Раздел 2. *Основы статистического анализа данных*

2.1. Определение случайной выборки

Пусть X – исследуемая случайная величина, $F_X(x) = P(X < x)$ – ее функция распределения, вообще говоря, неизвестная. В некоторых случаях может быть известен вид распределения случайной величины, а неизвестными являются один или несколько параметров, от которых зависит функция распределения. Ради краткости в записи $F_X(x)$ индекс может в дальнейшем опускаться. Условимся также указывать, непрерывной или дискретной является исследуемая случайная величина.

$\underline{def}|$ Пусть проводится серия n независимых наблюдений (измерений) случайной величины X в одних и тех же условиях (эксперимент). В результате эксперимента получают n чисел – значений $x_1, x_2, \ldots, x_n$, которые случайная величина X последовательно принимала в данной серии наблюдений. Эти числа будем считать значениями n одинаково распределенных независимых случайных величин $X_1, \ldots, X_n$, каждая из которых распределена так же, как исследуемая случайная величина X: $F_{X_i}(x) = F_X(x)$. Конечную последовательность n независимых одинаково распределенных случайных величин называют *случайной выборкой* $X_1, \ldots, X_n$ (короче – выборкой) из распределения $F_X(x)$, а указанные числа $x_1, x_2, \ldots, x_n$, полученные в данном эксперименте – *реализацией выборки*.

Отметим, что множество всех возможных значений исследуемой случайной величины называют генеральной совокупностью.

На основе выборок строят *оценки параметров распределения* исследуемой случайной величины X, таких как математическое ожидание, стандартное отклонение и других, а также судят о виде функции распределения $F_X(x)$.

Понятно, что числа $x_1, x_2, \ldots, x_n$ можно также рассматривать как значение n-мерной случайной величины $(X_1, \ldots, X_n)$, компоненты которой $X_1, \ldots, X_n$ независимы и одинаково распределены.

$\underline{def}|$ Всякую функцию выборки $\varphi(X_1, \ldots, X_n)$ называют статистикой. Статистика $\varphi(X_1, \ldots, X_n)$ – случайная величина, распределение которой зависит от распределения $F_X(x)$, из которого извлечена выборка, и от объема выборки n.

2.2. Закон распределения порядковых статистик

Пусть $X_1, \ldots, X_n$ – выборка объема n из распределения $F_X(x)$; $x_1, x_2, \ldots, x_n$ – некоторая ее реализация.

Упорядочим числа $x_1, x_2, \ldots, x_n$ по возрастанию и обозначим их следующим образом: $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$, где $x_{(1)} = min\,(x_1, x_2, \ldots, x_n)$, $x_{(n)} = max\,(x_1, x_2, \ldots, x_n)$, $x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$.

Представим, что упорядочены все возможные реализации выборки $X_1, \ldots, X_n$ и введем новую случайную величину $X_{(k)}$ – порядковую статистику порядка k $(k = 1, 2, \ldots, n)$.

Множество возможных значений случайной величины $X_{(k)}$ определим так: оно состоит из тех и только тех чисел $x^i_{(k)}$, которые оказываются на k-м месте при упорядочении любой реализации $x_1, x_2, \ldots, x_n$ выборки $X_1, \ldots, X_n$ (индекс $i = 1,2,\ldots$ – номер реализации).

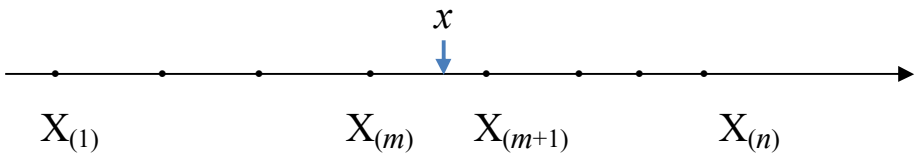
$$
\begin{array}{ccccc}
x^1_{(1)}, & \ldots & x^1_{(k)} & \ldots & x^1_{(n)} \\
\cdot & & \cdot & & \cdot \\
x^i_{(1)} & \ldots & x^i_{(k)} & \ldots & x^i_{(n)} \\
\cdot & & \cdot & & \cdot \\
\hline
X_{(1)} & \ldots & X_{(k)} & \ldots & X_{(n)}
\end{array}
$$

Таким образом, по выборке $X_1, \ldots, X_n$ построена последовательность $X_{(1)}, \ldots, X_{(k)}, \ldots, X_{(n)}$, называемая *вариационным рядом*. Элементы вариационного ряда – порядковые статистики удовлетворяют соотношениям: $X_{(1)} \le \ldots \le X_{(k)} \le \ldots \le X_{(n)}$, (это означает, что в любой *реализации* вариационного ряда числа $x^i_{(1)}, \ldots, x^i_{(k)}, \ldots, x^i_{(n)}$ связаны неравенствами $x^i_{(1)} \le \ldots \le x^i_{(k)} \le \ldots \le x^i_{(n)}$, где верхний индекс i – номер реализации, $i = 1, 2, \ldots$).

Найдем функцию распределения k-й порядковой статистики

$$X_{(k)}\colon F_{X_{(k)}}(x) = P(X_{(k)} < x) \quad (k = 1, 2, \ldots, n).$$

Эмпирической частотой $N_n(x)$ назовем случайную величину, равную числу элементов выборки $X_1, \ldots, X_n$, меньших x (иначе – числу элементов вариационного ряда $X_{(1)}, \ldots, X_{(n)}$, меньших x). Ясно, что возможные значения эмпирической частоты $N_n(x)$ – число осуществлений события $(X < x)$ на выборке $X_1, \ldots, X_n$ объема n – это числа $m = 0,1,\ldots, n$. Действительно,



$$(N_n(x) = 0) = (x \le X_{(1)});$$

$$(N_n(x) = m) = (X_{(m)} < x \le X_{(m+1)}) \quad \forall m = 1, \ldots, n-1;$$

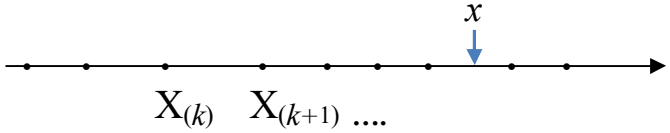
$$(N_n(x) = n) = (x > X_{(n)}).$$

Извлечение выборки из распределения $F_X(x)$ представляет собой серию n независимых испытаний (n измерений, регистраций значений) исследуемой случайной величины X. Для каждого из указанных испытаний вероятность события $(X < x)$ равна $P(X < x) = F_X(x)$.

Отсюда следует, что случайная величина $N_n(x)$ подчиняется биномиальному распределению:

$$P(N_n(x) = m) = C_n^m (F_X(x))^m (1 - F_X(x))^{n-m} \quad (m = 0,1,\ldots,n).$$

Заметим, что события $(X_{(k)} < x)$ и $(N_n(x) \ge k)$ равносильны,



то есть $(X_{(k)} < x) = (N_n(x) \ge k) = \sum_{m=k}^{n} (N_n(x) = m)$.

