Лабораторная работа №6

Обработка результатов эксперимента в MathCad

1. Законы распределения случайных чисел

- **Распределение** случайной величины – это функция, позволяющая определить вероятность появления заданного значения случайной величины.
- В теории вероятностей сформулировано несколько законов распределения как для дискретных, так и для непрерывных случайных величин.
- Наблюдаемые на практике случайные величины часто не вполне соответствуют теоретическим распределениям, но с некоторой точностью могут быть приближенно ими представлены.

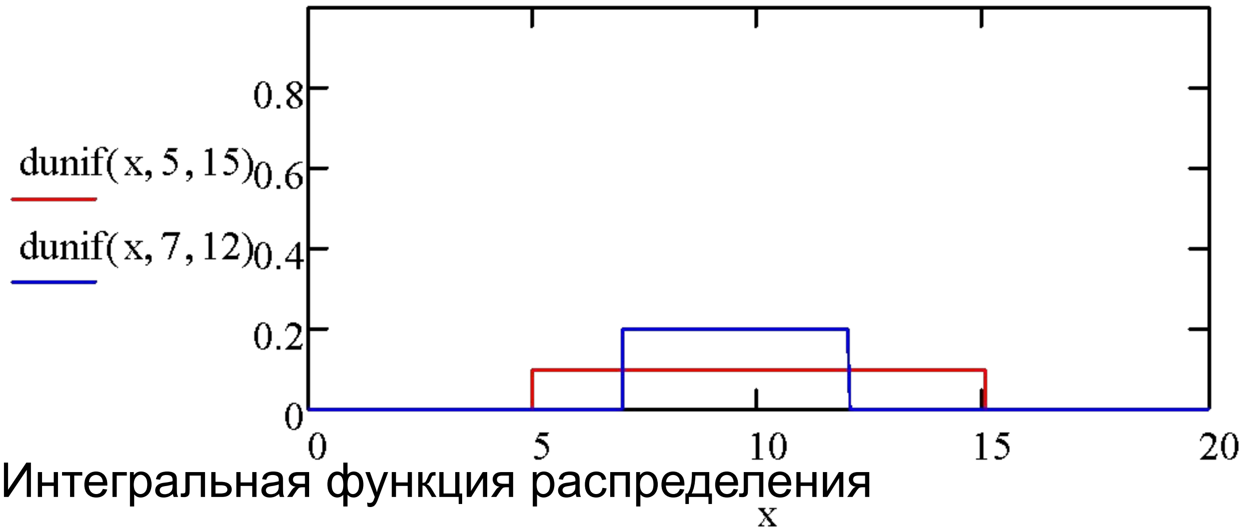
Законы распределения случайных чисел

Для **непрерывных случайных величин** рассмотрим следующие законы:

- Равномерное распределение
- Нормальное распределение
- Экспоненциальное распределение
- Гамма-распределение

Равномерное распределение

Плотность вероятности



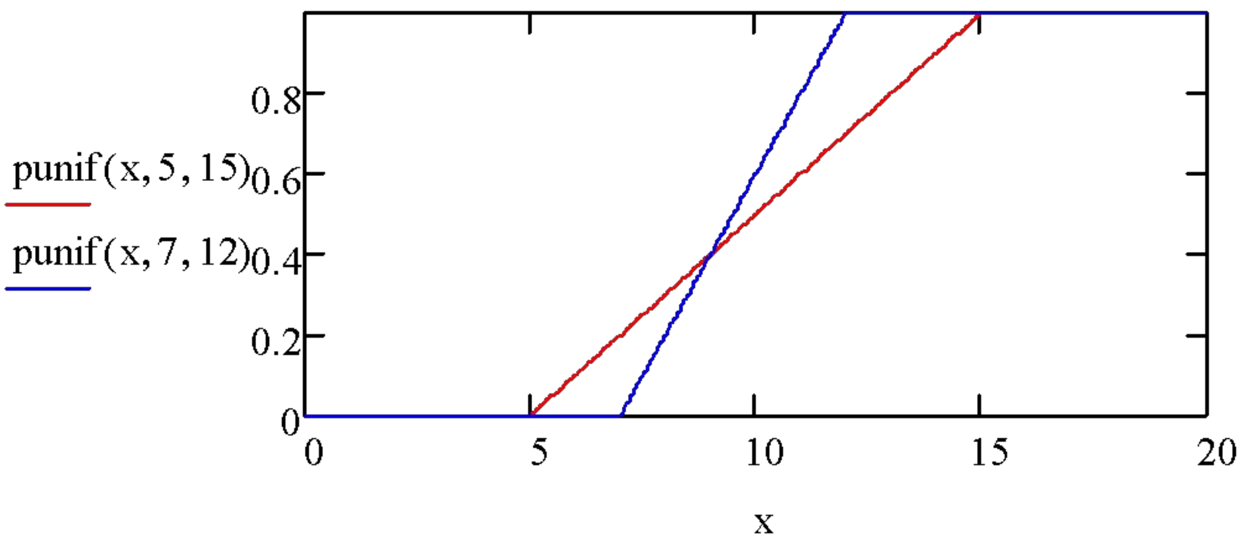
2 параметра:

