Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное
образовательное учреждение высшего образования
ПСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Институт инженерных наук
Кафедра информационно-коммуникационных технологий

ЛАБОРАТОРНАЯ РАБОТА №4

ИНТЕРВАЛЬНОЕ ОЦЕНИВАНИЕ ПАРАМЕТРОВ НОРМАЛЬНО РАСПРЕДЕЛЕННОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

Вариант 13

по дисциплине «Моделирование»

Выполнила: Разгонова Е.В.

Группа: 0432-04

Проверил: Миронов Т.С.

Псков
2021

Задание 4.1. Доверительные интервалы для математического ожидания. Доверительный интервал для дисперсии

Задание: Найдите доверительные интервалы для математического ожидания $M(\xi)$ и дисперсии $D(\xi)$ по выборке из нормального распределения.

x	904.3	910.2	916.6	928.8	935.0	941.2	947.4	953.6	959.8	966.0	972.2	978.4
n	3	1	2	7	8	10	4	2	4	1	1	1

$ORIGIN := 1 \qquad i := 1..12$

Выборка генеральной совокупности

$$D := \begin{pmatrix} 904.3 & 910.2 & 916.6 & 928.8 & 935.0 & 941.2 & 947.4 & 953.6 & 959.8 & 966.0 & 972.2 & 978.4 \\ 3 & 1 & 2 & 7 & 8 & 10 & 4 & 2 & 4 & 1 & 1 & 1 \end{pmatrix}$$

Объём выборки

$$n := \sum_{i=1}^{12} D_{2,i} \qquad n = 44$$

Точечные оценки мат. ожидания и дисперсии

$$Mx := \frac{1}{n} \cdot \sum_{i=1}^{12} \left(D_{1,i} \cdot D_{2,i}\right) \qquad Mx = 938.693$$

$$Dx := \frac{1}{n-1} \cdot \sum_{i=1}^{12} \left[D_{2,i} \cdot \left(D_{1,i} - Mx\right)^2\right] \qquad Dx = 282.988$$

95 %-ный доверительный интервал для математического ожидания (p = 1 - α, где α = 0,95)

$$t := qt\left(1 - \frac{0.05}{2}, n\right) \qquad t = 2.015$$

$$xl := Mx - t \cdot \sqrt{\frac{Dx}{n}} \qquad xl = 933.582 \qquad xr := Mx + t \cdot \sqrt{\frac{Dx}{n}} \qquad xr = 943.804$$

левая граница интервала правая граница интервала

90 %-ный доверительный интервал для дисперсии (p = 1 - α, α = 0,9)

$$hl := qchisq\left(\frac{0.1}{2}, n-1\right) \qquad hl = 28.965 \qquad hr := qchisq\left(1 - \frac{0.1}{2}, n-1\right) \qquad hr = 59.304$$

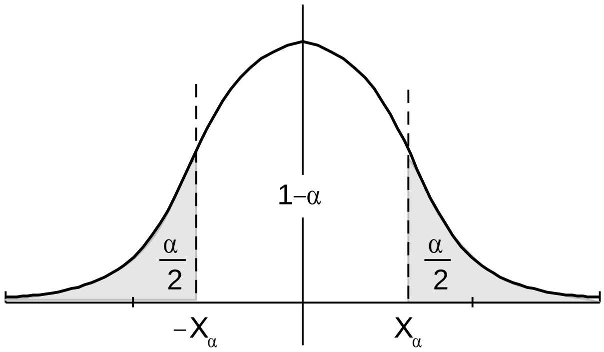
$$dl := Dx \cdot \frac{n-1}{hr} \qquad dl = 205.19 \qquad dr := Dx \cdot \frac{n-1}{hl} \qquad dr = 420.114$$

Рисунок 1. Фрагмент рабочего документа MathCAD

Пояснение:

Доверительным называют интервал, в который попадают измеренные в эксперименте значения, соответствующие доверительной вероятности p. Предпочтительнее, используется при выборках малого объёма.

В данном задании, в ряду использования вероятностного интервала, возникает понятие квантúля. Квантúль – значение случайной величины, которое все остальные значения этой величины не превышают с вероятностью p.



