

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ»

Ю.С. Радченко, А.В. Захаров, А.В. Зюльков

МЕТОДЫ ОБРАБОТКИ И ПЛАНИРОВАНИЯ ЭКСПЕРИМЕНТА

Часть 1 ОЦЕНКА РАСПРЕДЕЛЕНИЙ И ИХ ПАРАМЕТРОВ

Учебно-методическое пособие

Воронеж
Издательский дом ВГУ
2016

ОГЛАВЛЕНИЕ

Введение.....	4
1. Методы оценок числовых характеристик распределений случайных величин	5
2. Оценки при известном (с точностью до параметров) распределении выборки	13
3. Выборочные распределения и их анализ.....	18
4. Оценки, основанные на порядковых статистиках	27
5. Робастные оценки параметров.....	32
Библиографический список	38

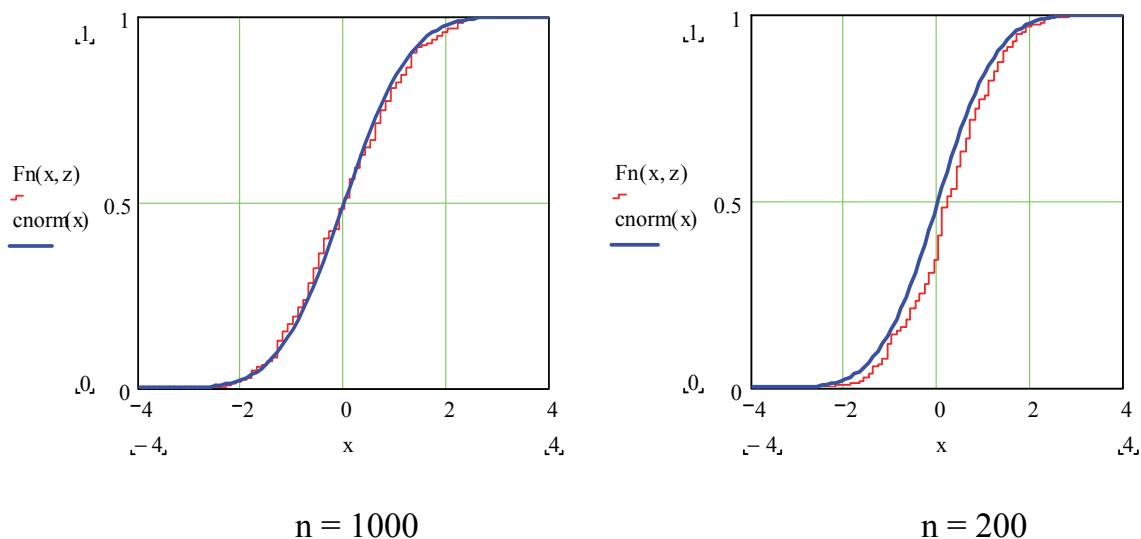


Рис. 1.1

На рис. 1.1 приведены выборочная функция распределения $F_n(x)$ (1.1) при разных n и теоретическая функция распределения.

Оценка числовых характеристик и параметров нормального распределения по независимым результатам наблюдений

Точечные оценки

Предположим, что случайная величина $X_i \sim N(m, \sigma_X^2)$ наблюдается непосредственно (без аномальных ошибок, обусловленных «засорением» выборки). Можно показать, что оценки параметров m и σ_X^2 будут определяться следующими выражениями для всех методов:

– оценка математического ожидания (выборочное среднее)

$$\tilde{m} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i ; \quad (1.3)$$

– оценка дисперсии (выборочная дисперсия):

при условии, что математическое ожидание известно,

$$\hat{\sigma}_X^2 \equiv s_X^2 = \frac{1}{n} \sum_{i=1}^n (X_i - m)^2 ; \quad (1.4)$$

при условии, что математическое ожидание неизвестно

$$\tilde{\sigma}_X^2 \equiv S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \left[\sum_{i=1}^n X_i^2 \right] - \frac{n}{n-1} \bar{X}^2. \quad (1.5)$$

Эти оценки (1.3)–(1.5) являются несмещенными и асимптотически эффективными. Их дисперсии определяются следующим образом:

$$D\{\bar{X}\} = \sigma_X^2 / n;$$

при условии, что математическое ожидание известно,

$$D\{S_X^2\} = \frac{2}{n} \sigma_X^4 \approx \frac{2}{n} S_X^4; \text{ (при } n \gg 1);$$

при условии, что математическое ожидание не известно,

$$D\{S_X^2\} = \frac{2}{n-1} \sigma_X^4 \approx \frac{2}{n-1} s_X^4 \text{ (при } n \gg 1).$$

Старшие начальные и центральные выборочные моменты определяются следующими выражениями:

$$m_k(\bar{X}) = \frac{1}{n} \sum_{i=1}^n x_i^k; \quad M_k(\bar{X}) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})^k; \quad (k = 3, 4, \dots),$$

которые дают состоятельные и эффективные оценки характеристик гауссовской случайной величины.