a, b – границы отрезка

$a = \min(X_i)$

$b = \max(X_i)$

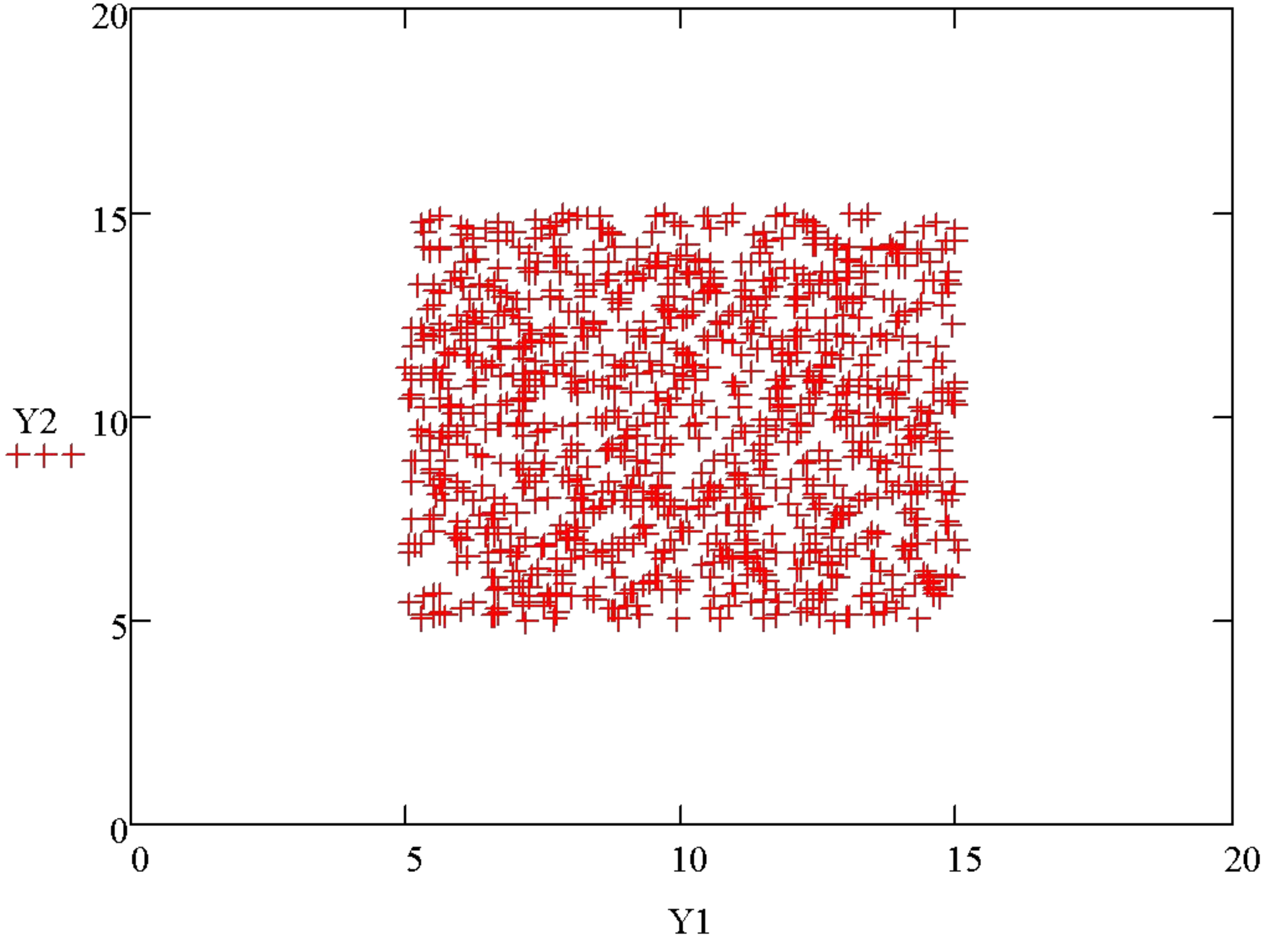
Интегральная функция распределения



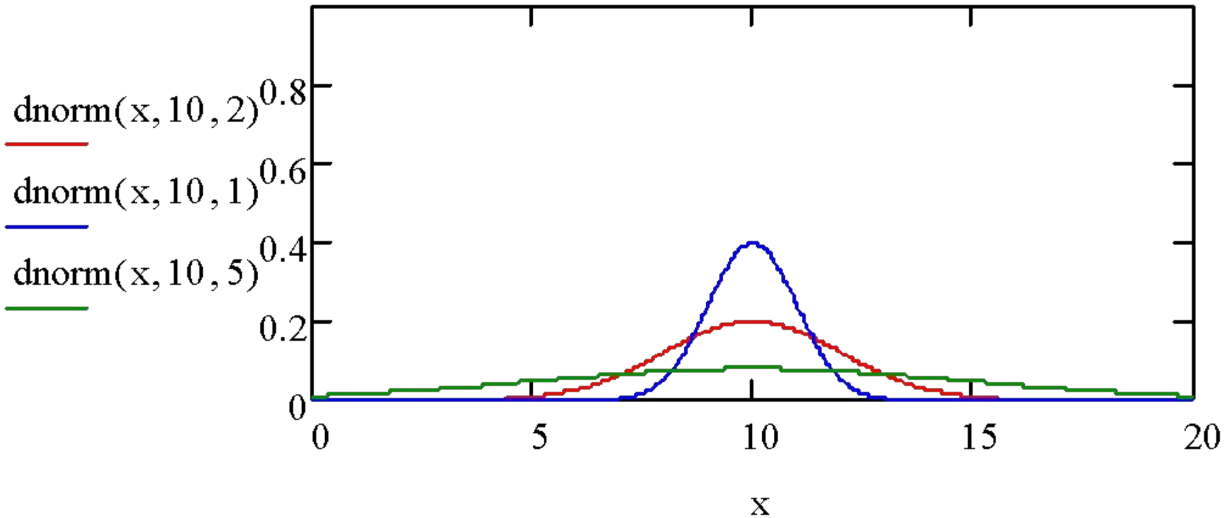
Равномерное распределение

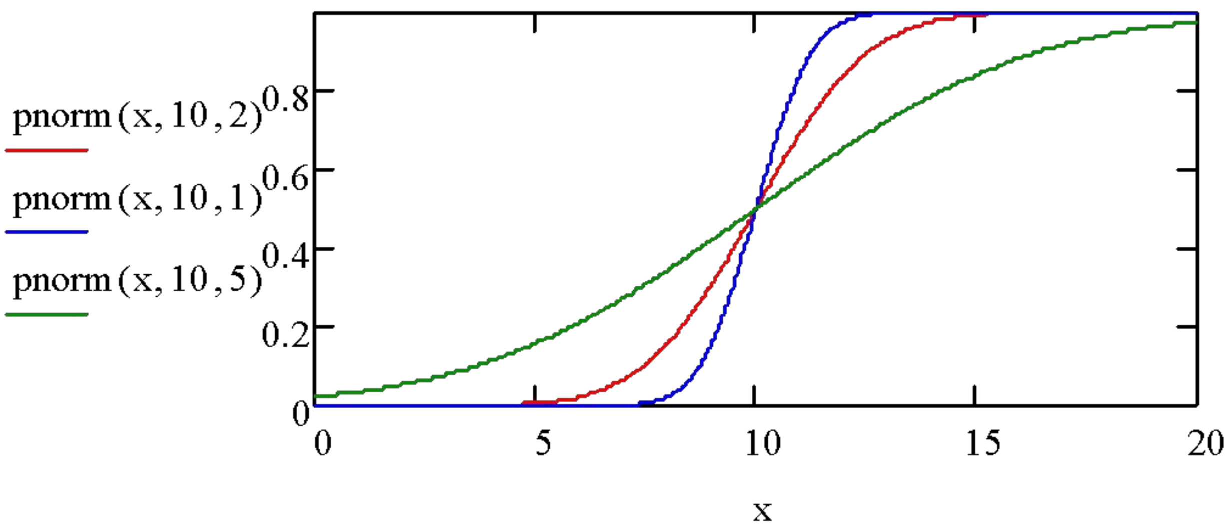
$t := 0..1000$

$Y1_t := \mathrm{runif}(1, 5, 15)_0$ $\quad Y2_t := \mathrm{runif}(1, 5, 15)_0$



Нормальное распределение





2 параметра:

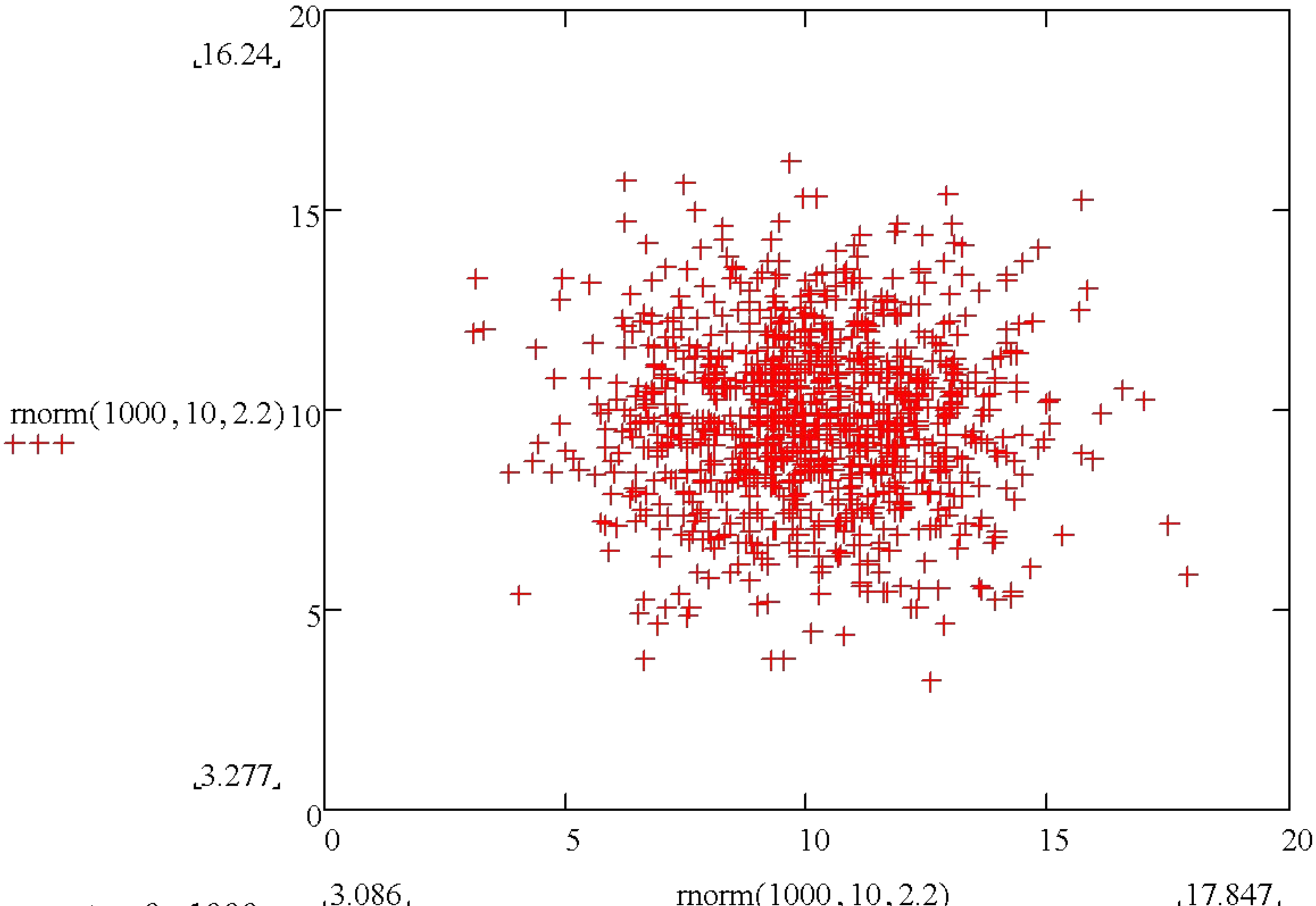
μ – мат. ожидание

σ – стандартное, или среднеквадратическое, отклонение

$$\mu = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

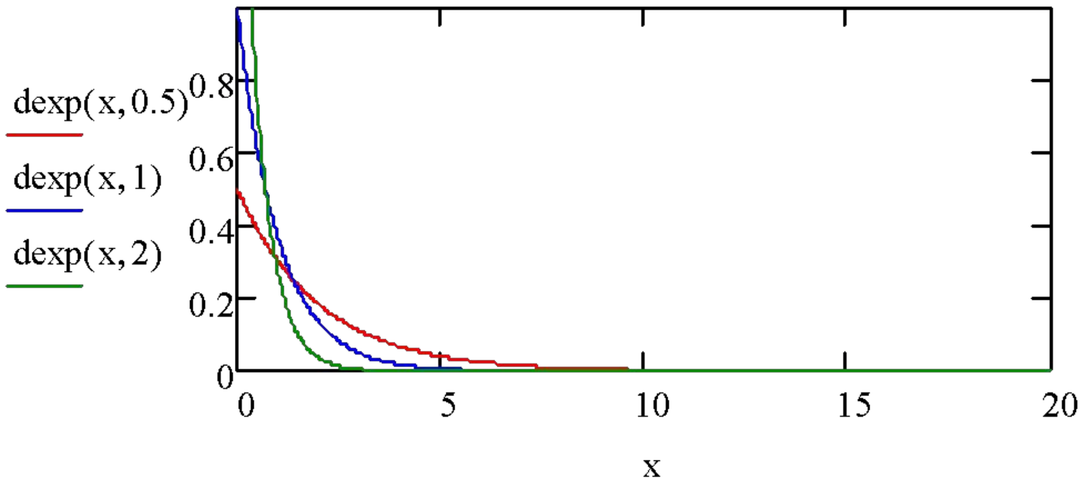
$$\sigma = \sqrt{\overline{(X^2)}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i)^2}$$