Таким образом, получаем закон распределения порядковых статистик:

$$P(X_{(k)} < x) = F_{X_{(k)}}(x) = \sum_{m=k}^{n} C_n^m (F_X(x))^m (1 - F_X(x))^{n-m}, \forall\ k = 1, \ldots, n$$

При $k = 1$ и $k = n$ имеем распределения *экстремальных порядковых статистик*:

минимальной $X_{(1)}$: $F_{X_{(1)}}(x) = 1 - (1 - F_X(x))^n$ и

максимальной $X_{(n)}$: $F_{X_{(n)}}(x) = (F_X(x))^n$.

2.3. Эмпирическая функция распределения

Пусть $X_1, \ldots, X_n$ – выборка из распределения $F_X(x)$, $X_{(1)}, \ldots X_{(k)}, \ldots, X_{(n)}$ – вариационный ряд, $N_n(x)$ – эмпирическая частота. $\underline{def}$ Случайная величина $F_n(x) = N_n(x)/n$ называется *эмпирической функцией распределения*; ее смысл – *относительная частота* числа элементов выборки $X_1, \ldots, X_n$, удовлетворяющих условию $X_i < x$.

Ясно, что множество возможных значений эмпирической функции распределения $F_n(x)$ есть: $0, 1/n, \ldots, m/n, \ldots, n/n$.

События $(F_n(x) = m/n)$ и $(N_n(x) = m)$ – равносильны, эмпирическая частота $N_n(x)$ распределена по биномиальному закону, поэтому

$$P(F_n(x) = m/n) = C_n^m (F_X(x))^m (1 - F_X(x))^{n-m} \qquad (m = 0,1,\ldots,n) -$$

– закон распределения эмпирической функции распределения $F_n(x)$.

Заметим, что для каждой реализации выборки реализация эмпирической функции распределения обладает всеми свойствами функции распределения. Действительно, пусть $x_1, x_2, \ldots, x_n$ – некоторая реализация выборки, $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ – соответствующая реализация вариационного ряда $(x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)})$. Среди чисел $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ выберем только различные и обозначим через

$x_{i_1}, x_{i_2}, \ldots, x_{i_k}$, где $x_{i_1} = x_{(1)}$, $x_{i_k} = x_{(n)}$, $x_{i_1} < \ldots < x_{i_m} < \ldots < x_{i_k}$, тогда:

x_{i_1}	$\ldots$	x_{i_m}	$\ldots$	x_{i_k}
n_1	$\ldots$	n_m	$\ldots$	n_k
n_1/n	$\ldots$	n_m/n	$\ldots$	n_k/n

Здесь n_m – *абсолютная частота* элемента x_{i_m} $(\sum_{m=1}^{k} n_m = n)$, n_m/n – *относительная частота* $(\sum_{m=1}^{k} \frac{n_m}{n} = 1)$.

Введем *дискретную* случайную величину X^* (*эмпирическую случайную величину*), заданную рядом распределения:

X^*	x_{i_1}	…	x_{i_m}	…	x_{i_k}
$\mathrm{P}(\mathrm{X}^* = x_{i_m})$	n_1 / n	…	n_m / n	…	n_k / n

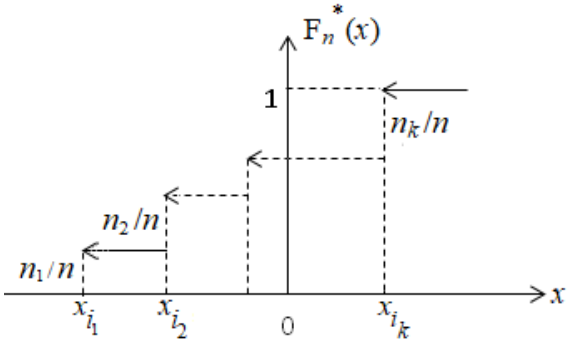
Заметим, что таким образом каждому элементу реализации выборки $x_1, x_2, \ldots, x_n$ сопоставлена вероятность $1 / n$.

Обозначим через $\mathrm{F}_n^*(x)$ *реализацию* эмпирической функции $\mathrm{F}_n(x)$ (*случайной величины*), отвечающую данной *реализации* выборки. Очевидно, $\mathrm{F}_n^*(x)$ – функция распределения эмпирической случайной величины X^*:

$$\mathrm{F}_n^*(x) = \mathrm{P}(\mathrm{X}^* < x) = \sum_{m:\, x_{i_m} < x} \frac{n_m}{n}.$$

График *реализации* $\mathrm{F}_n^*(x)$ эмпирической функции распределения $\mathrm{F}_n(x)$ (для некоторой реализации выборки) приведен ниже:

В каждой точке x, кроме точек x_{i_m}, функция $\mathrm{F}_n^*(x)$ непрерывна; в точках x_{i_m} – она непрерывна слева, величина скачка справа равна n_m / n (m = 1, 2, …, k).



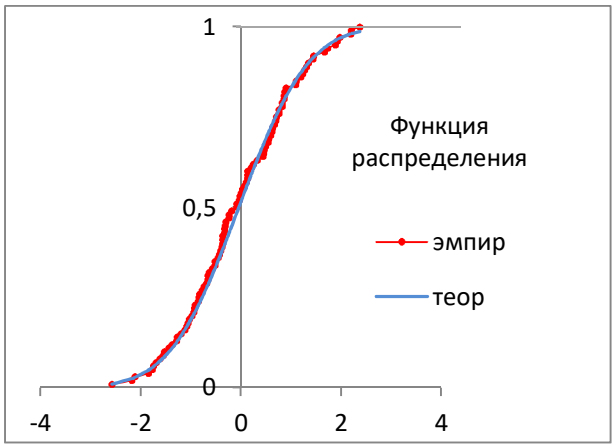
Эмпирическая функция распределения $\mathrm{F}_n(x) = N_n(x) / n$ – относительная частота числа осуществлений события $(\mathrm{X} < x)$ на выборке, сходится по вероятности к вероятности этого события $\mathrm{P}(\mathrm{X} < x) = \mathrm{F}_\mathrm{X}(x)$ при любом x:

$$\forall x \;\; \mathrm{F}_n(x) \xrightarrow[n \to +\infty]{P} \mathrm{F}_\mathrm{X}(x).$$

Поэтому, если объем выборки n достаточно велик, то значение эмпирической функции распределения $F_n(x)$ (функции выборки $X_1, \ldots, X_n$) в каждой точке x оказывается близким к соответствующему значению теоретической функции распределения $F_X(x)$, вообще говоря, неизвестной.

Доказано (*теорема Гливенко*): $\forall \varepsilon > 0 \; P\left(\sup_x |F_n(x) - F_X(x)| \le \varepsilon\right)_{n \to +\infty} 1.$

Отклонение *эмпирической* функции распределения $F_n(x)$ от *теоретической* функции распределения $F_X(x)$ с вероятностью единица стремится к нулю с ростом объема выборки n, при этом $F_n(x)$ служит равномерным приближением $F_X(x)$ на всей числовой оси. Заметим, что разность $(F_n(x) - F_X(x))$ асимптотически нормальна с нулевым математическим ожиданием.



