

9. Обработка экспериментальных данных

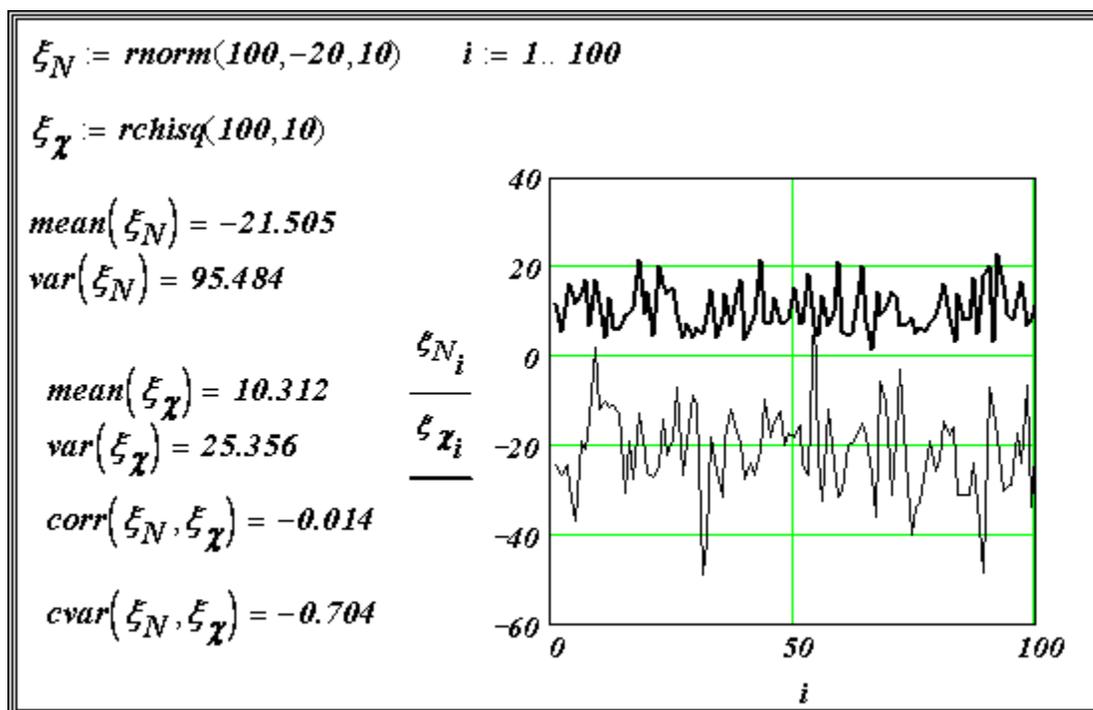
Моделирование псевдослучайных чисел

Функция $rnd(x)$ генерирует одно случайное число, равномерно распределенное в интервале $[0, x]$.

Функции MathCAD генерирования случайных векторов. m - число проекций случайного вектора

<p><i>Равномерное распределение</i></p> $\frac{1}{b-a}, \text{ если } x \in [a, b]$ $0, \text{ если } x \notin [a, b]$	$r\text{unif}(m, a, b)$
<p><i>Нормальное распределение</i></p> $\frac{1}{\sqrt{2\pi} \cdot \sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \sigma > 0$	$r\text{norm}(m, \mu, \sigma)$
<p>χ^2 - распределение</p> $0, \text{ если } x \leq 0;$ $\frac{1}{2^{n/2} \Gamma(n/2)} e^{-\frac{x}{2}} x^{n/2-1}, \text{ если } x > 0$ <p>($n > 0$ - число степеней свободы)</p>	$r\text{chisq}(m, n)$

Пример. Необходимо сгенерировать два случайных вектора: ξ_N - проекции имеют нормальное распределение (математическое ожидание равно -20 , дисперсия 100); ξ_x - проекции имеют χ^2 - распределение (с числом степеней свободы 10). Размерность векторов равна 100 .



Функции MathCAD вычисления выборочных значений числовых характеристик. К числовым характеристикам случайной величины относятся: математическое ожидание (или среднее), дисперсия, среднее квадратическое отклонение и т.д. Часто возникает необходимость оценить эти характеристики по выборке значений случайной величины объема m . Такие оценки называют выборочными значениями числовых характеристик.

В таблице приведены имена функций, вычисляющих выборочное значение часто используемых числовых характеристик. Здесь X, Y – векторы размерности m , составленные из значений случайной величины X и Y .

Числовые характеристики	Функция MathCAD
Математическое ожидание случайной величины X	$mean(X)$

Дисперсия случайной величины X	$var(X)$
Среднеквадратическое отклонение случайной величины X	$side(X)$
Медиана случайной величины X	$median(X)$
Мода случайной величины X	$mode(X)$
Корреляционный момент двух случайных величин X, Y	$cvar(X, Y)$
Коэффициент корреляции двух случайных величин X, Y	$ccor(X, Y)$

Пример. На приведенном выше рисунке показан фрагмент документа MathCAD, в котором вычисляются выборочные значения некоторых числовых характеристик.