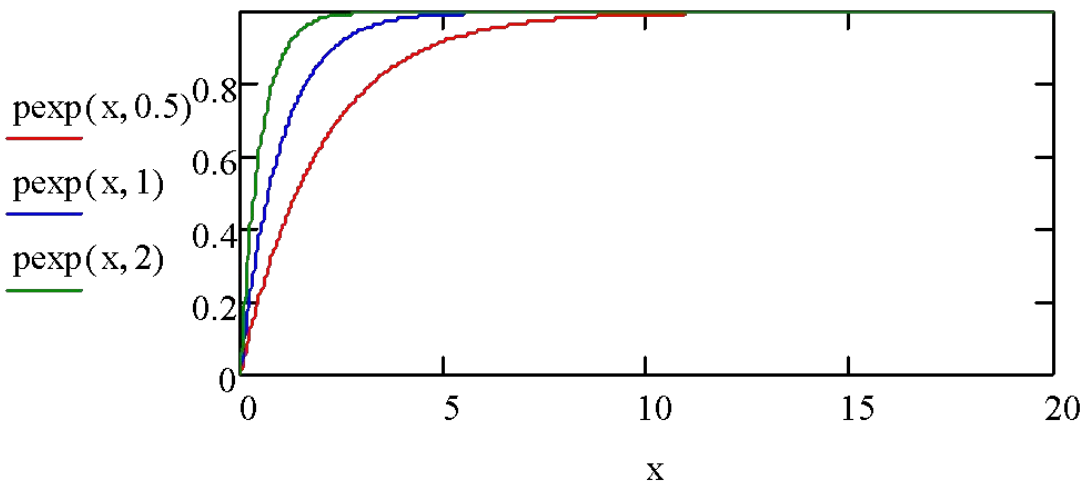
Нормальное распределение



t := 0.. 1000

Экспоненциальное распределение

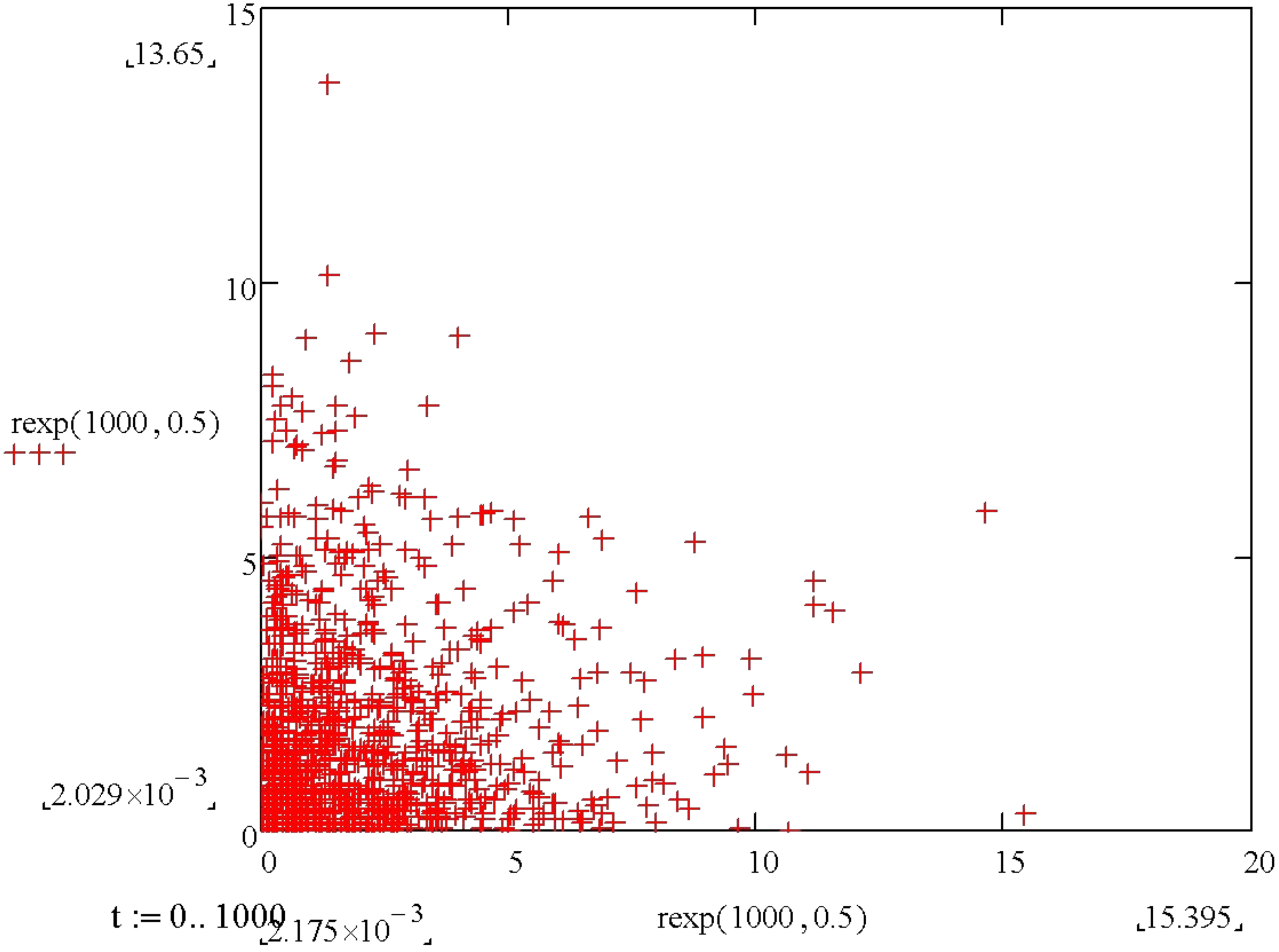




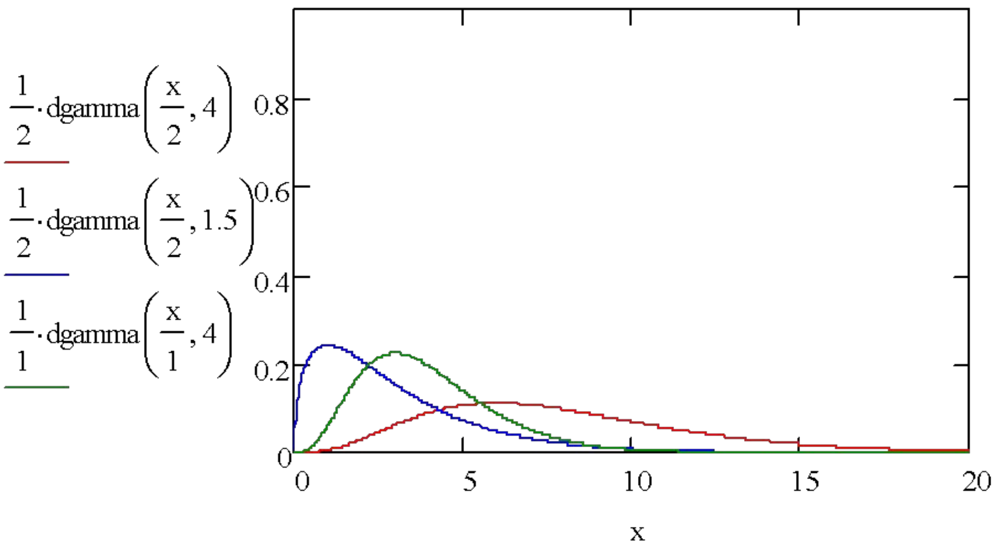
1 параметр масштаба λ

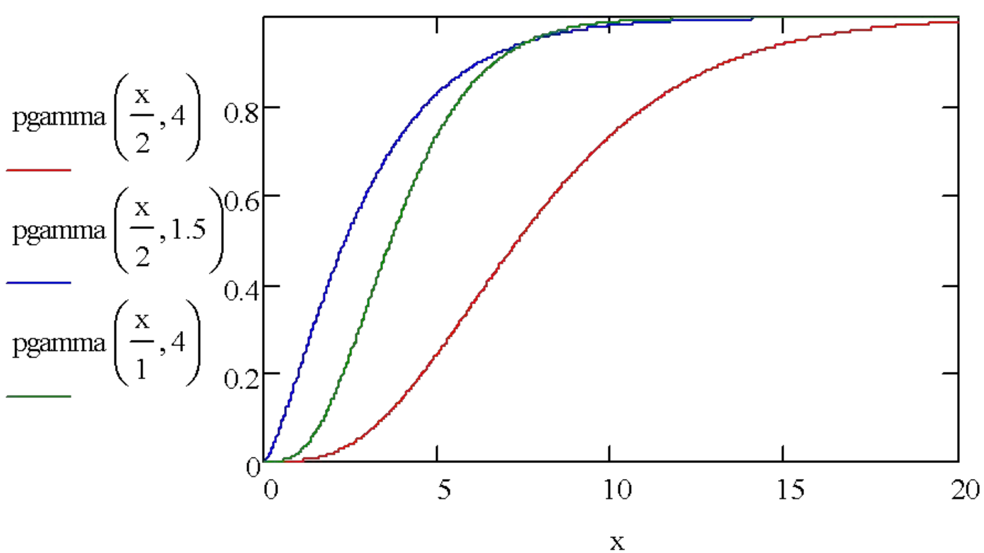
$\lambda = 1 / \mu$

Экспоненциальное распределение



