Галанов Ю.И.

Лабораторный практикум по мат. статистике

Проверка гипотезы о виде распределения с помощью критерия ХИ - квадрат

Задание.

- **Для выборки X на уровне значимости 0.05 проверить гипотезу о том, что она описывается показательным распределением.**
- **Применить метод хи-квадрат.**

Указание.

- **Провести разбиение области значений функции на интервалы равной вероятности.**
- **Проверить гипотезу для двух распределений: показательного и треугольного.**
- **Вывести услоие равенства математических ожиданий для модельных распределений.**

Плотность показательного распределения $f(x,L) := L \cdot \exp(-L \cdot x)$

Функция распределения $F(x,L) := 1 - \exp(-L \cdot x)$

Обратная функция $qexp(u,L) := \frac{-1}{L} \cdot \ln(1-u)$

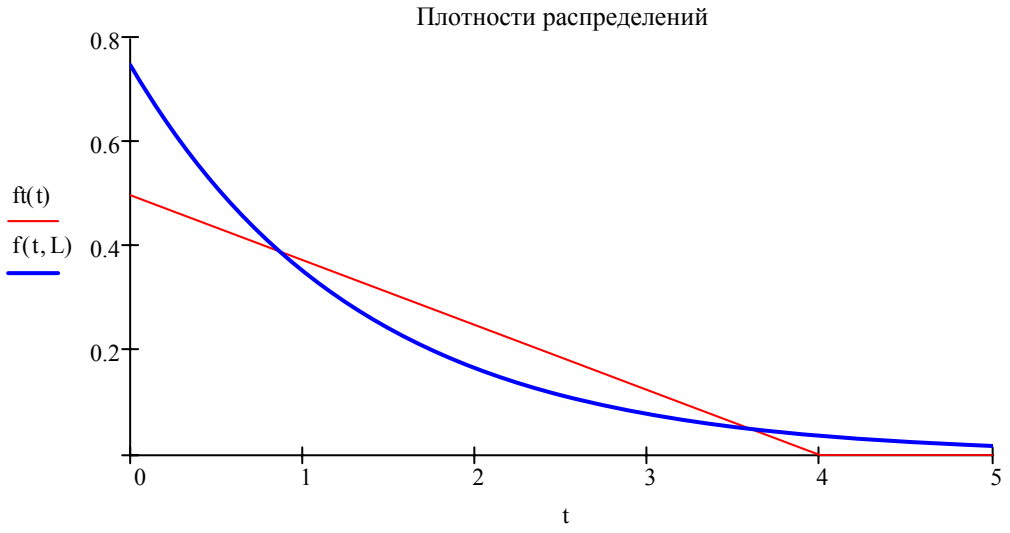
Параметры распределений задаем из условия равенства математических ожиданий

$a := 0.5$ $\quad b := \frac{2}{a}$ $\quad L := \frac{6}{a \cdot b^2}$ $\quad L = 0.75$

Плотность треугольного распределения $ft(x) := \mathrm{if}\left(x < 0, 0, \mathrm{if}\left(x > b, 0, a - \frac{a}{b} \cdot x\right)\right)$

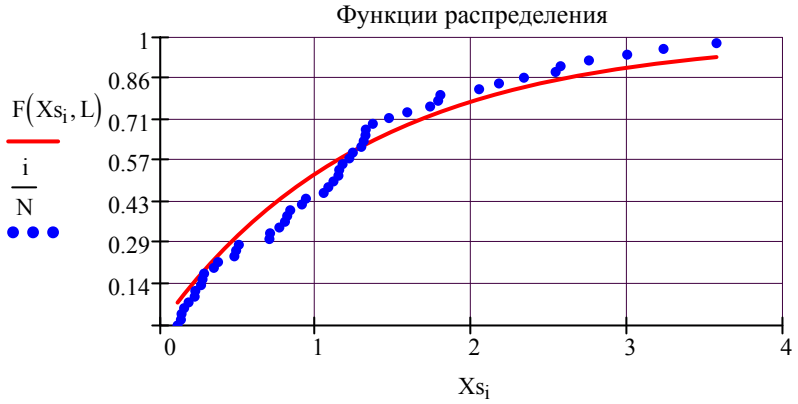
Функция распределения $Ft(x) := \mathrm{if}\left(x < 0, 0, \mathrm{if}\left(x > b, 1, a \cdot x - \frac{a}{b} \cdot \frac{x^2}{2}\right)\right)$

