МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО
ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«САМАРСКИЙ ГОСУДАРСТВЕННЫЙ АЭРОКОСМИЧЕСКИЙ
УНИВЕРСИТЕТ ИМЕНИ АКАДЕМИКА С.П. КОРОЛЕВА
(НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)» (СГАУ)

Коломиец Э.И.

МОДЕЛИРОВАНИЕ И СТАТИСТИЧЕСКИЙ АНАЛИЗ СЛУЧАЙНЫХ ДАННЫХ

Электронное учебное пособие

САМАРА
2011

УДК 517.2 (075)
ББК 22.171
К 612

Автор: **Коломиец Эдуард Иванович**

Коломиец, Э. И. Моделирование и статистический анализ случайных данных [Электронный ресурс]: электрон. учеб. пособие / Э. И. Коломиец; М-во образования и науки РФ, Самар. гос. аэрокосм. ун-т им. С. П. Королева (нац. исслед. ун-т). – Электрон. и граф. дан. (1,25 Мбайт). - Самара, 2011. - 1 эл. опт. диск (CD-ROM).

Учебное пособие содержит полное методическое обеспечение всех видов учебных занятий по модулю «Математическая статистика» курса «Теория вероятностей и математическая статистика». В состав учебного пособия входят: краткие теоретические сведения, методические указания по проведению практических занятий, варианты индивидуального задания и методические указания по его выполнению с использованием универсальных пакетов MCAD и MATLAB.

Учебное пособие предназначено для подготовки бакалавров, обучающихся на факультете информатики по направлению 010400.62 «Прикладная математика и информатика», при изучении дисциплины «Теория вероятностей и математическая статистика» в 4 семестре.

Разработано на кафедре технической кибернетики.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ………………………………………………………………… 4
1 ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ……………………………………….. 5
1.1 Выборка. Эмпирическая функция распределения Гистограмма. Выборочные числовые характеристики…………. 5
1.2 Оценивание неизвестных параметров распределений…………. 11
1.2.1 Точечные оценки. Методы нахождения точечных оценок 11
1.2.2 Интервальные оценки……………………………………... 15
1.3 Проверка статистических гипотез………………………………... 19
1.3.1 Проверка гипотезы о виде распределения……………….. 20
1.4 Изучение зависимости между случайными величинами……….. 22
1.4.1 Оценка коэффициента корреляции………………………. 23
1.4.2 Проверка гипотезы о независимости…………………….. 23
1.4.3 Эмпирические уравнения регрессии…………………….. 25
1.5 Моделирование случайных величин и векторов………………... 26
1.5.1 Моделирование непрерывных случайных величин……. 26
1.5.2 Моделирование гауссовского случайного вектора……... 31
2 ПРАКТИЧЕСКИЕ ЗАНЯТИЯ……………………………………….... 34
2.1 Первичная обработка статистических данных…………………. 34
2.2 Точечные оценки неизвестных параметров……………………... 36
2.3 Интервальные оценки неизвестных параметров………………... 42
2.4 Проверка статистических гипотез……………………………….. 44
3 ИНДИВИДУАЛЬНОЕ ЗАДАНИЕ «МОДЕЛИРОВАНИЕ И СТАТИСТИЧЕСКИЙ АНАЛИЗ СЛУЧАЙНЫХ ДАННЫХ»………… 48
3.1 Содержание задания……………………………………………… 48
3.2 Исходные данные к заданию……………………………………... 49
3.3 Методические указания по выполнению задания……………… 49
3.4 Требования к оформлению отчета………………………………. 52
ЗАКЛЮЧЕНИЕ………………………………………………………….. 54
ЛИТЕРАТУРА…………………………………………………………… 55
Приложение А. Варианты индивидуальных заданий…………………. 56
Приложение Б. Нормальное распределение…………………………… 66
Приложение В. Распределение Стьюдента $S(n)$ ……………………… 69
Приложение Г. Распределение хи-квадрат $\chi^2(n)$ ……………………… 70
Приложение Д. Образец оформления титульного листа отчета…………. 72

ВВЕДЕНИЕ

Тезис о том, что «критерий истины есть практика» имеет самое непосредственное отношение к математической статистике,- науке, занимающейся анализом случайных данных. Именно эта наука изучает методы (в рамках точных математических моделей), которые позволяют отвечать на вопрос, соответствует ли практика, представленная в виде результатов эксперимента, данному гипотетическому представлению о природе явления или нет. При этом имеются в виду не эксперименты, которые позволяют делать однозначные, детерминированные выводы о рассматриваемых явлениях, а эксперименты, результатами которых являются случайные события. С развитием науки задач такого рода становится все больше и больше, поскольку с увеличением точности экспериментов становится все труднее избежать «случайного фактора», связанного с различными помехами и ограниченностью наших измерительных и вычислительных возможностей. Вот почему за последнее время статистические методы, проникнув в самые разнообразные области науки и техники, стали широко использоваться при анализе и обработке опытных данных. Этот процесс находит отражение и в обучении по направлению «Прикладная математика и информатика», в соответствии с учебным планом которого существенное время отводится на изучение дисциплин вероятностного цикла, что обусловлено неуклонным возрастанием их практической значимости.