Гамма-распределение





2 параметра:
k – параметр формы
θ – параметр масштаба

$$k = \frac{\mu}{\sigma^2}$$

$$\theta = \mu$$

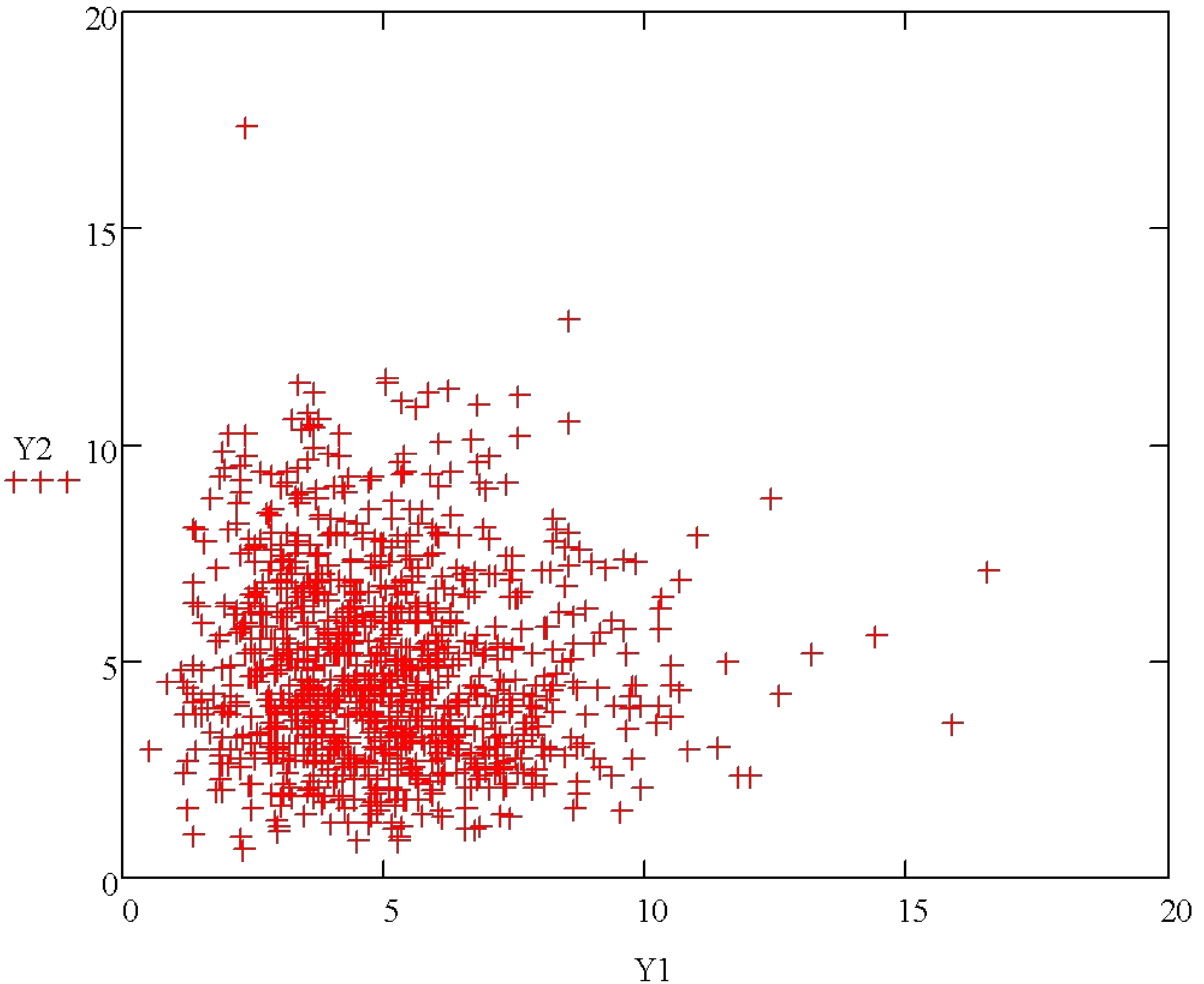
При k = 1 получается экспоненциальное распределение, где λ = 1 / θ

При k → ∞ получается нормальное распределение с параметрами $k\cdot\theta$ и $k\cdot\theta^2$

