Выборкой из генеральной совокупности объёма n с распределением P называется последовательность $X_1, \dots, X_n$ н. о. р. с. в. (независимых одинаково распределённых случайных величин) с распределением P

Обозначение: $Low(X_i) = P$, или $X_i \sim P$.

$X = (X_1, \dots, X_n)^T$

$x_i = X_i(\omega)$, ω (элементарное событие) одно и то же при всех i, мы определяем Ω (пространство элементарных событий) с помощью проведения опыта

$x = (x_1, \dots, x_n)^T$ - всё, с чем имеет дело матстатистика, x тоже называют выборкой, но после опыта

Упорядоченная выборка называется вариационным рядом $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$, члены этого ряда называются порядковыми статистиками: $X_{(k)}$ - k-ая порядковая статистика

Если ω фиксированно, то на данной выборке можно построить распределение, оно и его функция распределения будут обычными. Если ω не зафиксированно, то распределение и функция распределения будут случайными

$F_n(x) = \frac{1}{n}\sum\limits_{i=1}^{n} \chi(x - X_i) = \frac{K}{n}$ - оценка распределения, где K - число элементов выборки, меньших x, χ - ступенька Хевисайда, $\chi(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases}$

$F_n(x)$ - эмпирическая функция распределения

$EF_n(x) = \frac{1}{n}\sum\limits_{i=1}^{n} E\chi(x - X_i) = \frac{1}{n}\sum\limits_{i=1}^{n} \mathbb{P}(\{X_i < x\}) = \frac{1}{n}\sum\limits_{i=1}^{n} F(x) = F(x)$ - матожидание функции равномерного распределения совпадает с функцией распределения генеральной совокупности

$\mathbb{P}(\{X_{(k)} \in x,[x+dx))) = *k-1$ элементов $< x$, один элемент в интервале (всего элементов $n)* = n\mathbb{P}(\{X_1 \in x, x+dx;$ ровно $k-1$ элементов выборки $< x\}) = P([x, x+dx))C_{n-1}^{k-1}(F(x))^{k-1}(1-F(x))^{n-k} + o(dx)$, устремим $dx \to 0$ и получим плотность k-ой порядковой статистики

$f_{(k)} = nf(x)C_{n-1}^{k-1}(F(x))^{k-1}(1-F(x))^{n-k}$

Свойства $F_n(x)$ - эмпирической функции распределения:

1)$EF_n(x) = F(x)\ \forall x \in \mathbb{R}^1$

Свойство 1) - свойство несмещённости оценки, $F_n(x)$ в среднем совпадает с $F(x)$

2) По ЗБЧ (закону больших чисел) $F_n(x) \to^{\mathbb{P}} F(x)$

т. е. $\forall \epsilon > 0\ \mathbb{P}(\{|F_n(x) - F(x)| > \epsilon\}) \to_{n\to\infty} 0$

Свойство 2) - свойство состоятельности оценки, $F_n(x)$ близко к $F(x)$ в каком-то смысле

3)$\sqrt{n}(F_n(x) - F(x)) = \frac{1}{\sqrt{n}}\sum\limits_{i=1}^{n}(\chi(x - X_i) - F(x)) \to^d N(0, \sigma^2)$ по центральной предельной теореме, где $\sigma^2 = E\chi^2(x - X_i) - (F(x))^2 = F(x)(1 - F(x))$

Свойство 3) - свойство асимптотической нормальности, позволяет оценивать погрешность

4) Теорема Гливенко

$\sup\limits_{x\in\mathbb{R}^1} |F_n(x) - F(x)| \to_{n\to\infty} 0$ с вероятностью 1

5) Теорема Колмогорова

$\sqrt{n}\sup\limits_{x\in\mathbb{R}^1} |F_n(x) - F(x)| \to^d \varsigma \sim K$

ς имеет распределение Колмогорова, $\mathbb{P}(\varsigma < t) = \sum\limits_{j=-\infty}^{+\infty} (-1)^i e^{-2j^2t^2}$