Обратная функция $qtr(t) := b - \sqrt{\frac{2 \cdot b \cdot (1-t)}{a}}$



Моделируем выборку из показательного распределения

$i := 0 .. N - 1$ $\quad X_i := qexp(rnd(1), L)$ $\quad$ **Сортируем** $\quad Xs := sort(X)$

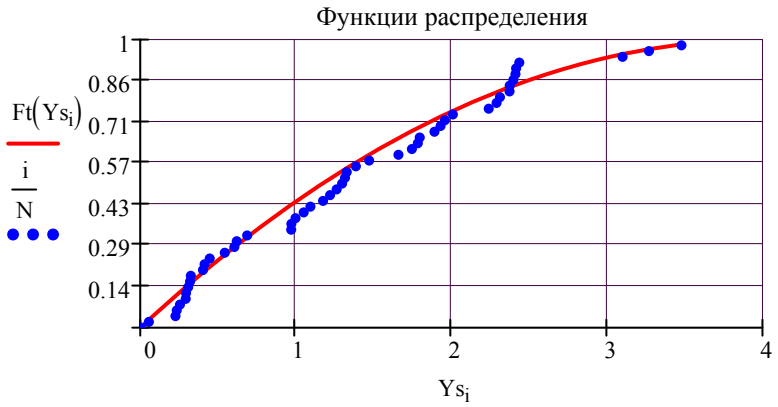


$L1 := \frac{1}{mean(X)}$

$L1 = 0.847$

Моделируем выборку из треугольного распределения

$Y_i := qtr(rnd(1))$ $\quad Ys := sort(Y)$ $\quad L2 := \frac{1}{mean(Y)}$



Задаем число интервалов группировки и формируем интервалы равной вероятности

$M := 1 + \mathrm{trunc}(\log(N,2))$ $\quad M = 6$ $\quad k := 0..M-1$ $\quad l_k := \frac{k}{M}$

$dx_k := \mathrm{qexp}(l_k, L1)$ $\quad dx_M := \max(X)\cdot 2$ $\quad dx^T = (0 \;\; 0.215 \;\; 0.478 \;\; 0.818 \;\; 1.296 \;\; 2.114 \;\; 7.149)$

$dy_k := \mathrm{qexp}(l_k, L2)$ $\quad dy_M := \max(Y)\cdot 2$ $\quad dy^T = (0 \;\; 0.246 \;\; 0.547 \;\; 0.935 \;\; 1.481 \;\; 2.416 \;\; 6.955)$

Подсчитываем число точек, попавших в интервалы группировки

$v := \mathrm{hist}(dx, X)$ $\quad vy := \mathrm{hist}(dy, Y)$

$v^T = (5 \;\; 8 \;\; 7 \;\; 12 \;\; 10 \;\; 8)$ $\quad vy^T = (4 \;\; 10 \;\; 3 \;\; 13 \;\; 16 \;\; 4)$

Теоретическое число точек, попавших в интервалы группировки

$vt := \frac{N}{M}$ $\quad vt = 8.333$

Число степеней свободы

$r := M - 1 - 1 \quad r = 4$

Критическая точка

$C := \mathrm{qchisq}(0.95, r)$ $\quad C = 9.488$

Статистики критерия

$hiX := \sum_k \frac{(v_k - vt)^2}{vt}$ $\quad hiX = 3.52$ $\quad hiY := \sum_k \frac{(vy_k - vt)^2}{vt}$ $\quad hiY = 17.92$

Функция принятия решения (для ленивых)

$\mathrm{Crit}(t) := \mathrm{if}(t < C, \text{"Ho --- Prinimaem "}, \text{"Ho --- Otwergaem"})$

$N \equiv 50$ $\quad \mathrm{Crit}(hiX) = \text{"Ho --- Prinimaem "}$ $\quad \mathrm{Crit}(hiY) = \text{"Ho --- Otwergaem"}$

УКАЗАНИЕ: оцените ошибку второго рода и мощность критерия; результаты представьте в виде таблицы

N	β	$1-\beta$	Критическая точка
50			
100			
200			
500			
1000			