Для того, чтобы говорить об доверительном интервале, необходимо определится с распределением, которое позволяет получить любую интервальную оценку. К примеру, квантиль стандартного нормального распределения используется для оценки идеального среднего (оно же мат. ожидание). Применив его, получим следующую формулу для доверительного интервала:

$$\bar{x} - \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}} \leq \xi \leq \bar{x} + \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}}$$

В данной лабораторной работе используется квантиль Стьюдента. Так как стандартное нормальное распределение и распределение Стьюдента при определённых условиях совпадают, мы от одной формулы можем перейти к другой, посредством применения квантиля Стьюдента $t = \frac{\alpha - \bar{x}}{S}\sqrt{n}$ и получить следующую интервальную оценку:

$$\bar{x} - \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}} \leq \xi \leq \bar{x} + \frac{S}{\sqrt{n}} t_{1-\frac{\alpha}{2}},$$

где S – выборочная дисперсия (в случае нормального распределения – идеальная дисперсия).

В ходе работы, для расчёта квантиля для доверительного интервала мат. ожидания, использовалась функция распределения Стьюдента $qt(p, n)$, где p – это вещественная вероятность попадания в интервал (область принятия), а n

– степень свободы. В случае нахождения границ интервала для дисперсии используется другая функция – $qchisq(p, n)$. Они позволяют по заданной вероятности вычислить такое значение x, при котором вероятность равна или меньше заданного значения p.

В случае, если известно распределение выборочной дисперсии S^2, то можно провести интервальную оценку генеральной дисперсии σ^2. Распределение величины S^2 можно получить с помощью **распределения $\boldsymbol{\chi^2}$** с k степенями свободы — распределение суммы квадратов k независимых стандартных нормальных случайных величин.

Вывод: в ходе выполнения задания, найдены 95%-ный доверительный интервал для математического ожидания $(933.582; 943.804)$ и 90%-ный доверительный интервал для дисперсии $(205{,}19; 420.114)$.

Примечание: при использовании функции нахождения квантиля, параметр вероятности определяется через выражение $p = 1 - \alpha$, где α – уровень значимости. Уровень значимости соответствует практически невозможному событию. Уровень значимости и уровень достоверности в сумме дают единицу. В данной работе достоверная вероятность равна 95% и 90% (0.95 и 0.9 соответственно).

Задание 4.2. Доверительный интервал для параметра пуассоновского распределения

Задание: Найдите доверительный интервал для параметра $\lambda = 3$ по заданной выборке $x_1, x_2 \dots x_n$ из пуассоновского распределения.

Пояснение: как известно, при моделировании и исследовании выборок применяются различные функции, по которым можно провести анализ значений и извлечь некоторую информацию. В случае работы с выборками фигурируют два вида функций – выборочная функция распределение и функция плотности вероятности.

Согласно заданию, нам необходимо построить доверительный интервал для выборки, подчиняющейся пуассоновскому распределению.

<u>Распределение Пуассона</u> – распределение случайной величины, характеризующееся числом событий, произошедших за фиксированное время, при условии, что данные события происходят с некоторой интенсивностью и не зависят друг от друга. Интенсивность в данной работе определяется буквой λ (здесь оно равно 3).

$ORIGIN := 1$

$\alpha := 0.1 \quad \lambda := 3 \quad N := 500 \quad P := rpois(N, \lambda)$

Интервал построения функции плотности вероятности (Pmax; Pmin)

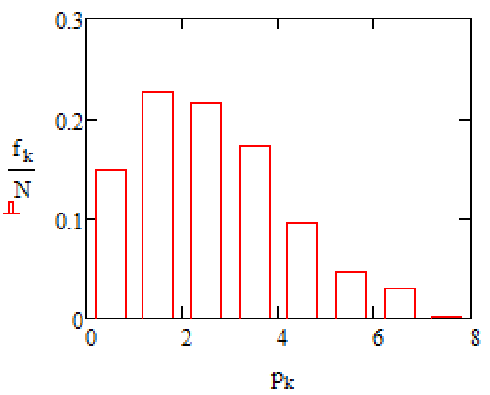
$Pmax := max(P) = 9 \quad Pmin := min(P) = 0 \quad R := Pmax - Pmin = 9$

$PS := sort(P)$

Для построения плотности, необходимо подобрать m таким образом, чтобы Δ была приближена к 1

$m := 9 \quad \Delta := \frac{R}{m} \quad j := 1..m \quad k := 1..m-1$

$p_j := Pmin + \frac{\Delta}{2}\cdot(2\cdot j - 1) \quad f := hist(p, PS)$



$$p = \begin{pmatrix} 0.5 \\ 1.5 \\ 2.5 \\ 3.5 \\ 4.5 \\ 5.5 \\ 6.5 \\ 7.5 \\ 8.5 \end{pmatrix}$$

$p_j =$

0.5
1.5
2.5
3.5
4.5
5.5
6.5
7.5
8.5

$P =$

	1
1	0
2	2
3	3
4	1
5	2
6	3
7	1
8	8
9	4
10	4
11	5
12	3
13	3
14	3
15	3
16	...