Цель данного учебного пособия – привить студентам практические навыки обработки экспериментальных случайных данных с использованием теоретических методов классической математической статистики и современных программных пакетов со встроенными статистическими функциями, а также предоставить студентам методическую поддержку при самостоятельной работе.

Учебное пособие содержит полное методическое обеспечение всех видов учебных занятий по разделу «Математическая статистика» и в его состав входят: краткие теоретические сведения, методические указания по проведению практических занятий, варианты индивидуального задания для расчетно-графической работы или для курсового проекта (в зависимости от действующего учебного плана) и методические указания по его выполнению, примеры выполнения задания с использованием универсальных пакетов MCAD и MATLAB.

1 ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

1.1 Выборка. Эмпирическая функция распределения. Гистограмма. Выборочные числовые характеристики

В математической статистике имеют дело со стохастическими экспериментами, состоящими в проведении повторных независимых наблюдений над некоторой случайной величиной X, имеющей неизвестное распределение вероятностей, т.е. неизвестную функцию распределения $F_X(x) = F(x)$. В этом случае множество всех возможных значений наблюдаемой случайной величины X называют ***генеральной совокупностью***, имеющей функцию распределения $F(x)$. Числа $(x_1,...,x_n)$, являющиеся результатом n повторных независимых наблюдений над случайной величиной X, называют ***выборкой*** из генеральной совокупности или ***выборочными*** (статистическими) данными. Число наблюдений n называется ***объемом*** выборки.

Основная задача математической статистики состоит в том, как по выборке $(x_1,...,x_n)$, извлекая из нее максимум информации, сделать обоснованные выводы относительно вероятностных характеристик наблюдаемой случайной величины X.

Замечание: Выборка $(x_1,...,x_n)$ является исходной информацией для статистического анализа и принятия решений о неизвестных вероятностных характеристиках наблюдаемой случайной величины X. Однако на основе конкретной выборки обосновать качество статистических выводов принципиально невозможно. Для этого на выборку следует смотреть априорно как на ***случайный вектор*** $(X_1,...,X_n)$, координаты которого являются независимыми, распределенными так же как и X, случайными величинами, и который еще не принял конкретного значения в результате эксперимента. Переход от выборки конкретной $(x_1,...,x_n)$ к выборке случайной $(X_1,...,X_n)$ будет неоднократно использоваться далее при решении теоретических вопросов и задач для получения выводов, справедливых для любой выборки из генеральной совокупности.

В зависимости от дальнейших целей существует несколько способов представления статистических данных. Простейший из них - в виде статистического ряда:

Номер наблюдения i	1	2	…	n
Результат наблюдения x_i	x_1	x_2	…	x_n

Если среди выборочных значений имеются совпадающие, то статистический ряд удобнее записывать в виде таблицы, называемой ***таблицей частот***:

Выборочные значения y_i	y_1	y_2	…	y_r
Частоты m_i	m_1	m_2	…	m_r
Относительные частоты $p_i^* = \frac{m_i}{n}$	$p_1^* = \frac{m_1}{n}$	$p_2^* = \frac{m_2}{n}$	…	$p_i^* = \frac{m_i}{n}$

где $(y_1,...,y_r)$, $r \le n$ - различные значения среди $(x_1,...,x_n)$; m_i - частота значения y_i; $p_i^* = \frac{m_i}{n}$ - относительная частота значения y_i. Очевидно, что $\sum_{i=1}^{r} m_i = n$, $\sum_{i=1}^{r} p_i^* = 1$. Поэтому совокупность пар $\left(y_i, p_i^*\right)$, $i = \overline{1,r}$ называют ***эмпирическим законом распределения***.

Выборочные значения $(x_1,...,x_n)$, упорядоченные по возрастанию, носят название ***вариационного ряда***:

$$x_{(1)} \le x_{(2)} \le ... \le x_{(n)},$$

где $x_{(1)} = \min(x_1,...,x_n)$, $x_{(n)} = \max(x_1,...,x_n)$.