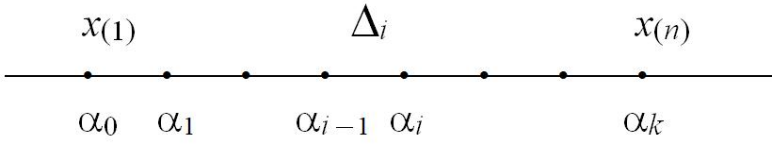
2.4. Группирование выборочных данных, гистограмма

Эмпирическая функция распределения является характеристикой выборки, позволяющей наглядно представлять статистические данные и выдвигать предположения о виде неизвестной функции распределения исследуемой (наблюдаемой) случайной величины.

Другой способ представления статистического материала – это построение группированного статистического ряда и гистограммы.

Пусть исследуемая случайная величина X – непрерывна. Если выборка достаточно большая (обычно в статистике большими считают выборки объемом $n \geq 100$), то ее реализацию $(x_1, x_2, \ldots, x_n)$ упорядочивают $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$ и подвергают группировке следующим образом.

Отрезок $[x_{(1)}; x_{(n)}]$, где $x_{(1)} = min\ (x_1, x_2, \ldots, x_n)$, $x_{(n)} = max\ (x_1, x_2, \ldots, x_n)$, содержащий все элементы выборки, разбивают на k равных интервалов Δ_i (обычно $5 \leq k \leq 15$):



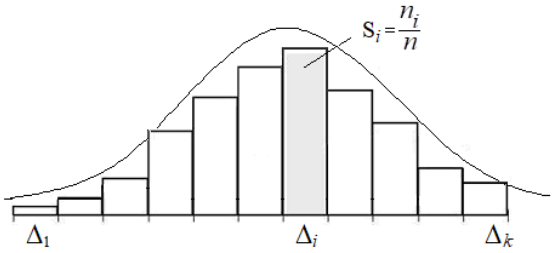
$$\alpha_0 = x_{(1)}\ ,\ \alpha_k = x_{(n)}\ ,\ \ \Delta_i = (\alpha_{i-1}; \alpha_i)\ \ (i = 1, \ldots, k)\ ;$$

$$|\Delta_i| = \frac{x_{(n)} - x_{(1)}}{k} = h\ -\ \text{шаг разбиения.}$$

Число n_i – частота, $\frac{n_i}{n}$ – относительная частота числа элементов реализации выборки, попавших в i-й интервал $(\sum_{i=1}^{k} n_i = n,\ \sum_{i=1}^{k} \frac{n_i}{n} = 1)$.

Группированный статистический ряд – это совокупность интервалов $\Delta_1, \ldots, \Delta_k$ и соответствующих им частот $n_1, \ldots, n_k$ (или относительных частот $\frac{n_1}{n}, \ldots, \frac{n_k}{n}$).

Наглядное графическое представление группированного статистического ряда дает гистограмма. Гистограммой называют ступенчатую фигуру, построенную следующим образом: на каждом интервале $\Delta_i = (\alpha_{i-1};\ \alpha_i)$ строят прямоугольник, площадь S_i которого равна относительной частоте $\frac{n_i}{n}$ числа элементов выборки, попавших в интервал Δ_i (основание прямоугольника и его высота равны, соответственно, $|\Delta_i| = h$ и $\frac{n_i}{nh}$).



Относительная частота события по вероятности сходится к вероятности этого события, поэтому если длина интервалов разбиения h достаточно мала, то $\forall x \in \Delta_i\ \frac{n_i}{n} \cong f_X(x) \cdot h$. При больших n верхний контур гистограммы (ступенчатый график) служит приближением графика плотности вероятности $f_X(x)$ (вообще говоря, неизвестной). Таким образом, разумно построенная гистограмма позволяет выдвинуть гипотезу о виде распределения исследуемой случайной величины X. Заметим, что слишком малое или слишком большое число интервалов разбиения k при построении гистограммы может привести к ее недостаточной информативности.

Число интервалов k при разбиении отрезка $[x_{(1)}; x_{(n)}]$ обычно определяют по формуле $k = 1 + 3{,}32 \lg n$ (формула Старджесса), либо по формуле $k = 1{,}72\, n^{1/3}$.

2.5. Определение и свойства точечных оценок параметров распределения: состоятельность, несмещенность, эффективность

Пусть θ – некоторый параметр распределения $F_X(x,\theta)$. Информация, необходимая для нахождения оценки $\hat{\theta}$ неизвестного параметра θ, содержится в выборке $X_1, \ldots, X_n$ из данного распределения. Таким образом, возникает задача построения оценки $\hat{\theta}$ параметра распределения как функции случайной выборки

$$\hat{\theta} = \hat{\theta}\,(X_1, \ldots, X_n).$$

Заметим, что оценка параметра распределения является случайной величиной (статистикой). В результате проведения эксперимента (серии n независимых наблюдений) получают реализацию выборки – числа $x_1, x_2, \ldots, x_n$. При этом оценка $\hat{\theta}$ принимает соответствующее числовое значение $\hat{\theta}_e = \hat{\theta}(x_1, x_2, \ldots, x_n)$, которое является приближенным значением неизвестного параметра θ. Оценки указанного типа называют точечными, их применение целесообразно при достаточно больших выборках. При малых объемах выборок используют интервальные оценки, которые будут рассмотрены далее.

Ниже приведены определения некоторых свойств точечных оценок: *состоятельность*, *несмещенность*, *эффективность*. Каждое из этих свойств определенным образом характеризует меру близости оценки $\hat{\theta}$ (случайной величины) к истинному значению (неслучайной величине) неизвестного параметра θ распределения $F_X(x,\theta)$.

def| *Состоятельность.* Оценка $\hat{\theta}$ называется состоятельной оценкой параметра θ, если оценка $\hat{\theta}$ по вероятности сходится к оцениваемому параметру: $\forall \varepsilon > 0 \;\; P\,(|\hat{\theta} - \theta| \geq \varepsilon) \underset{n \to +\infty}{\to} 0$ (символически это записывают так: $\hat{\theta} \underset{n \to +\infty}{\overset{P}{\to}} \theta$).

Иными словами, состоятельная оценка обладает свойством: с увеличением объема выборки n уменьшается вероятность того, что абсолютная величина отклонения оценки от оцениваемого параметра θ превзойдет любое наперед заданное $\varepsilon > 0$.

def | *Несмещенность.* Оценка $\hat{\theta}$ называется несмещенной оценкой параметра θ, если ее математическое ожидание равно оцениваемому параметру: $M\hat{\theta} = \theta$.

Если $M\hat{\theta} \neq \theta$, то имеет место систематическая ошибка, величину $|M\hat{\theta} - \theta|$ называют смещением.

Оценка $\hat{\theta}$ называется асимптотически несмещенной, если

$$M\hat{\theta} \underset{n\to+\infty}{\to} \theta$$

В качестве упражнения докажем следующее *утверждение*:

Пусть $\hat{\theta}$ – несмещенная оценка параметра распределения θ, причем $D\hat{\theta} \underset{n\to+\infty}{\to} 0$, тогда $\hat{\theta}$ – состоятельна.