Рисунок 2. Рабочий документ MathCAD с вычислениями (1)

Изучив распределение случайной величины, можно провести оценку параметра λ и построить доверительный интервал.

Точечные оценки мат. ожидания и дисперсии для параметра выглядят как: $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i$, следовательно $M[\hat{\lambda}] = \lambda$ и $D[\hat{\lambda}] = \frac{\lambda}{n}$.

Чтобы образовать случайную величину стандартного распределения, которое не зависит от значения параметра, можно показать, что при объёме выборки $n \to \infty$ случайная величина $\frac{2\sqrt{\hat{\lambda}}-2\sqrt{\lambda}}{\sqrt{\frac{1}{n}}}$ стремится к случайной величине со стандартным нормальным распределением, т.е. нормально распределена асимптотически с параметрами 0 и 1 (что и используется при вычислении интервала в лаб. работе).

Предположим, что если величина x_α это решение уравнения $\phi(x_\alpha) = 1 - 0.5\alpha$, то для любого значения λ, где вероятность попадания значения в интервал приближена к доверительной вероятности, интервал можно записать в следующем виде:

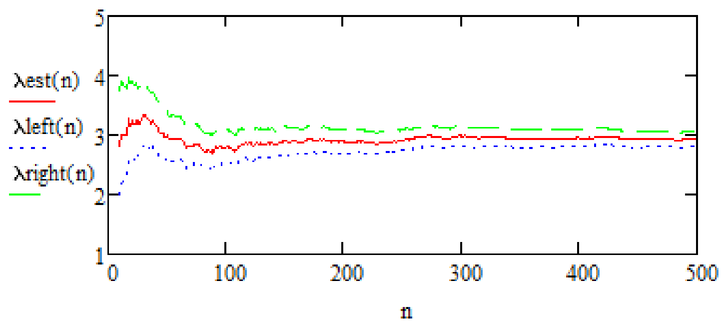
$$\left(\left(\sqrt{\hat{\lambda}} - \frac{x_\alpha}{2\sqrt{n}}\right)^2; \left(\sqrt{\hat{\lambda}} + \frac{x_\alpha}{2\sqrt{n}}\right)^2\right)$$

$x\alpha := \mathrm{qnorm}\left(1 - \frac{\alpha}{2}, 0, 1\right)$ $\quad x\alpha = 1.645$ $\quad n := 10..500$

$\lambda est(n) := \frac{1}{n} \cdot \sum_{j=1}^{n} P_j$ $\quad \lambda est(50) = 3$ $\quad \lambda est(100) = 2.77$ $\quad$ Вычисление оценок параметра λ

$\lambda left(n) := \left(\sqrt{\lambda est(n)} - \frac{x\alpha}{2 \cdot \sqrt{n}}\right)^2$ $\quad \lambda right(n) := \left(\sqrt{\lambda est(n)} + \frac{x\alpha}{2 \cdot \sqrt{n}}\right)^2$

Расчёт левой и правой границ доверительного интервала



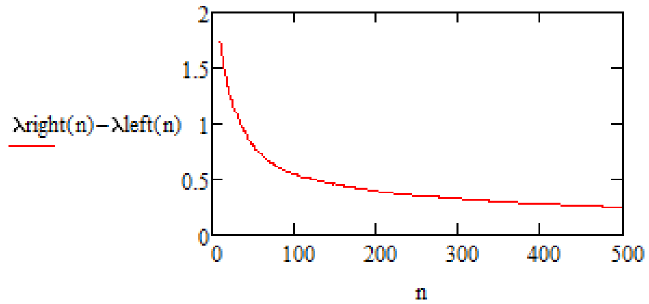


Рисунок 3. Рабочий документ MathCAD с вычислениями (2)

Вывод: исходя из первого графика, можно понять, что оцениваемый параметр λ пуассоновского распределения при всех n покрыт доверительным интервалом. Второй график показывает, что с уменьшением длины доверительного интервала оценка случайной величины становится лучше и точнее.

Задание 4.3. Доверительный интервал для вероятности

Задание: Найдите доверительный интервал для вероятности события по заданным значениям числа испытаний n (80) и числа m (35) появлений события в серии из n испытаний.