Гамма-распределение

$t := 0..1000$

$Y1_t := rgamma(3,5)_0$ $\quad$ $Y2_t := rgamma(3,5)_0$

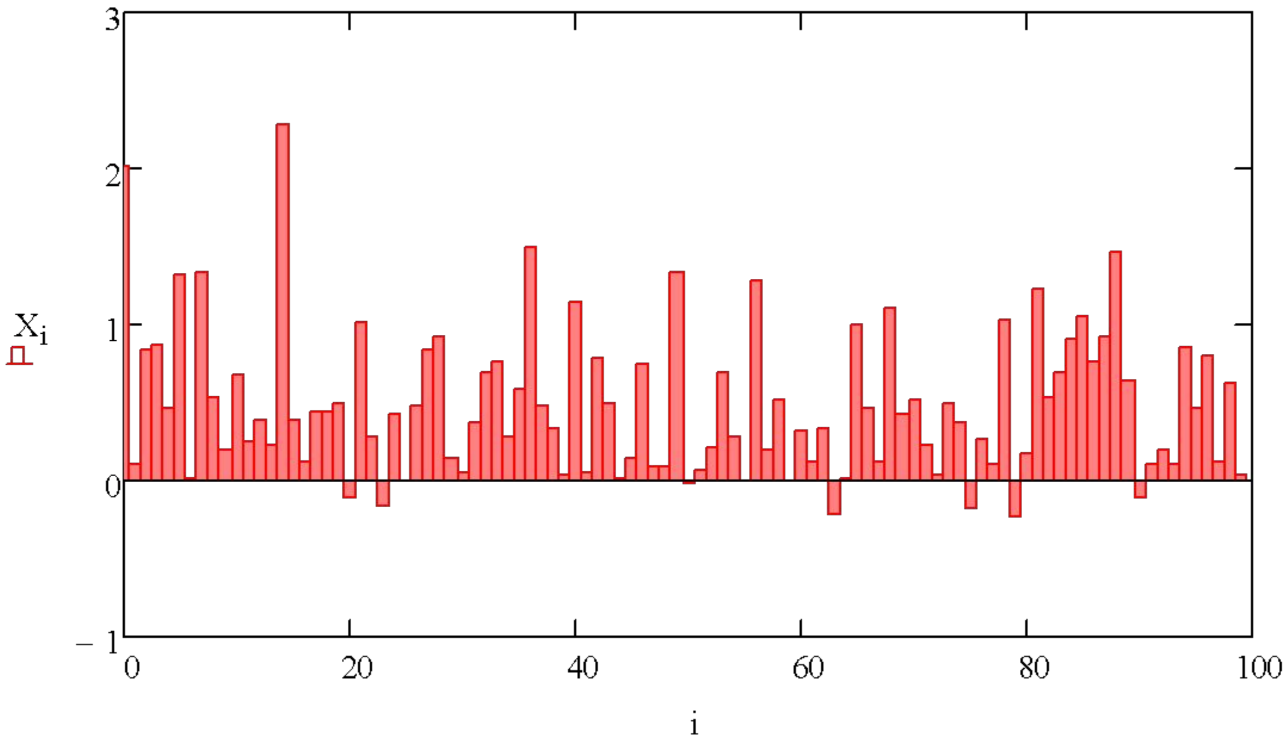


2. Построение гистограмм плотности вероятности и интегральной функции распределения

Исходный вектор значений случайной величины:

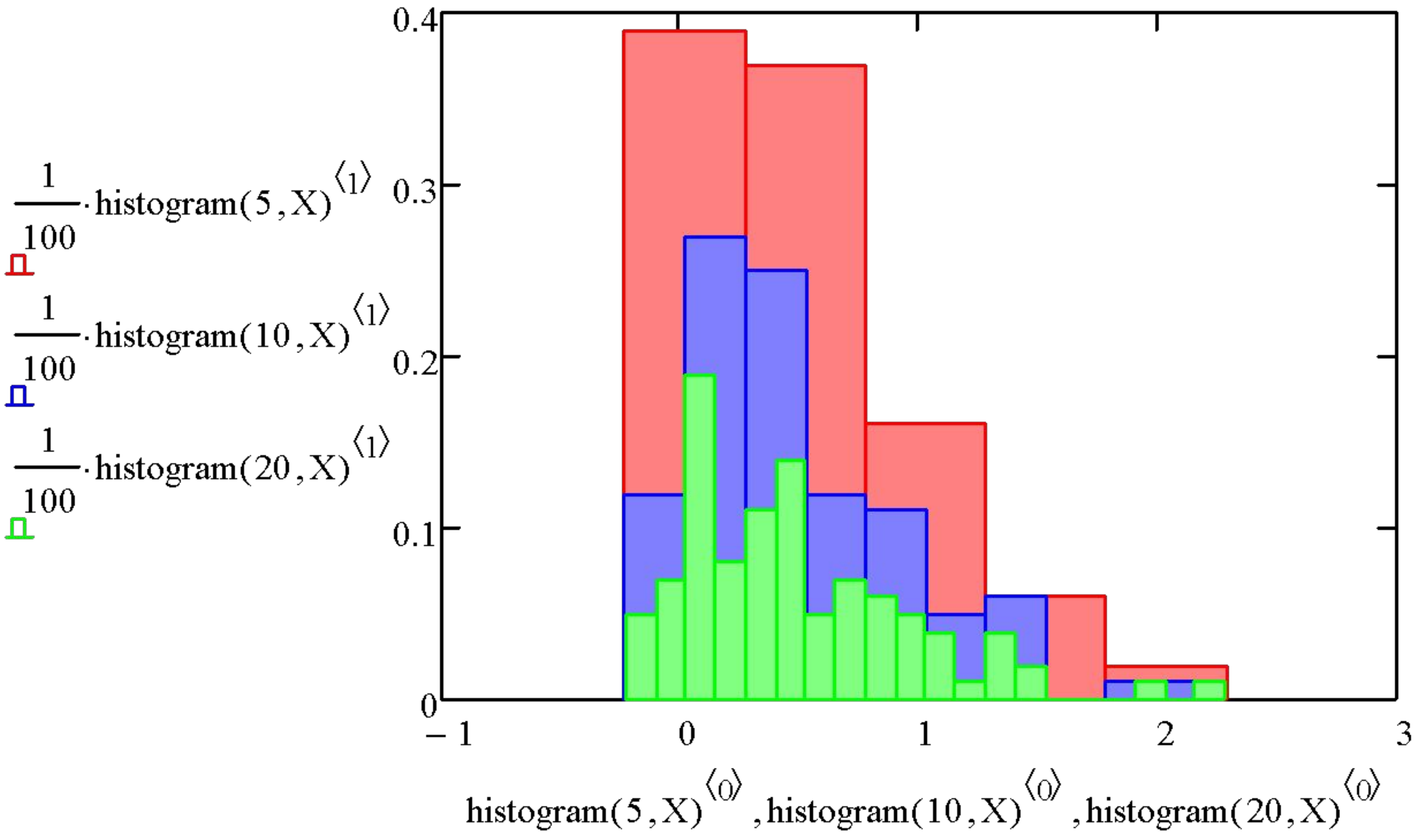
$i := 0..99$

$X_i := \mathrm{rnd}(1) \cdot \mathrm{rnorm}(1, 1, 0.7)_0$



Построение гистограмм плотности вероятности и интегральной функции распределения

Построение гистограммы плотности вероятности:



Построение гистограмм плотности вероятности и интегральной функции распределения

Для выбора числа интервалов (бинов) у гистограммы рекомендуется использовать формулу Стерджесса

$$n = 1 + 3.322 lgN = log_2 N + 1$$

N	15-24	25-44	45-89	90-179	180-359	360-719
n	5	6	7	8	9	10

Ширина каждого из интервалов

$$i = (X_{max} - X_{min}) / n$$

Ширину интервалов рекомендуется округлять.