Запишем неравенство Чебышева: $\forall \varepsilon > 0 \quad P(|\hat{\theta} - M\hat{\theta}| \geq \varepsilon) \leq \frac{D\hat{\theta}}{\varepsilon^2}$.

Учтем, что по условию $\hat{\theta}$ – несмещенная оценка ($M\hat{\theta} = \theta$), тогда, переходя к пределу при $n \to +\infty$, по теореме о сжатой переменной имеем: $\forall \varepsilon > 0 \; P(|\hat{\theta} - \theta| \geq \varepsilon) \underset{n\to+\infty}{\to} 0$, символически $\hat{\theta} \underset{n\to+\infty}{\overset{P}{\to}} \theta \;\Leftrightarrow$

$\Leftrightarrow$ $\hat{\theta}$ – состоятельная оценка.

def | *Эффективность.* Оценку $\hat{\theta}$ называют эффективной оценкой параметра θ в классе оценок, имеющих дисперсию, если дисперсия $D\hat{\theta}$ минимальна. Из двух оценок, принадлежащих данному классу, более эффективна та, которая имеет меньшую дисперсию.

Рассмотрим в качестве примера свойства относительной частоты как оценки вероятности в схеме Бернулли.

Пусть случайная величина X подчиняется биномиальному распределению с параметрами n и p.

Относительная частота $\frac{\mathrm{X}}{n} = \hat{p}$ как оценка вероятности p обладает следующими свойствами:

1) $\hat{p}$ – состоятельная, что следует из закона больших чисел:

$$\hat{p} = \frac{\mathrm{X}}{n} \xrightarrow[n \to +\infty]{P} p .$$

2) $\hat{p}$ – несмещенная, так как:

$$\mathrm{M}(\hat{p}) = \mathrm{M}(\frac{1}{n}\mathrm{X}) = \frac{1}{n}\ \mathrm{MX} = \frac{1}{n} np = p .$$

3) Дисперсия $\hat{p}$ – бесконечно малая при $n \to +\infty$:

$$\mathrm{D}(\hat{p}) = \mathrm{D}(\frac{\mathrm{X}}{n}) = \frac{1}{n^2}\ npq = \frac{pq}{n} \xrightarrow[n \to +\infty]{} 0.$$

4) Согласно теореме Муара-Лапласа эта оценка является асимптотически нормальной: $\hat{p} = \frac{\mathrm{X}}{n} \underset{n \to +\infty}{\sim} N(p; \sqrt{\frac{pq}{n}})$.

2.6. Оценки основных числовых характеристик распределения и их свойства

Пусть $(X_1, \dots, X_n)$ – выборка из распределения $F_X(x)$ исследуемой одномерной случайной величины, $(x_1, x_2, \dots, x_n)$ – реализация выборки.

Числовая характеристика случайной величины X	Оценка числовой характеристики (функция выборки $(X_1, \dots, X_n)$)	Реализация оценки (число) – значение оценки в точке $(x_1, x_2, \dots, x_n)$
$\alpha_k = M(X^k)$ начальный момент k-го порядка	$\hat{\alpha}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ выборочный начальный момент k-го порядка	$\hat{\alpha}_{ke} = \frac{1}{n}\sum_{i=1}^{n} x_i^k = a_k$ выборочный начальный момент k-го порядка
$MX = m_X = m = \alpha_1$ математическое ожидание	$\bar{X} = \hat{m} = \frac{1}{n}\sum_{i=1}^{n} X_i = \hat{\alpha}_1$ выборочное среднее	$\bar{x} = \hat{m}_e = \frac{1}{n}\sum_{i=1}^{n} x_i = \hat{\alpha}_{1e}$ выборочное среднее
$\mu_k = M(X - MX)^k$ центральный момент k-го порядка	$\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^k$ выборочный центральный момент k-го порядка	$\hat{\mu}_{ke} = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^k = m_k$ выборочный центральный момент k-го порядка
$DX = M(X - MX)^2 =$ $= \sigma_X^2 = \mu_2$ – дисперсия $\sigma_X = \sigma$ – стандартное отклонение	$S^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^2 = \hat{\mu}_2 = \hat{\sigma}^2$ выборочная дисперсия $S^{*2} = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \bar{X})^2$ исправленная выборочная дисперсия (несмещенная)	$s^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2 = \hat{\mu}_{2e} = \hat{\sigma}_e^2$ выборочная дисперсия $s^{*2} = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2$ исправленная выборочная дисперсия

Подчеркнем еще раз, что числовая характеристика случайной величины – неслучайная величина, ее оценка (функция выборки) – случайная величина, реализация оценки (значение оценки, которое она принимает в точке $(x_1, x_2, \dots, x_n)$) – неслучайная величина (число).

Замечание. Допустим (ради простоты), что в реализации выборки все числа x_i различны $x_1 < x_2 < \dots < x_n$, тогда ряд распределения эмпирической случайной величины X^*:

X^*	x_1	x_2	$\cdots$	x_n
$P(X^* = x_i)$	$1/n$	$1/n$	$\dots$	$1/n$

Выборочный начальный момент k-го порядка случайной величины X^*

$M((X^*)^k) = \frac{1}{n}\sum_{i=1}^{n} x_i^k = a_k = \hat{\alpha}_{ke}$, в частности, $M(X^*) = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Следующие теоремы раскрывают свойства указанных оценок.

Т| Выборочный начальный момент k-го порядка $\hat{\alpha}_k$ ($\hat{\alpha}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$) является несмещенной, состоятельной, асимптотически нормальной оценкой начального момента k-го порядка α_k ($\alpha_k = M(X^k)$) случайной величины X (при условии, что существуют конечные α_k и α_{2k}).

Док| Из определения выборки (см. п. 2.1.) следует:

$$MX_i = MX \text{ и } M(X_i^k) = M(X^k) \ (i = 1, \ldots, n).$$

Имеем: $M(\hat{\alpha}_k) = M(\frac{1}{n}\sum_{i=1}^{n} X_i^k) = \frac{1}{n}(nM(X^k)) = \alpha_k$ – *несмещенность*.

Далее, $D(\hat{\alpha}_k) = D(\frac{1}{n}\sum_{i=1}^{n} X_i^k) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i^k) = \frac{1}{n^2} nD(X_i^k) =$

$$= \frac{1}{n}(M(X^{2k}) - (M(X^k))^2) = \frac{\alpha_{2k} - \alpha_k^{\,2}}{n} \underset{n \to +\infty}{\to} 0.$$

Таким образом $\hat{\alpha}_k$ – несмещенная оценка, $D(\hat{\alpha}_k)$ бесконечно малая при $n \to +\infty$, откуда следует *состоятельность* $\hat{\alpha}_k$ (см. п. 2.5. утверждение о несмещенной оценке, дисперсия которой – бесконечно малая при $n \to +\infty$): $\hat{\alpha}_k \underset{n \to +\infty}{\xrightarrow{P}} \alpha_k$.