$ORIGIN := 1 \quad \alpha := 0.1 \quad n := 80 \quad m := 35$

$x\alpha := qnorm\left(1 - \frac{\alpha}{2}, 0, 1\right) \quad x\alpha = 1.645 \quad p := \frac{m}{n} \quad p = 0.438$

$$pleft := \left(\sin\left(asin\left(\sqrt{p}\right) - \frac{x\alpha}{2\cdot\sqrt{n}}\right)\right)^2 \qquad pleft = 0.348$$

$$pright := \left(\sin\left(asin\left(\sqrt{p}\right) + \frac{x\alpha}{2\cdot\sqrt{n}}\right)\right)^2 \qquad pright = 0.529$$

$n := 10..80 \quad m(n) := rbinom(10, n, p)_1 \quad p1 := p \quad p(n) := \frac{m(n)}{n}$

$$pleft(n) := \left(\sin\left(asin\left(\sqrt{p(n)}\right) - \frac{x\alpha}{2\cdot\sqrt{n}}\right)\right)^2$$

$$pright(n) := \left(\sin\left(asin\left(\sqrt{p(n)}\right) + \frac{x\alpha}{2\cdot\sqrt{n}}\right)\right)^2$$

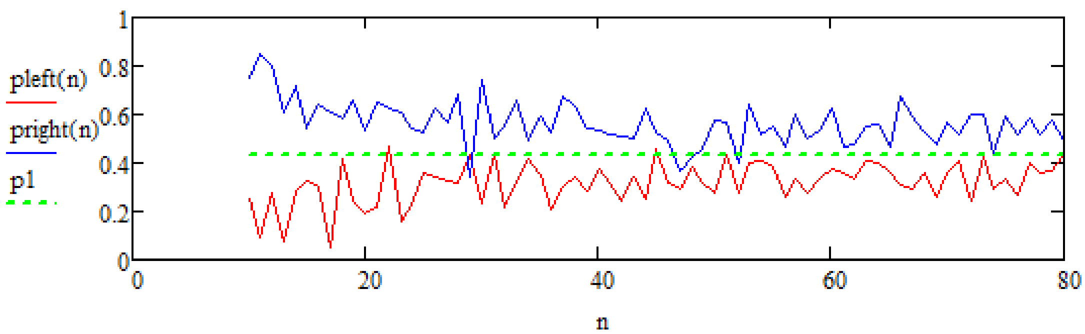


Рисунок 4. Рабочий документ MathCAD с вычислениями

Вывод: с ростом объема выборки n растет и точность вычисления границ доверительного интервала, а сам интервал полностью покрывает вероятность p.

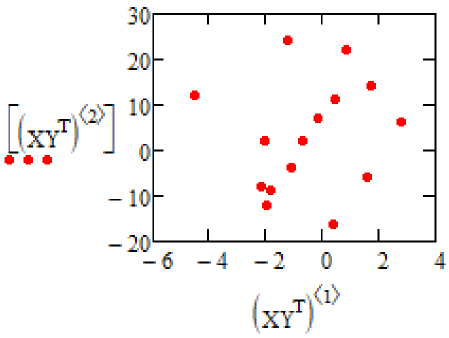
Задание 4.4. Доверительный интервал для коэффициента корреляции

Задание: найдите доверительный интервал для коэффициента корреляции по заданной выборке $(x_1, y_1), (x_2, y_2), \dots (x_n, y_n)$ из двумерной случайной величины.

$$ORIGIN := 1$$

$$\alpha := 0.1 \qquad n := 15 \qquad x\alpha := qnorm\left(1 - \frac{\alpha}{2}, 0, 1\right) \qquad x\alpha = 1.645$$

$$XY := \begin{pmatrix} -2.07 & 0.404 & -1.96 & 2.793 & 1.692 & -1.82 & -2.143 & 0.822 & -1.079 & -0.725 & 0.447 & 1.559 & -0.194 & -1.236 & -4.506 \\ 2.249 & -16.375 & -12.108 & 6.09 & 14.304 & -8.687 & -7.723 & 22.213 & -3.647 & 2.218 & 11.291 & -5.649 & 7.16 & 24.184 & 11.975 \end{pmatrix}$$



Вычисление выборочных средних, m, δ2x и δ2y

$$Xmean := \frac{1}{n} \cdot \sum_{i=1}^{n} XY_{1,i} \qquad Ymean := \frac{1}{n} \cdot \sum_{i=1}^{n} XY_{2,i}$$

$$m := \frac{1}{n} \cdot \sum_{i=1}^{n} \left[\left(XY_{1,i} - Xmean\right) \cdot \left(XY_{2,i} - Ymean\right)\right]$$

$$\delta 2x := \frac{1}{n} \cdot \sum_{i=1}^{n} \left(XY_{1,i} - Xmean\right)^2 \qquad \delta 2x = 3.296$$

$$\delta 2y := \frac{1}{n} \cdot \sum_{i=1}^{n} \left(XY_{2,i} - Ymean\right)^2 \qquad \delta 2y = 139.786$$