Величина $R = x_{(n)} - x_{(1)}$ называется ***размахом выборки***.

Эмпирической функцией распределения, соответствующей выборке $(x_1,...,x_n)$, называется функция

$$F_n^*(x) = \frac{1}{n}\sum_{i=1}^{n} I(x_i < x) = \frac{1}{n}\,\nu_n(x),$$

где $I(A)$ - индикатор множества A, а $\nu_n(x)$ - число выборочных значений, не превосходящих x.

Для заданной выборки $(x_1,...,x_n)$ эмпирическая функция распределения $F_n^*(x)$ обладает всеми свойствами обычной функции распределения: принимает значения между 0 и 1, является неубывающей и непрерывной слева. График $F_n^*(x)$ имеет ступенчатый вид, причем:

если все значения $x_1,...,x_n$ различны, то

$$F_n^*(x) = \frac{k}{n} \text{ при } x \in \left[x_{(k)}, x_{(k+1)}\right), \ x_{(0)} = -\infty, \ x_{(n+1)} = \infty;$$

если $(y_1,...,y_r)$ - различные значения среди $(x_1,...,x_n)$, то

$$F_n^*(x) = \sum_{i:\, y_i < x} \frac{m_i}{n}.$$

Принципиальное отличие эмпирической функции распределения $F_n^*(x)$ от обычной функции распределения состоит в том, что она может изменяться от выборки к выборке и притом случайным образом. Важнейшим свойством эмпирической функции распределения $F_n^*(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i < x)$ как случайной функции (см. замечание выше) является то, что она для любого $x \in (-\infty, \infty)$ при увеличении объема выборки n сближается (в смысле сходимости по вероятности) с истинной функцией распределения $F(x)$. Поэтому говорят, что эмпирическая функция распределения $F_n^*(x)$ является статистическим аналогом (оценкой) неизвестной функции распределения $F(x)$, которую называют при этом теоретической.

Если $(x_1,...,x_n)$ - выборка объема n из генеральной совокупности, имеющей непрерывное распределение с неизвестной плотностью вероятностей $f_X(x) = f(x)$, то для получения статистического аналога $f(x)$ следует предварительно произвести группировку данных. Она состоит в следующем:

1. По данной выборке $(x_1,...,x_n)$ строят вариационный ряд

$$x_{(1)} \le x_{(2)} \le ... \le x_{(n)}.$$

2. Промежуток $\left[x_{(1)}, x_{(n)}\right]$ разбивают точками $x_{(1)} = u_0 < u_1 < \ldots < u_N = x_{(n)}$ на N непересекающихся интервалов $J_k = \left[u_{k-1}, u_k\right)$ (на практике $N << n$).
3. Подсчитывают частоты ν_k попадания выборочных значений в k-ый интервал J_k.
4. Полученную информацию заносят в следующую таблицу, которую называют ***интервальным статистическим рядом:***

Интервалы J_k	$\left[u_0, u_1\right)$	$\left[u_1, u_2\right)$	$\cdots$	$\left[u_{N-1}, u_N\right]$
Частоты ν_k	ν_1	ν_2	$\cdots$	ν_N
Относительные частоты $\tilde{p}_k^* = \dfrac{\nu_k}{n}$	$\tilde{p}_1^* = \dfrac{\nu_1}{n}$	$\tilde{p}_2^* = \dfrac{\nu_2}{n}$	$\cdots$	$\tilde{p}_N^* = \dfrac{\nu_N}{n}$

Очевидно, что $\sum\limits_{k=1}^{N} \nu_i = n, \quad \sum\limits_{k=1}^{N} \tilde{p}_k^* = 1$. Поэтому совокупность пар $(\tilde{u}_k, \tilde{p}_k^*)$, где $\tilde{u}_k = \dfrac{1}{2}(u_k + u_{k-1})$ - середина интервала J_k, $k = \overline{1, N}$, называют ***эмпирическим законом распределения, полученным по сгруппированным данным.***

Далее в прямоугольной системе координат на каждом интервале J_k как на основании длиной $\Delta u_k = u_k - u_{k-1}$ строят прямоугольник с высотой $h_k = \dfrac{\nu_k}{n \cdot \Delta u_k}$, $k = \overline{1, N}$. Получаемую при этом ступенчатую фигуру называют ***гистограммой***.