Заметим, что выборочные центральные моменты $\hat{\mu}_k$ – также являются состоятельными оценками центральных моментов μ_k случайной величины X.

Докажем, что $\hat{\alpha}_k$ – асимптотически нормальная оценка начального момента α_k.

Действительно, выборочный начальный момент $\hat{\alpha}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ – сумма одинаково распределенных независимых случайных величин, имеющих конечное математическое ожидание и дисперсию.

Согласно центральной предельной теореме:

$$\frac{\hat{\alpha}_k - M(\hat{\alpha}_k)}{\sqrt{D(\hat{\alpha}_k)}} = \frac{\hat{\alpha}_k - \alpha_k}{\sqrt{\frac{\alpha_{2k} - \alpha_k^2}{n}}} \underset{n\to+\infty}{\sim} N(0;1).$$

Таким образом, $\hat{\alpha}_k$ – асимптотически нормальная случайная величина с математическим ожиданием и стандартным отклонением, равными, соответственно: α_k и $\sqrt{\frac{\alpha_{2k} - \alpha_k^2}{n}}$. Символически это можно записать так: $\hat{\alpha}_k \underset{n\to+\infty}{\sim} N(\alpha_k; \sqrt{\frac{\alpha_{2k} - \alpha_k^2}{n}})$.

Следствие Выборочное среднее $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \hat{\alpha}_1$ является несмещенной, состоятельной, асимптотически нормальной оценкой математического ожидания MX случайной величины X.

Т| Выборочная дисперсия $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ – состоятельная, асимптотически несмещенная оценка дисперсии DX случайной величины X.

Док| Запишем следующее очевидное равенство: $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 =$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i^2 - 2X_i\bar{X} + \bar{X}^2) = \frac{1}{n}\sum_{i=1}^{n}X_i^2 - \bar{X}^2 = \hat{\alpha}_2 - \hat{\alpha}_1^2.$$

В предыдущей теореме доказана состоятельность оценки $\hat{\alpha}_k$ k-го начального момента α_k:

$$\hat{\alpha}_k \underset{n\to+\infty}{\xrightarrow{P}} \alpha_k.$$

Известные теоремы о пределе суммы и произведения функций справедливы и для сходимости по вероятности, поэтому имеем:

$$S^2 = (\hat{\alpha}_2 - \hat{\alpha}_1^2) \xrightarrow[n\to+\infty]{P} (\alpha_2 - \alpha_1^2) = \mathrm{M}(\mathrm{X}^2) - (\mathrm{MX})^2 = \mathrm{DX}.$$

Таким образом, S^2 – состоятельная оценка дисперсии DX:

$$S^2 \xrightarrow[n\to+\infty]{P} \mathrm{DX} = \sigma_{\mathrm{X}}^2.$$

Докажем теперь, что выборочная дисперсия S^2 является *асимптотически несмещенной* оценкой дисперсии DX.

Поскольку $S^2 = \hat{\alpha}_2 - \hat{\alpha}_1^2$, имеем: $\mathrm{M}(S^2) = \mathrm{M}(\hat{\alpha}_2) - \mathrm{M}(\hat{\alpha}_1^2)$.

Выборочные начальные моменты $\hat{\alpha}_k$ – несмещенные оценки соответствующих начальных моментов α_k, поэтому

$$\mathrm{M}(\hat{\alpha}_2) = \alpha_2 = \mathrm{M}(\mathrm{X}^2).$$

Далее запишем:

$$\mathrm{M}(\hat{\alpha}_1^2) = \mathrm{M}(\bar{\mathrm{X}}^2) = \mathrm{D}\bar{\mathrm{X}} + (\mathrm{M}\bar{\mathrm{X}})^2 = \frac{1}{n}\mathrm{DX} + (\mathrm{M}\bar{\mathrm{X}})^2 = \frac{1}{n}\mathrm{DX} + (\mathrm{MX})^2$$

(здесь использованы очевидные равенства $\mathrm{M}\bar{\mathrm{X}} = \mathrm{MX}$, $\mathrm{D}\bar{\mathrm{X}} = \frac{1}{n}\mathrm{DX}$).

В итоге получаем:

$$\mathrm{M}S^2 = \mathrm{M}\hat{\alpha}_2 - \mathrm{M}(\hat{\alpha}_1^2) = \mathrm{M}(\mathrm{X}^2) - \frac{1}{n}\mathrm{DX} - (\mathrm{MX})^2 = \mathrm{DX} - \frac{1}{n}\mathrm{DX} = \frac{n-1}{n}\mathrm{DX}.$$

Выборочная дисперсия S^2 не является несмещенной, однако эта оценка – *асимптотически* несмещенная, поскольку

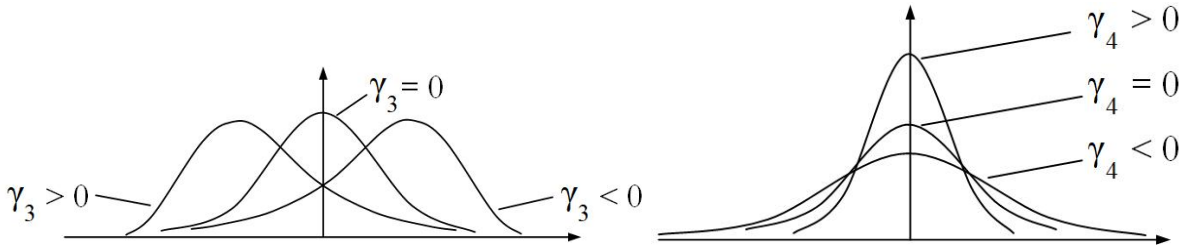
$$\mathrm{M}S^2 = \frac{n-1}{n}\mathrm{DX} \xrightarrow[n\to+\infty]{} \mathrm{DX}.$$

Наряду с выборочной дисперсией S^2 в качестве оценки дисперсии DX используют также *исправленную* выборочную дисперсию

$S^{*2} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathrm{X}_i - \bar{\mathrm{X}})^2 = \frac{n}{n-1}\ S^2$ – несмещенную состоятельную оценку дисперсии DX.

Действительно: $\mathrm{M}S^{*2} = \frac{n}{n-1}(\frac{n-1}{n}\mathrm{DX}) = \mathrm{DX}$ и $S^{*2} = (\frac{n}{n-1}\ S^2)\ \underset{n\to+\infty}{\xrightarrow{P}} \mathrm{DX}$.

Замечание



def | *Коэффициент асимметрии (асимметрия)*

$\gamma_3 = \frac{\mu_3}{\sigma^3}$, где $\mu_3 = \mathrm{M}(\mathrm{X} - \mathrm{MX})^3 = \int_{-\infty}^{\infty}(\mathrm{X}-\mathrm{MX})^3 f_{\mathrm{X}}(x)\,dx$ – третий центральный момент. Симметричные распределения (в частности, нормальное распределение) имеют коэффициент асимметрии $\gamma_3 = 0$.

def | *Коэффициент эксцесса (эксцесс)* $\gamma_4 = \frac{\mu_4}{\sigma^4} - 3$ характеризует "островершинность" графика плотности вероятности.