Рисунок 6. Рабочий документ MathCAD с вычислениями (1)

Вычисляем выборочный коэффициент корреляции и доверительный интервал

$$k := \frac{m}{\sqrt{\delta 2x \cdot \delta 2y}} \qquad k = 0.134$$

$$kleft := \tanh\left(atanh(k) - \frac{x\alpha}{\sqrt{n-3}}\right) \qquad kleft = -0.328$$

$$kright := \tanh\left(atanh(k) + \frac{x\alpha}{\sqrt{n-3}}\right) \qquad kright = 0.544$$

$$k := \frac{\sum_{i=1}^{n} \left(XY_{1,i} \cdot XY_{2,i}\right) - \frac{1}{n} \cdot \sum_{i=1}^{n} XY_{1,i} \cdot \sum_{i=1}^{n} XY_{2,i}}{\sqrt{\left[\sum_{i=1}^{n} \left(XY_{1,i}\right)^2 - \frac{1}{n} \cdot \left[\sum_{i=1}^{n} \left(XY_{1,i}\right)\right]^2\right] \left[\sum_{i=1}^{n} \left(XY_{2,i}\right)^2 - \frac{1}{n} \cdot \left[\sum_{i=1}^{n} \left(XY_{2,i}\right)\right]^2\right]}}$$

$$k = 0.134$$

$$kmin := \tanh\left(\left(atanh(k) - \frac{x\alpha}{\sqrt{n-3}}\right)\right) \qquad kmin = -0.328$$

$$kmax := \tanh\left(\left(atanh(k) + \frac{x\alpha}{\sqrt{n-3}}\right)\right) \qquad kmax = 0.544$$

Рисунок 7. Рабочий документ MathCAD с вычислениями (2)

Пояснение:

Случайную зависимость случайных величин обычно называют корреляцией, которая является одним из вариантов зависимости числовых значений. Корреляция – это, простыми словами, взаимосвязанное изменение показателей. Эта взаимосвязь значений проявляется в следующем: при изменении одного показателя имеет место изменение другого.

У корреляции два коэффициента измерения: корреляционный момент (или ковариация) и коэффициент корреляции. Корреляционным моментом называется математическое ожидание от произведения отклонений между значениями случайных величин и своим средним значением (или мат. ожиданием) и значением другой случайной величины и своим мат. ожиданием:

$$M\left[(y - M_y)(x - M_x)\right] \neq 0$$

Т.е. допустим, что у нас имеется два массива случайных величин X и Y, которые имеют собственные мат. ожидания. Именно «общее» мат. ожидание и будет являться корреляционным моментом.

В практическом применении корреляционный момент – это некое среднее, т.е. сумма произведений между каждым значением и его средним, делённое на количество пар n (именно пары, а не объём массива).

$$m = \frac{\sum_1^n (y - M_y)(x - M_x)}{n}$$

Чаще всего корреляция измеряется коэффициентом корреляции, а не корреляционным моментом.

$$\rho = \frac{M\left[(y - M_y)(x - M_x)\right]}{\sqrt{D_y * D_x}} = \frac{\sigma^2{}_{yx}}{\sigma_y \sigma_x}$$

Следовательно, коэффициент будет являться безразмерной величиной, что отличает его от корреляционного момента, т.к. во втором случае существует зависимость между образующимися парами, потому что как только они меняются, то возникает функциональная зависимость, которую мы хотим избежать.

Корреляционный момент изменяется в пределах $-\infty < m < \infty$ и определяет два показателя корреляции случайной величины – направление и силу. Говоря о направлении, подразумевается какие значения имеет корреляционный момент (или коэффициент корреляции), положительные или отрицательные. Например, если $m < 0$, то направление зависимости обратное, в противном случае – прямое. Силу можно определить только по коэффициенту корреляции $(-1 \leq m \leq 1)$, если он ближе к 1, то зависимость сильная, если ближе к 0, то слабая.

Вывод: в ходе выполнения лабораторной работы был получен достаточно доверительный интервал для коэффициента корреляции $k = 0.134$ $(-0.328\,; 0.544)$. Данный интервал достаточно широкий, поэтому нельзя утверждать, что полученный выборочный коэффициент корреляции является точной оценкой генерального коэффициента корреляции. Зависимость исследуемых величин – прямая (т.к. коэффициент положителен) слабая (лежит вблизи 0) корреляционная зависимость.