Задание. Сгенерируйте случайный вектор размерности 200, проекции которого равномерно распределены в интервале $[0,1]$. Вычислите его числовые характеристики.

Увеличить размерность до 1000 и снова вычислить его числовые характеристики. Сравнить выборочные числовые характеристики с теоретическими.

Функции MathCAD вычисления частот значений случайной величины (построение гистограмм). Введём некоторые определения.

Предположим, что дана выборка $\{x_i\}, i = \overline{1, N}$ случайной величины X (N – объём выборки). Введём $L+1$ точку

$$z_1 < z_2 < \dots < z_L < z_{L+1}$$

при этом:

$$z_1 \leq \min\{x_i\}; z_{L+1} \geq \max\{x_i\}$$

Тогда число значений x_i , попавших в интервал $[z_k, z_{k+1}]$, $k = 1, \dots, L$, обозначим через n_k и назовём частотой. Очевидно, что

$$\sum_{k=1}^L n_k = N$$

Величину

$$w_k = \frac{n_k}{N}$$

назовём относительной частотой, для которой выполняется условие

$$\sum_{k=1}^L w_k = 1$$

В качестве *оценки плотности распределения вероятности* непрерывной случайной величины X используют гистограмму относительных частот, т.е. систему прямоугольников, k -й из которых основанием имеет $z_{k+1} - z_k$, а высота p_k^* определяется по формуле

$$p_k^* = \frac{w_k}{z_{k+1} - z_k}, \quad k = 1, \dots, L$$

и имеет место приближенное тождество

$$p_k^* \cong p(x^*),$$

где x^* — некоторое число из интервала $[z_k, z_{k+1}]$.

Возникает вопрос: как сформировать интервалы $[z_k, z_{k+1}]$? Количество интервалов L рекомендуется вычислять по формуле

$$L = \lceil 1 + 3.222 \lg(N) \rceil + 1,$$

где $[Q]$ – целая часть числа Q .

Значения w_k , P_k^* вычисляются по частотам n_k . Поэтому для определения n_k по выборке $\{x_i\}$ в MathCAD включены две функции:

$hist(int, X)$, $histogram(int, X)$.

Параметры функции $hist(int, X)$:

- int – массив длины $(L+1)$, составленный из значений z_k , $k = 1, \dots, L+1$. Если параметр int задать целым числом, равным числу интервалов L , то при выполнении функции формируется рабочий массив узлов $\{z_k\}$;

- X – массив длиной N , составленный из значений выборки $\{x_i\}$.

Результатом работы функции является одномерный массив $\{n_k\}$, $k = 1, \dots, L$.

Параметры функции $histogram(int, X)$:

- int – массив длины $(L+1)$, составленный из значений z_k , $k = 1, \dots, L+1$. Если int задать целым числом, равным числу интервалов L , то при выполнении функции формируется рабочий массив узлов $\{z_k\}$;

- X – массив длиной N , составленный из значений выборки $\{x_i\}$.

Результатом работы функций является матрица размером $L \times 2$, первый столбец содержит значения d_k (середины отрезков $[z_k, z_{k+1}]$, $k = 1, \dots, L$, а второй столбец – значения n_k .

Пример. Построить гистограммы относительных частот по выборкам случайных величин ξ_N , ξ_X , определенных в рассмотренном ранее примере. Объем выборки $N = 1000$.

На рисунке А показано построение гистограммы для случайной величины ξ_N , а на рисунке Б – для случайной величины ξ_X с использованием функции $histogram$ при $L = 11$.

Середины отрезков d_k «откладываются» по оси абсцисс, а для отображения гистограммы задаётся параметр `solidbar` (команда **Формат** контекстного меню, закладка *Метки*). Точками на рисунках показаны значения соответствующих плотностей распределений, вычисленных

при $x = d_k$.