Для X~$N(m; \sigma)$ $\mu_4 = \int_{-\infty}^{\infty}(\mathrm{X}-\mathrm{MX})^4 f_{\mathrm{X}}(x)\,dx = 3\sigma^4 \Rightarrow \gamma_4 = 0$.

Оценками указанных числовых характеристик служат *выборочная асимметрия* $\hat{\gamma}_3 = \hat{\mu}_3 / S^3$ и *выборочный эксцесс* $\hat{\gamma}_4 = \hat{\mu}_4 / S^4 - 3$.

Реализации этих оценок: $\hat{\gamma}_{3e} = \hat{\mu}_{3e}/s^3$, $\hat{\gamma}_{4e} = \hat{\mu}_{4e}/s^4 - 3$.

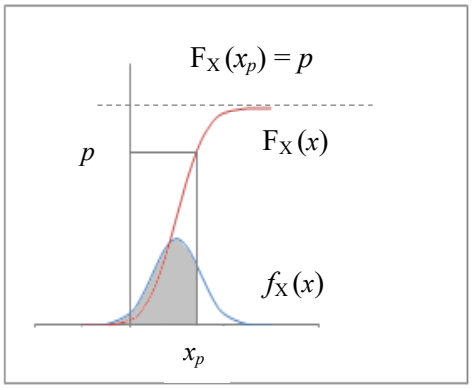
Выборочные коэффициенты эксцесса и асимметрии можно использовать для грубой проверки выборки "на нормальность",

а именно, для *отклонения* гипотезы о нормальности распределения. Если отличие от нуля эксцесса ($\hat{\gamma}_{3e}$) или асимметрии ($\hat{\gamma}_{4e}$) оказывается существенным, то гипотезу о нормальности распределения следует отвергнуть.

Заметим, что в ряде статистических модулей прикладных программ (в частности, в Excel) реализованы *несмещенные* выборочные оценки числовых характеристик распределений, в том числе, асимметрии и эксцесса.

2.7. Выборочные квантили

Напомним (п. 1.22.): *квантиль* x_p *порядка* p $(0 < p < 1)$ одномерного распределения – это корень уравнения $F_X(x) = p$, где $F_X(x)$ – функция распределения.



Порядок p квантили x_p – это *вероятность* того, что случайная величина X примет значение *левее точки* x_p:

$$F_X(x_p) = P(X < x_p) = p = \int_{-\infty}^{x_p} f_X(x)\,dx$$

Функция распределения $F_X(x)$ – неубывающая. Если X непрерывная случайная величина и $F_X(x)$ строго монотонна, то уравнение $F_X(x) = p$ имеет единственное решение x_p. Вообще говоря, это уравнение может иметь более одного решения, тогда в качестве квантили x_p берут наименьшее из них (либо их среднее арифметическое). Квантили порядка 1/4 и 3/4: $x_{1/4}$ и $x_{3/4}$ – это *нижняя и верхняя квартили*, соответственно. *Интерквантильная широта* $x_{3/4} - x_{1/4}$ может служить мерой рассеяния распределения случайной величины.

$\underline{def}$ Выборочной квантилью $Z_{n,p}$ порядка p называется следующая статистика:

$$Z_{n,p} = \begin{cases} X_{([np]+1)}, \text{ если } np \text{ - дробное } (\textit{здесь } [np] \textit{ – целая часть } np) \\ X_{(np)}, \quad \text{если } np \text{ - целое} \end{cases} .$$

Напомним: целой частью данного числа называют наибольшее целое число, не превосходящее данное; $X_{(k)}$ – k-й элемент вариационного ряда (k-я порядковая статистика).

Из определения следует, что $Z_{n,p}$ – это максимальная из порядковых статистик, обладающая свойством: *левее* нее располагаются члены вариационного ряда, доля которых $\frac{[np]}{n}$ *не превосходит* p.

Таким образом, выборочная квантиль является статистическим аналогом квантили x_p исследуемой случайной величины X.

Частные случаи.

а) Значению $p = 1/2$ отвечает выборочная медиана

$$Z_{n,1/2} = Med = \begin{cases} X_{([n/2]+1)}, \text{ если } n/2 \text{ - дробное,} \\ \dfrac{X_{(n/2)} + X_{(n/2+1)}}{2}, \quad \text{если } n/2 \text{ - целое.} \end{cases}$$

Выборочная медиана Med – оценка медианы MeX распределения случайной величины X. Реализацию *med* выборочной медианы вычисляют по реализации вариационного ряда $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$.

б) Значениям $p = 1/4$ и $p = 3/4$ отвечают выборочные *квартили* $Z_{n,1/4}$ и $Z_{n,3/4}$ (оценки нижней и верхней квартилей $x_{1/4}$ и $x_{3/4}$),

$Z_{n,1/4} = X_{(i)}$, $i = \begin{cases} [n/4]+1, \text{если } n/4 - \textit{дробное} \\ n/4, \textit{ если } n/4 - \textit{целое} \end{cases}$; $Z_{n,3/4} = X_{(j)}$, $j = n - i + 1$,

реализации $Z_{n,1/4}$ и $Z_{n,3/4}$ обозначают $z_{n,1/4}$ и $z_{n,3/4}$, соответственно.

Замечание При наличии выбросов при измерениях или в случае “зашумленных” выборок в качестве оценок математического ожидания MX и дисперсии DX симметричных распределений, целесообразным оказывается использование также оценок, перечисленных ниже.

Оценки MX (положение центра распределения):

Med – выборочная медиана, $\hat{\theta}_R = \frac{1}{2}(X_{(1)} + X_{(n)})$ – среднее арифметическое экстремальных статистик, $\hat{\theta}_Q = \frac{1}{2}(Z_{n,1/4} + Z_{n,3/4})$ – среднее арифметическое выборочных квартилей.

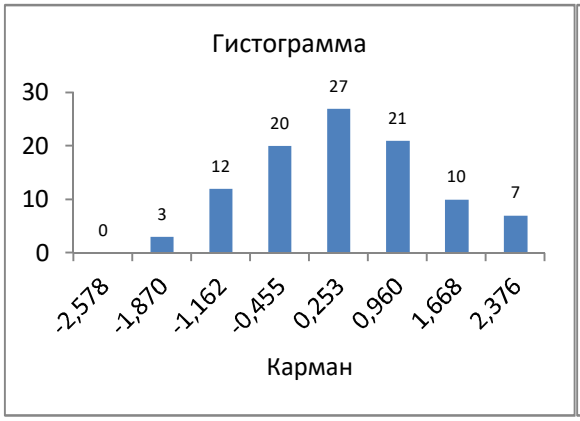
Оценки DX (мера рассеяния распределения):

