & x_{(1)} \le x < x_{(2)}, \\ \frac{2}{N}, & x_{(2)} \le x < x_{(3)}, \\ \frac{3}{N}, & x_{(3)} \le x < x_{(4)}, \\ \ldots\ldots\ldots\ldots\ldots\ldots, \\ 1, & x \ge x_{(N)}. \end{cases}$$

Эмпирическая функция распределения для выборки дискретной случайной величины

$$F_N^{Э}(x) = \sum_{x_i^* \le x} w_i = \begin{cases} 0, & x < x_1^*, \\ w_1, & x_1^* \le x < x_2^*, \\ w_1 + w_2, & x_2^* \le x < x_3^*, \\ w_1 + w_2 + w_3, & x_3^* \le x < x_4^*, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \\ 1, & x \ge x_m^*. \end{cases}$$

Теорема В.И.Гливенко (1933 г.)

Пусть $\mathbf{X} = (X_1, X_2, \ldots, X_N)$ – случайная выборка из распределения с.в. ξ,

$F_N(x, \mathbf{X})$ – эмпирическая функция распределения,

$F(x)$ – теоретическая функция распределения.

Тогда

$$\sup\{|F_N(x,\mathbf{X}) - F(x)| : -\infty < x < +\infty\} \xrightarrow[N\to\infty]{\text{п.н.}} 0$$

Равенство почти-наверное (п.н.) (почти всюду)

$$\xi \overset{\text{п.н.}}{=} \eta \Leftrightarrow P(\underbrace{\xi = \eta}_{\{\omega \mid \xi(\omega) = \eta(\omega)\}}) = 1$$

$\xi : \Omega \to \mathbb{R}$

$\eta : \Omega \to \mathbb{R}$

Сходимость почти всюду

$$\xi_n \xrightarrow[n\to\infty]{\text{п.н.}} \eta \Leftrightarrow \{\omega \mid \lim_{n\to\infty} \xi_n(\omega) = \eta(\omega)\} = 1$$

Свойства эмпирической функции распределения

$$NF_N(x,\mathbf{X}) = \sum_{k=1}^{N} I_{(X_k \le x]} - \text{число } X_k,$$

попавших в интервал $(-\infty, x]$

Значит $NF_N(x,\mathbf{X}) \sim Bi(N,p)$, где

$p = P(\xi \le x) = F(x)$

$$P(NF_N(x,\mathbf{X}) = k) = C_N^k F(x)^k (1-F(x))^{N-k}$$

$$F_N(x,\vec{X}) = \frac{1}{N} \sum_{j=1}^{N} I_{(X_j \le x)}$$

$\{I_{(X_j \le x)}\}_{j=1}^{N}$ — Н.О.Р.С.В.

$I_{(X_j \le x)}$	0	1
P	q	p

$q = 1-p$

$p = P(X_j \le x) = F(x)$

Математическое ожидание

$$\mathrm{M}(NF_N(x,\mathbf{X})) = Np = N\,F(x)$$

$$\mathrm{M}(F_N(x,\mathbf{X})) = p = F(x)$$

Дисперсия

$$\mathrm{D}(NF_N(x,\mathbf{X})) = Npq = N\,F(x)(1-F(x))$$

$$\mathrm{D}(F_N(x,\mathbf{X})) = \frac{F(x)(1-F(x))}{N} \xrightarrow[N\to\infty]{} 0$$

Значит по ЗБЧ: $F_N(x,\mathbf{X}) \xrightarrow[N\to\infty]{\mathrm{P}} F(x)$

$I_{(X_k \le x]}$ – н.о.р.с.в., имеющие распределение Бернулли

$$p = P(I_{(X_k \le x]} = 1) = F(x)$$

$$NF_N(x,\mathbf{X}) = \sum_{k=1}^{N} I_{(X_k \le x]}$$ – сумма н.о.р.с.в.

По ЦПТ нормированная сумма н.о.р.с.в. слабо сходится к $N(0,1)$:

$$\frac{NF_N(x,\mathbf{X}) - N\,F(x)}{\sqrt{N\,F(x)(1-F(x))}} \xrightarrow[N\to\infty]{w} N(0,1)$$

$I_{(X_k \le x]}$ – н.о.р.с.в., имеющие распределение Бернулли

$p = P(I_{(X_k \le x]} = 1) = F(x)$

$NF_N(x, \mathbf{X}) = \sum_{k=1}^{N} I_{(X_k \le x]}$ – сумма н.о.р.с.в.

По ЦПТ нормированная сумма н.о.р.с.в.
слабо сходится к $N(0,1)$:

$$\frac{NF_N(x, \mathbf{X}) - N\,F(x)}{\sqrt{N\,F(x)(1 - F(x))}} \xrightarrow[N \to \infty]{w} N(0,1)$$