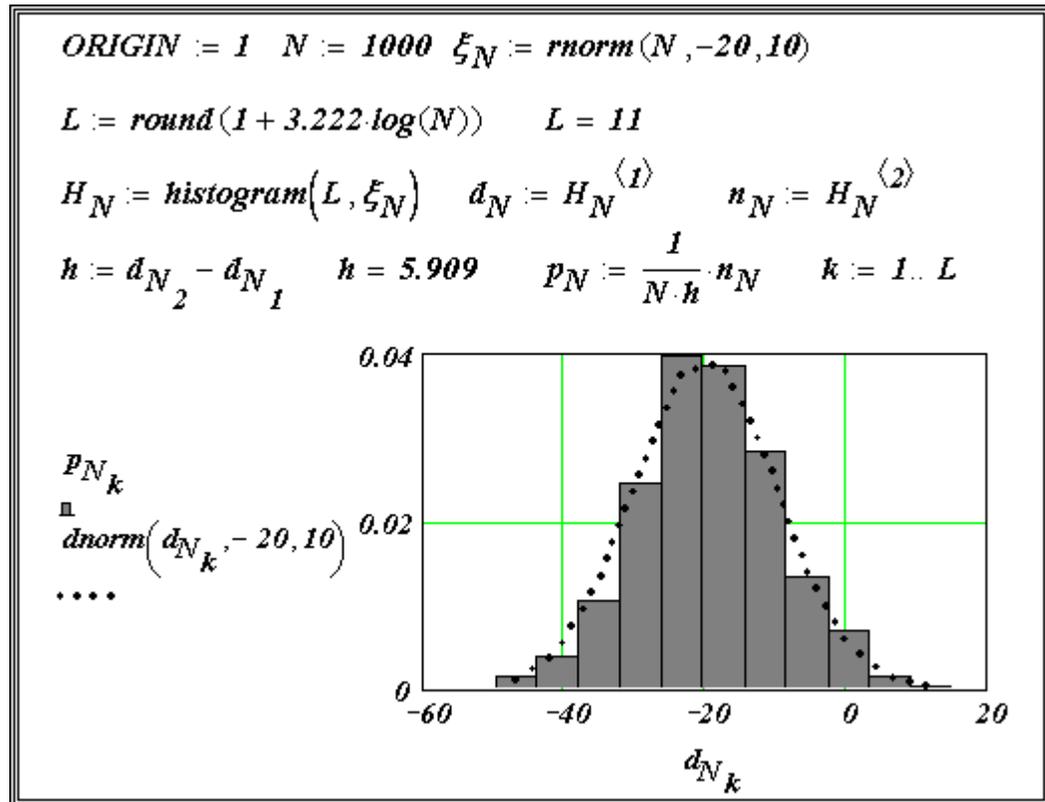


Рис. А

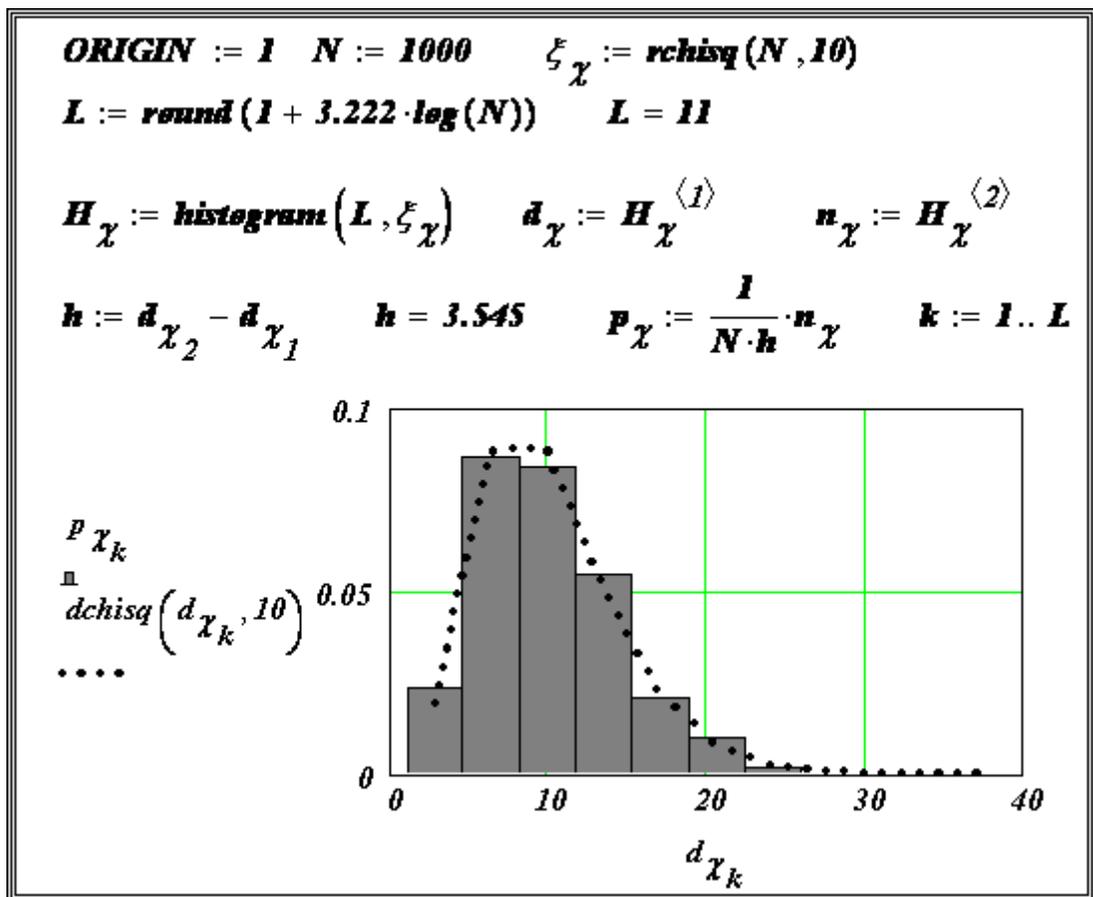


Рис. Б.

Задание. По двум выборкам равномерно распределенных случайных чисел (объемом 200 и 2000) построить гистограммы.

Сделать вывод о влиянии объема выборки на точность оценивания плотности распределения случайной величины.