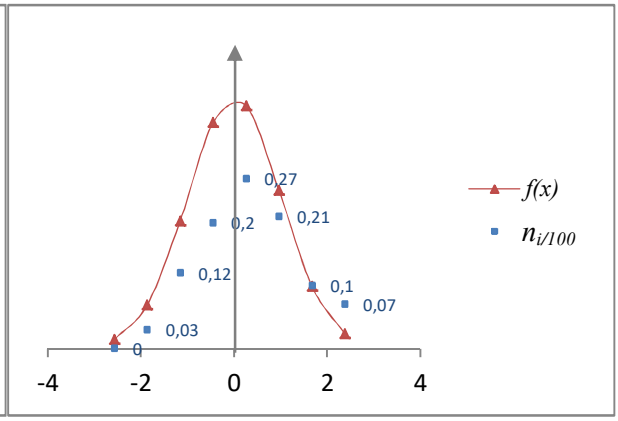
$D = \frac{1}{n}\sum_{i=1}^{n} |X_i - Med|$ – среднее абсолютное отклонение,

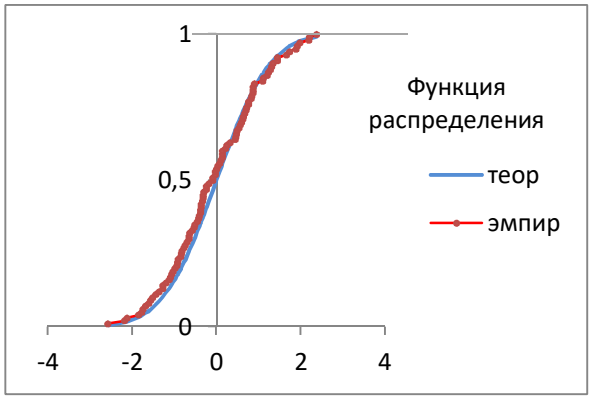
$R = X_{(n)} - X_{(1)}$ – размах, $Q = Z_{n,\,3/4} - Z_{n,\,1/4}$ – интерквартильная широта.

Пример первичной обработки данных в Excel: реализация выборки из $N(0;1)$, объем $n = 100$; $k = 7$

Описательная статистка			*Гистограмма*			
Среднее	$\overline{x}$	- 0,0405	α_i	n_i	n_i/n	X~$N(0;1)$ $f(x)$
Стандартная ошибка	$s^*/\sqrt{n}$	0,1086	-2,57758	0	0	0,014395
Медиана	med	- 0,0849	-1,86998	3	0,03	0,069436
Мода		#Н/Д	-1,16237	12	0,12	0,203012
Стандартное отклонение	s^*	1,0856	-0,45477	20	0,2	0,359751
Дисперсия выборки	s^{*2}	1,1786	0,25284	27	0,27	0,386392
Эксцесс	$\hat{\gamma}_3$	- 0,4757	0,960445	21	0,21	0,251537
Асимметричность	$\hat{\gamma}_4$	0,0907	1,66805	10	0,1	0,099248
Интервал	$x_{(100)} - x_{(1)}$	4,9532	2,375655	7	0,07	0,023735
Минимум	$x_{(1)} = \alpha_1$	- 2,5776				
Максимум	$x_{(100)} = \alpha_k$	2,3757				
Сумма	$x_1 + \ldots + x_{100}$	- 4,0485				
Счет	n	100				
Наибольший(25)	$z_{100,3/4}$	0,7576				
Наименьший(25)	$z_{100,1/4}$	- 0,8399				







2.8. Метод максимального правдоподобия

Пусть $X_1, \ldots, X_n$ – выборка из распределения $F_X(x, \theta)$, зависящего от одного неизвестного параметра θ и стоит задача построить оценку этого параметра. Один из методов нахождения оценок параметров распределений – *метод максимального правдоподобия.*

а) пусть X – непрерывная исследуемая случайная величина, $X_1, \ldots, X_n$, – выборка из распределения с плотностью вероятности $f_X(x, \theta)$, зависящей от неизвестного параметра θ, причем вид функции f известен и $x_1, x_2, \ldots, x_n$ – некоторая реализация выборки.

$\underline{def}$ Функция $L(x_1, x_2, \ldots, x_n, \theta) = f_X(x_1, \theta) \cdot \ldots \cdot f_X(x_n, \theta) = \prod_{i=1}^{n} f_X(x_i, \theta)$,

рассматриваемая в фиксированной точке $(x_1, x_2, \ldots, x_n)$ как функция параметра θ, называется функцией правдоподобия.

Вероятностный смысл функции правдоподобия $L(x_1, x_2, \ldots, x_n, \theta)$ – значение плотности вероятности n-мерной случайной величины $X_1, \ldots, X_n$, вычисленное в точке $x_1, x_2, \ldots, x_n$ (*апостериорное значение* плотности вероятности).

б) пусть теперь $x_1, x_2, \ldots, x_n$ – некоторая реализация выборки $X_1, \ldots, X_n$ из распределения дискретной случайной величины, множество возможных значений которой $\{x_i\}$ $i = 1, 2, \ldots$, причем распределение $P(X = x_i) = p_i(\theta)$ зависит от параметра θ.

Пусть в данной реализации $x_1, x_2, \ldots, x_n$ значение x_m встречается n_m раз (здесь $m = 1, 2, \ldots k$; причем $n_1 + \ldots + n_k = n$).

В случае дискретного распределения функцию правдоподобия определяют так:

$$\mathrm{L}(x_1, x_2, \ldots, x_n, \theta) = p_1^{n_1}(\theta) \ \ldots \ p_k^{n_k}(\theta) = \prod_{m=1}^{k} p_m^{n_m}(\theta).$$

Вероятностный смысл функции правдоподобия для случая дискретного распределения состоит в следующем: это вероятность того, что случайная выборка $\mathrm{X}_1, \ldots, \mathrm{X}_n$ примет значение, равное именно *данной реализации* выборки $x_1, x_2, \ldots, x_n$.

Понятно, что чем ближе значение переменной θ к истинному (неизвестному) значению параметра распределения $\mathrm{F}_{\mathrm{X}}(x, \theta)$, тем выше вероятность при проведении эксперимента получить данную реализацию выборки $x_1, x_2, \ldots, x_n$.

$\underline{def}$ *Оценкой максимального правдоподобия* $\hat{\theta}_{\mathrm{МП}}$ неизвестного параметра θ (точнее – *значением* оценки, отвечающим данной конкретной реализации выборки $x_1, x_2, \ldots, x_n$) называется такое значение переменной θ, которое доставляет максимум функции правдоподобия $\mathrm{L}(x_1, x_2, \ldots, x_n, \theta)$.

Функция правдоподобия L, определенная выше, представляет собой произведение ряда сомножителей, поэтому при поиске точки максимума L целесообразно перейти к $\ln \mathrm{L}$ (очевидно, что $\ln \mathrm{L}$ и L имеют максимум в одной и той же точке) и оценку максимального правдоподобия $\hat{\theta}_{\mathrm{МП}}$ параметра θ находить из *уравнения правдоподобия* $\dfrac{\partial}{\partial \theta}(\ln \mathrm{L}) = 0.$

В случае, когда неизвестными являются m параметров $\theta_1, \ldots, \theta_m$, оценки $\hat{\theta}_{1\mathrm{МП}}, \ldots, \hat{\theta}_{m\mathrm{МП}}$ находят из соответствующей системы m уравнений.