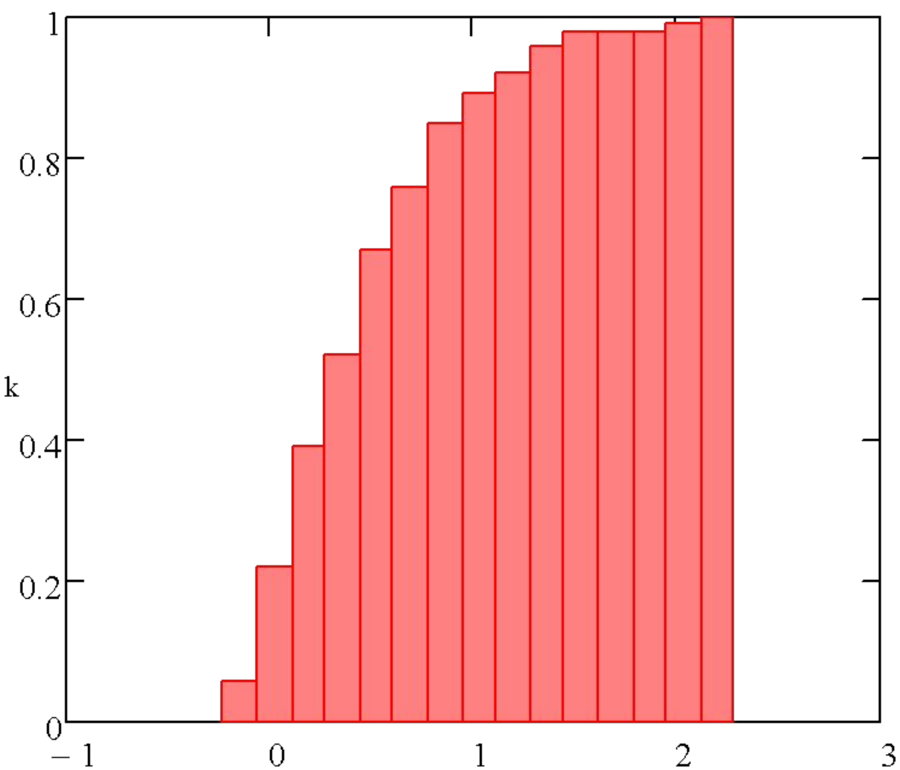
Функция histogram выбирает ширину автоматически.

Построение гистограмм плотности вероятности и интегральной функции распределения

Построение гистограммы интегральной функции распределения:

$$j := 0..\,\mathrm{length}\left(\mathrm{histogram}(15,X)^{\langle 0\rangle}\right) - 1$$

$$\frac{1}{100}\cdot\sum_{k=0}^{j}\left(\mathrm{histogram}(15,X)^{\langle 1\rangle}\right)_k$$



$$\left(\mathrm{histogram}(15,X)^{\langle 0\rangle}\right)_j$$

4. Вычисление математического ожидания, стандартного отклонения, дисперсии

- Математическое ожидание случайной величины вычисляется как её среднее значение, в mathCad вычисляется функцией **mean**
- Среднеквадратическое (стандартное) отклонение – корень из дисперсии, в mathCad вычисляется функцией **stdev** Обозначается **σ**
- Дисперсия – среднее значение квадрата отклонений от среднего значения ($\boldsymbol{\sigma^2}$), в mathCad вычисляется функцией **var**

$$\mathrm{mean}(X) = 0.486$$

$$\mathrm{var}(X) = 0.225$$

$$\mathrm{stdev}(X) = 0.475$$

5. Критерии достоверности гипотез

Гипотеза – предположение о виде или параметрах неизвестного распределения.

Например: гипотеза «случайная величина X подчиняется нормальному закону распределения»

Для каждой гипотезы есть вероятность p, что она верна, и вероятность 1 – p, что гипотеза ошибочна.

При проверке гипотез заранее задают **уровень значимости** $\alpha = 1 - p$, то есть вероятность недостоверности гипотезы.

Критерии достоверности гипотез

Для проверки гипотез вычисляют значение **критерия**, зависящее от значений проверяемой случайной величины, и проверяют его на нахождение в области значений, соответствующей достоверности гипотезы при заданном уровне значимости.

Наиболее часто используют критерий Колмогорова и критерий Пирсона (критерий «хи-квадрат» - χ^2).

Использование критерия Колмогорова

- Упорядочить случайные числа по возрастанию.
- Вычислить значения D_i и выбрать максимальное из них D

$D_n = \max\left|F(x_i) - \frac{2i-1}{2n}\right| + \frac{1}{2n}$, где x_i - члены вариационного ряда.

Значение критерия $\lambda = D\sqrt{n}$

- Найти вероятность совпадения законов распределения $P(\lambda)$.

Использование критерия Колмогорова при заданном уровне значимости

- Задавшись α и зная n, выбрать критическое значение критерия $D_{кр}$.
- Упорядочить случайные числа по возрастанию.
- Вычислить значения критерия D_i и выбрать максимальное из них D
- Гипотезу о принадлежности случайной величины распределению можно принять, если $D < D_{кр}$

Поиск критической точки для критерия Колмогорова

Для нахождения критической величины критерия $D_{кр}$ надо знать уровень значимости α и число опытов n

Для заданного α выбираем $\lambda_{кр}$:

α	0.4	0.3	0.2	0.1	0.05	0.025	0.01	0.005	0.001	0.0005
$\lambda_{кр}$	0.89	0.97	1.07	1.22	1.36	1.48	1.63	1.73	1.95	2.03

При больших n (n > 35)

$$D_{кр} = \frac{\lambda_{кр}}{\sqrt{n-1}}$$

Для малых n пользоваться табличными значениями $D_{кр}$

Критерий Колмогорова: пример

Пассажир, приходящий в случайные моменты времени на автобусную остановку, в течение пяти поездок фиксировал своё время ожидания автобуса: 5,1; 3,7; 1,2; 9,2; 4,8 мин. Проверить гипотезу о том, что время ожидания автобуса равномерно распределено на отрезке [0; 10] на уровне значимости 0,05.

Решение

На практике статистика Колмогорова вычисляется по формуле

$D_n = \max\left|F(x_i) - \frac{2i-1}{2n}\right| + \frac{1}{2n}$, где x_i - члены вариационного ряда.