Для определения коэффициентов асимметрии и эксцесса можно воспользоваться следующими выражениями:

$$\hat{A}\{\bar{X}\} = \frac{1}{n(\sqrt{S_X^2})^3} \sum_{i=1}^n (X_i - m)^3;$$

$$\hat{E}\{\bar{X}\} = \frac{1}{n} \sum_{i=1}^n (X_i - m)^4 / (S_X^4) - 3.$$

Интервальные оценки параметров

Для совместного анализа точности и надежности рассмотренных оценок используют доверительные интервалы, соответствующие заданной доверительной вероятности P_0 :

$$P[d_1(\bar{X}) \leq \theta \leq d_2(\bar{X})] = P_0. \quad (1.6)$$

Пусть $\gamma(\bar{X})$ – точечная оценка. Тогда (для симметричного распределения $\gamma(\bar{X})$)

$$\begin{aligned} d_1(\bar{X}) &= \gamma(\bar{X}) - \Delta_p, \\ d_2(\bar{X}) &= \gamma(\bar{X}) + \Delta_p. \end{aligned} \quad (1.7)$$

Тогда можно записать (1.6) с учетом (1.7)

$$P[-\Delta_p \leq \gamma(\bar{X}) - \theta \leq \Delta_p] = P_0. \quad (1.8)$$

Таким образом, зная распределение случайной величины $\chi(\bar{X}) - \theta$, можно найти Δ_p .

Следует отметить, что наилучшее качество интервального оценивания гарантируется только при **гауссовском** распределении и резко падает при отклонениях от него.

Оценка математического ожидания при известной дисперсии выборки

Оценка математического ожидания, определенная в соответствии с выражением (1.3), является несмещенной и имеет гауссовское распределение $\bar{X} \sim N(m, \frac{\sigma_x^2}{n})$. Поэтому, как правило, рассматривается симметричный доверительный интервал

$$\begin{aligned} P[-\Delta_p \leq \bar{X} - m \leq \Delta_p] &= P_0 \\ P\left[-\frac{\Delta_p}{\sigma_x} \sqrt{n} \leq \frac{\bar{X} - m}{\sigma_x} \sqrt{n} \leq \frac{\Delta_p}{\sigma_x} \sqrt{n}\right] &= P_0. \\ P[-\delta \leq \eta \leq \delta] &= P_0 \end{aligned} \quad (1.9)$$

При заданной дисперсии σ_x^2 статистика $\eta = \frac{\bar{X} - m}{\sigma_x} \sqrt{n} \sim N(0,1)$ имеет стандартное гауссовское распределение $\delta = \Delta_p \sqrt{n} / \sigma_x$. Откуда:

$$\frac{1}{\sqrt{2\pi}} \int_{-\delta}^{\delta} \exp\left(-\frac{x^2}{2}\right) dx = \Phi(\delta) - \Phi(-\delta) = P_0.$$

Учитывая, что $\Phi(-\delta) = 1 - \Phi(\delta)$, получаем

$$2\Phi(\delta) - 1 = P_0 \text{ или } \delta = \Phi^{-1}\left(\frac{1 + P_0}{2}\right).$$

Обозначим λ_q – квантиль порядка q распределения $N(0,1)$, где $q = (1 + P_0) / 2$. Чаще всего берут $P_0 = 0.9$, следовательно $\lambda_p = 1.645$

$$d_1 = \bar{X} - \sigma_x \lambda_{(1+p)/2} / \sqrt{n}; \quad (1.10)$$

$$d_2 = \bar{X} + \sigma_x \lambda_{(1+p)/2} / \sqrt{n}. \quad (1.11)$$

Очевидно, что $d_2 - d_1 = 2\Delta_p$. Тогда из полученных соотношений (1.10)

$$\frac{2\Delta_p}{\sigma_x} = \frac{\lambda_{(1+p)/2}}{\sqrt{n}}. \quad (1.12)$$

На рис. 1.2 приведен график коридора $v(n) = \Delta / \sigma_x = \frac{\lambda_{(1+p)/2}}{\sqrt{n}}$.