Заметим, что метод максимального правдоподобия всегда приводит к состоятельным оценкам, распределенным асимптотически

нормально, имеющим наименьшую возможную дисперсию среди других асимптотически нормальных оценок. Однако на практике он может осложняться трудностями, связанными с решением систем уравнений правдоподобия.

2.9. Примеры нахождения оценок максимального правдоподобия (МП- оценок) параметров распределений

Пример 1_мп (нормальное распределение).

Пусть имеется выборка $X_1, \ldots, X_n$ из нормального распределения $N(m; \sigma)$ и $x_1, x_2, \ldots, x_n$ – некоторая реализация выборки. Найдем оценки максимального правдоподобия $\hat{m}_{\text{МП}}$ и $\hat{\sigma}_{\text{МП}}$ параметров распределения m и σ.

Функция правдоподобия в точке $(x_1, x_2, \ldots, x_n)$ равна

$$L(x_1, x_2, \ldots, x_n, m, \sigma) = \prod_{i=1}^{n} f_X(x_i, m, \sigma) = \prod_{i=1}^{n} \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i - m)^2}{2\sigma^2}}\right),$$

ее логарифм $\ln L = \left(-\frac{n}{2}\right) \ln 2\pi - n \ln \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - m)^2$.

Решая систему уравнений правдоподобия

$\frac{\partial \ln L}{\partial m} = 0$, $\frac{\partial \ln L}{\partial \sigma} = 0$ относительно неизвестных m и σ,

получаем: $m = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}$; $\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - m)^2 = s^2$.

Полученное решение – значение реализаций оценок параметров m и σ, соответствующее данной реализации выборки, однако *все приведенные рассуждения справедливы для любой реализации выборки*, поэтому искомые оценки равны, соответственно:

$$\hat{m}_{\text{МП}} = \bar{X}; \ \hat{\sigma}_{\text{МП}} = S^2.$$

Пример 2_мп (распределение Пуассона)

Найдем оценку максимального правдоподобия $\hat{a}_{\text{МП}}$ параметра a распределения Пуассона $P(X = i) = p_i = \frac{a^i}{i!} e^{-a}$ $(i = 0, 1, 2, \ldots)$.

Пусть $x_1, x_2, \ldots, x_n$ – некоторая реализация выборки $X_1, \ldots, X_n$ из распределения Пуассона, так что числа $x_1, x_2, \ldots, x_n$ – целые неотрицательные. Обозначим через k наибольшее из них и подсчитаем число раз, которое каждое из чисел $0, 1, \ldots, k$ встретилось в данной реализации выборки:

$$0 - n_0 \text{ раз}, \; 1 - n_1 \text{ раз}, \ldots, \; m - n_m \text{ раз}, \ldots, \; k - n_k \text{ раз},$$

$$\text{при этом } \sum_{m=0}^{k} n_m = n, \quad \sum_{m=0}^{k} m n_m = \sum_{i=1}^{n} x_i .$$

$$\text{Далее, } \mathrm{L}(x_1, x_2, \ldots, x_n, a) = \prod_{m=0}^{k} p_m^{n_m}(a), \qquad \ln \mathrm{L} = \sum_{m=0}^{k} n_m \ln p_m(a),$$

$$\ln p_m(a) = \ln\left(\frac{a^m}{m!} e^{-a}\right) = m \ln a - a - \ln(m!).$$

$$\frac{\partial \ln \mathrm{L}}{\partial a} = \sum_{m=0}^{k} n_m \left(\frac{m}{a} - 1\right) = 0, \quad \frac{1}{a} \sum_{m=0}^{k} m n_m - \sum_{m=0}^{k} n_m = 0, \quad \frac{1}{a} \sum_{i=1}^{n} x_i = n,$$

$$\text{откуда } a = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x} .$$

Таким образом, оценка максимального правдоподобия параметра a распределения Пуассона равна $\hat{a}_{\text{МП}} = \bar{\mathrm{X}}$. Заметим, что эта оценка несмещенная состоятельная асимптотически нормальная.

Пример 3_мп (равномерное распределение)**.**

Пусть $x_1, x_2, \ldots, x_n$ – некоторая реализация выборки $X_1, \ldots, X_n$ из равномерного распределения с параметрами a и b, а $x_{(1)} \le \ldots \le x_{(n)}$ – соответствующая реализация вариационного ряда. Найдем оценки максимального правдоподобия параметров a и b.

Функция правдоподобия в точке $(x_1, x_2, \ldots, x_n)$ равна:

$$\mathrm{L}(x_1, x_2,\ldots,x_n; a, b) = \left(\frac{1}{b-a}\right)^n, \quad \text{если } \forall i \ x_i \in [a; b]$$

Если хотя бы одно число из $x_1, x_2,\ldots,x_n$ лежит вне $[a; b]$, то $\mathrm{L}(x_1, x_2,\ldots,x_n; a, b) = 0$.

Ясно, что функция правдоподобия $\mathrm{L}(x_1, x_2,\ldots,x_n; a, b) = \left(\frac{1}{b-a}\right)^n$ максимальна (как функция параметров a и b) при условии, что величина разности $b-a$ минимальна. Таким образом, поскольку $a \le x_{(1)}$, $b \ge x_{(n)}$, то значения параметров, доставляющие максимум функции правдоподобия $a = x_{(1)}$, $b = x_{(n)}$, а искомые оценки

$$\hat{a}_{\text{МП}} = \mathrm{X}_{(1)}, \ \hat{b}_{\text{МП}} = \mathrm{X}_{(n)}.$$

Пример 4_мп (распределение Бернулли)

Найдем оценку максимального правдоподобия вероятности p (вероятности успеха) в каждом испытании при проведении n независимых испытаний по схеме Бернулли. Индикатор X_i появления успеха в i-м испытании – случайная величина, принимающая два возможных значения 1 или 0, а именно, $\mathrm{P}(\mathrm{X}_i = 1) = p$, если в результате i - го испытания осуществился успех и $\mathrm{P}(\mathrm{X}_i = 0) = 1 - p = q$ если результат i - го испытании – неуспех (распределение Бернулли):

X_i	0	1
P	$q = 1 - p$	p

Пусть в результате данной серии n испытаний получена реализация выборки из распределения Бернулли, в которой значение

“1” встретилось точно m раз (m успехов в n испытаниях), а значение “0”, соответственно, $n-m$ раз. Функция правдоподобия имеет вид: $\mathrm{L}(x_1, x_2, \ldots, x_n, p) = p^m q^{n-m} = p^m (1-p)^{n-m}$.

$$\frac{\partial \ln \mathrm{L}}{\partial p} = \frac{m}{p} - \frac{n-m}{1-p} = 0, \text{ откуда } p = \frac{m}{n}.$$

Таким образом, искомой оценкой максимального правдоподобия вероятности p является относительная частота $\hat{p}_{\text{МП}} = \frac{\mathrm{X}}{n}$ числа успехов при проведении n независимых испытаний (свойства $\hat{p}_{\text{МП}}$ см. п.2.5.).