$F(x) = \frac{x-a}{b-a}$

	x_i	$F(x_i)$	$\frac{2i-1}{2n}$	$\left\|F(x_i) - \frac{2i-1}{2n}\right\|$
1	1,2	0,12	0,1	0,02
2	3,7	0,37	0,3	0,07
3	4,8	0,48	0,5	0,02
4	5,1	0,51	0,7	0,18
5	9,2	0,92	0,9	0,02

Таким образом, значение статистики Колмогорова составляет $D_n = 0{,}18 + 0{,}1 = 0{,}28$.

По таблице критических точек находим $\lambda_{кр} = 0{,}56$.

Поэтому гипотеза о равномерном распределении выборки ПРИНИМАЕТСЯ.

Решение задачи в MathCAD

$n := 5$

$i := 0..n-1 \qquad a := 0 \qquad b := 10$

$$X := \begin{pmatrix} 1.2 \\ 3.7 \\ 4.8 \\ 5.1 \\ 9.2 \end{pmatrix} \qquad FV := \frac{(X-a)}{(b-a)} \qquad FV = \begin{pmatrix} 0.12 \\ 0.37 \\ 0.48 \\ 0.51 \\ 0.92 \end{pmatrix} \qquad F_i := \frac{(2\cdot i+1)}{2\cdot n} \qquad F = \begin{pmatrix} 0.1 \\ 0.3 \\ 0.5 \\ 0.7 \\ 0.9 \end{pmatrix}$$

$$h := \frac{(b-a)}{n}$$

$$FV - F = \begin{pmatrix} 0.02 \\ 0.07 \\ -0.02 \\ -0.19 \\ 0.02 \end{pmatrix}$$

$$0.19\cdot\sqrt{5} = 0.425$$

Критическое значение статистики Колмогорова 0.56 больше наблюдаемого 0.43, НУЛЕВУЮ ГИПОТЕЗУ о том, что распределение времени ожидания автобуса равномерное, ПРИНИМАЕМ.

Использование критерия Пирсона

Критерий используется для дискретных величин, либо непрерывных величин, разбитых на интервалы.

Например, он может быть использован, если построена гистограмма результатов эксперимента.

Использование критерия Пирсона

- Определить число степеней свободы
  $k = l - r - 1$, где
  l – число интервалов гистограммы
  r – число параметров предполагаемого распределения, оцениваемых по выборке
  (2 для нормального, 1 для экспоненциального…)

- Найти критическое значение критерия:
  $\chi^2_{кр}$ = qchisq(1 – α, k)

- Вычислить критерий χ^2 по экспериментальным данным. Гипотеза верна, если
  $\chi^2 < \chi^2_{кр}$

Использование критерия Пирсона

Вычисление χ^2:

$$\chi^2 = \sum_{i=1}^{l} \frac{(n_i - np_i)^2}{np_i}$$

где

- n_i – эмпирические частоты (фактическое количество попаданий случайной величины в заданный интервал гистограммы)
- np_i – теоретические частоты (количество попаданий случайной величины в заданный интервал гистограммы, вычисленное по предполагаемому закону её распределения)

Пример

Измерены интервалы в минутах между 100 поездами метро, прибывшими на станцию. Результаты измерений представлены статистическим рядом:

№ интервала	Интервал	Середина интервала	n_i
1	5,05-5,15	5,1	5
2	5,15-5,25	5,2	8
3	5,25-5,35	5,3	12
4	5,35-5,45	5,4	20
5	5,45-5,55	5,5	26
6	5,55-5,65	5,6	15
7	5,65-5,75	5,7	10
8	5,75-5,85	5,8	4

На уровне значимости $\alpha = 0,05$ проверить гипотезу о том, что интервалы можно описать нормальным распределением.

$$P(a < \xi < b) = \Phi(\frac{b - \bar{x}}{s}) - \Phi(\frac{a - \bar{x}}{s})$$

x_i	n_i	np_i	$n_i - np_i$	$(n_i - np_i)^2/(np_i)$
$\le 5{,}15$	5	3,67	1,33	0,47
5,15-5,25	8	7,59	0,41	0,02
5,25-5,35	12	15,00	-3,00	0,60
5,35-5,45	20	21,43	-1,43	0,09
5,45-5,55	26	22,14	3,86	0,67
5,55-5,65	15	16,53	-1,53	0,14
5,65-5,75	10	8,93	1,07	0,13
$\ge 5{,}75$	4	4,70	-0,70	0,10
Наблюдаемое значение статистики критерия 2,22				

$$k = 8 - 2 - 1 = 5; \quad \chi^2_{кр} = 11{,}1 > \chi^2_{набл} = 2{,}22$$

Неравенство **$\chi^2 < \chi^2_{кр}$** выполнено, гипотезу можно принять.

6. Коэффициент линейной корреляции

Коэффициент линейной корреляции – величина, показывающая наличие линейной связи между значениями двух случайных величин.

Для линейно зависящих величин он равен 1 или -1, для независимых величин – 0.

В MathCad вычисляется как функция от двух векторов случайных чисел.

$$X := \mathrm{rgamma}(100, 2)$$

$$X1 := \mathrm{rgamma}(100, 2)$$

$$\mathrm{corr}(X, 2X + 5) = 1$$

$$\mathrm{corr}(X1, -X1) = -1$$

$$\mathrm{corr}(X, X1) = 0.151$$

$$\mathrm{corr}\left(X, X^2\right) = 0.948$$

$$\mathrm{corr}\left(X1, X1^2\right) = 0.948$$

$$\mathrm{corr}\left(X, X^3\right) = 0.868$$

$$\mathrm{corr}\left(X1, X1^3\right) = 0.863$$

$$\mathrm{corr}(X, \exp(X)) = 0.686$$

$$\mathrm{corr}(X1, \exp(X1)) = 0.661$$

$$\mathrm{corr}(X, \sin(X)) = -0.468$$

$$\mathrm{corr}(X1, \sin(X1)) = -0.547$$

Задание к работе 6

1. Сгенерировать случайные числа по своему варианту два раза, векторы по 50 и 500 шт.
2. Найти мат. ожидание, стандартное отклонение, дисперсию для обоих наборов случайных чисел
3. Определить параметры распределения случайной величины, предполагая, что она распределена по известному закону распределения, для каждого из рассмотренных законов.
4. Построить гистограммы плотности вероятности и интегральной функции распределения.
   Построить функции плотности вероятности и интегральных функций распределения с найденными параметрами на тех же графиках.
5. Найти вероятности достоверности гипотез о принадлежности случайной величины к каждому из 4 распределений из п. 4 по критерию Колмогорова. Проверить те же гипотезы по критерию Пирсона для уровня значимости 0,1.
6. Найти коэффициент корреляции между 1 вектором и первыми 50 числами второго вектора.