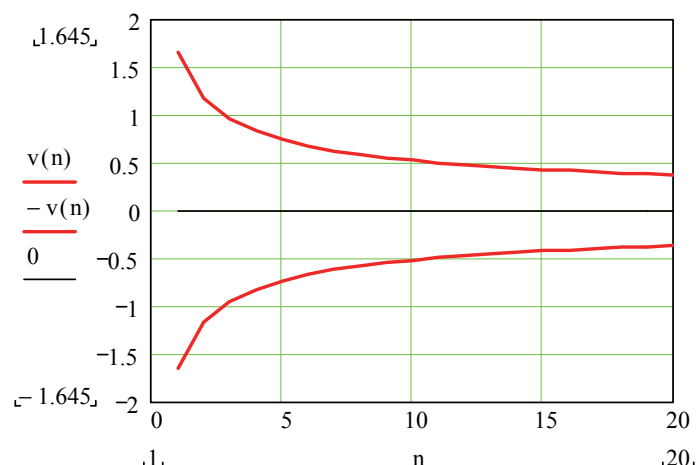


Рис. 1.2

Как видно из графика на рис. 1.2, уменьшение доверительного интервала с увеличением n замедляется. Так уменьшение доверительного интервала в 10 раз требует увеличения выборки в 100 раз.

Оценка математического ожидания при неизвестной дисперсии выборки

В таком случае вместо σ_x^2 берут ее оценку $S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$.

Если дисперсия неизвестна, статистика $\xi = \frac{\bar{X} - m}{S_X} \sqrt{n}$ имеет t – распределение **Стьюдента** с $\nu = n - 1$ степенями свободы. Границы доверительных интервалов получаются следующие:

$$d_1 = \bar{X} - S_X t_{n-1, (1+P)/2} / \sqrt{n}; \quad (1.13)$$

$$d_2 = \bar{X} + S_X t_{n-1, (1+P)/2} / \sqrt{n}; \quad (1.14)$$

Следует отметить, что при достаточно больших n ($n \geq 60 \dots 100$) $t_{n-1, q} \approx \lambda_q$ и вместо выражений (1.13)–(1.14) могут быть использованы (1.10)–(1.11), в которые вместо истинных значений σ_x подставляются их оценки S_X .

Расчет длины выборки n

Если **задана** ширина двустороннего интервала $2\Delta_p$, в котором с вероятностью P_0 должно содержаться истинное значение математического ожидания m , диапазон отклонений $\Delta \sim |\bar{X} - m|$, то из выражений (1.12), (1.10), (1.11) можно определить n :

– при условии известной дисперсии σ_x^2 для оценки m

$$n \geq \left(\lambda_p \sigma_x^2 / \Delta_p \right)^2 = \lambda_p \left(\sigma_x^2 / \Delta_p \right)^2 ; \quad (1.15)$$

– при условии неизвестной дисперсии σ_x^2

$$n \geq s_{\bar{X}}^2 \left(t_{n-1, (1+P)/2} / \Delta_p \right)^2 . \quad (1.16)$$

Расчет допустимой погрешности измерения σ_x

Заданы: Пусть задана доверительная вероятность P_0 ($P_0 = 0.9$), объем выборки n , диапазон отклонений $\Delta \sim |\bar{X} - m|$. Тогда из (1.12) получаем расчетную формулу

$$\sigma_x = \frac{\Delta \sqrt{n}}{\lambda_q} .$$

Расчет доверительной области

Заданы: Задана доверительная вероятность P_0 ($P_0 = 0.9$), объем выборки n , дисперсия ошибки единичного измерения. Тогда из (1.12) получаем

$$\Delta = \frac{\sigma_x \lambda_{(1+P)/2}}{\sqrt{n}} .$$

В таком случае диапазон нахождения параметра $m = \bar{X} \pm \Delta$.

Оценка числовых характеристик случайных величин по сгруппированным данным

Пусть исходная выборка $\vec{X} = \{X_i\}$ сгруппирована в M интервалов длиной h , \bar{x}_j – центр j -го интервала. Соответствующие частоты (оценки вероятностей попадания в i подинтервал) $\{p_j\}$ ($j = \overline{1, M}, M \ll n$).

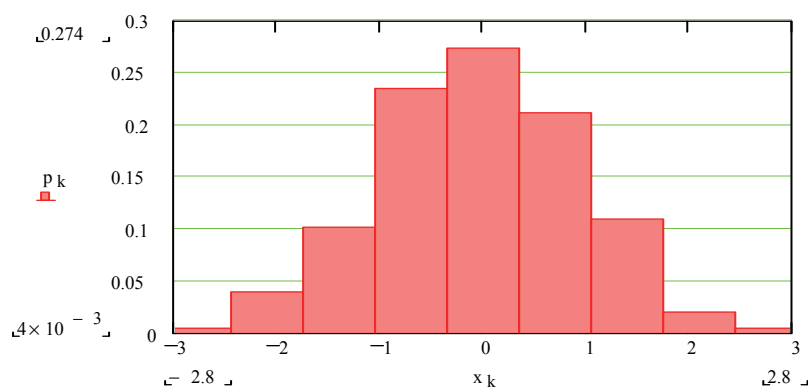


Рис. 1.3