Поскольку при больших n в соответствии с теоремой Бернулли $\dfrac{\nu_k}{n} \approx p_k$, где $p_k = \mathrm{P}(u_{k-1} \le X < u_k)$ - истинная вероятность попадания случайной величины X в интервал J_k, а $p_k = \int\limits_{u_{k-1}}^{u_k} f(x)dx \approx f(\tilde{u}_k)\Delta u_k$, то справедливо приближенное равенство $h_k \approx f(\tilde{u}_k)$. Поэтому верхняя граница гистограммы является статистическим аналогом (оценкой) неизвестной плотности вероятностей $f(x)$.

На практике при группировке данных обычно берут интервалы одинаковой длины $\Delta u = \text{const}$, а число интервалов группировки определяют с помощью, так называемого, правила Стургерса, согласно которому полагается $N = [1 + 3{,}32 \lg n] + 1$.

Ломаная с вершинами в точках $(\tilde{u}_k, h_k)$ называется ***полигоном частот*** и для гладких плотностей является более точной оценкой, чем гистограмма. Пример гистограммы и полигона частот приведен на рисунке 1.

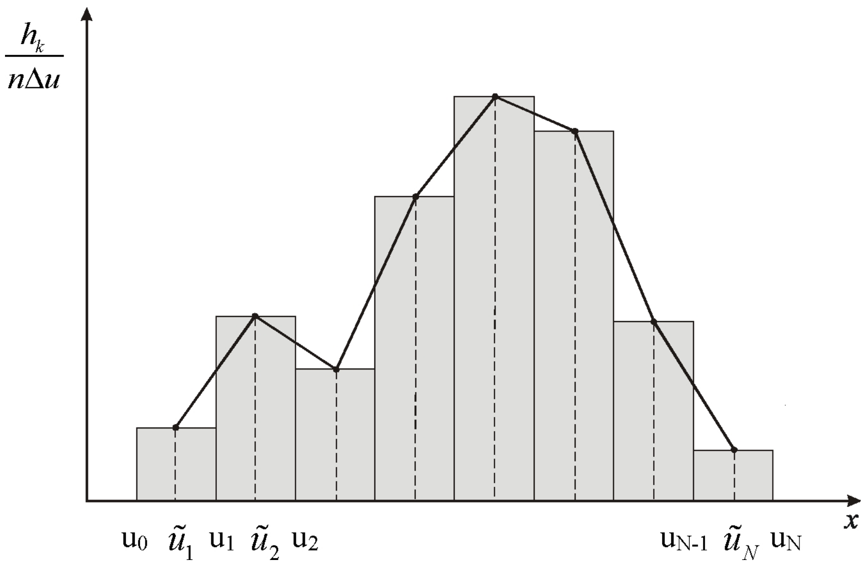

Рисунок 1 - Гистограмма и полигон частот

Пусть $(x_1, ..., x_n)$ - выборка из генеральной совокупности, имеющей функцию распределения $F(x)$. Аналогично тому, как теоретической функции распределения $F(x)$ ставят в соответствие эмпирическую функцию распределения $F_n^*(x)$, любой теоретической характеристике $g = Mg(X) = \int\limits_{-\infty}^{\infty} g(x) dF(x)$ можно поставить в соответствие ее статистический аналог - ***выборочную (эмпирическую) числовую характеристику g****, определяемую как среднее арифметическое значений функции $g(x)$ для элементов выборки $(x_1, ..., x_n)$:

$$g^* = \int\limits_{-\infty}^{+\infty} g(x) dF_n^*(x) = \frac{1}{n} \sum_{i=1}^{n} g(x_i).$$

В частности, выборочный ***начальный момент*** k-го порядка есть величина

$$\alpha_k^* = \frac{1}{n}\sum_{i=1}^{n} x_i^k .$$

При $k=1$ величину α_1^* называют ***выборочным средним*** и обозначают $\overline{x}$:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i .$$

Выборочный ***центральный момент*** k-го порядка есть величина

$$\mu_k^* = \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^k .$$

При $k=2$ величину μ_2^* называют ***выборочной дисперсией*** и обозначают s^2:

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^2 .$$

Между выборочными начальными и выборочными центральными моментами сохраняются те же соотношения, что и между теоретическими. Например, справедливо равенство

$$s^2 = \alpha_2^* - (\overline{x})^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 ,$$

являющееся аналогом известного равенства

$$\mu_2 = \mathrm{D}X = \alpha_2 - \alpha_1^2 = \mathrm{M}X^2 - (\mathrm{M}X)^2 .$$

Являясь для заданной выборки числами, в общем случае выборочные числовые характеристики являются случайными величинами и обозначаются соответствующими заглавными буквами:

$$G^* = \frac{1}{n}\sum_{i=1}^{n} g(X_i);\ A_k^* = \frac{1}{n}\sum_{i=1}^{n} X_i^k;\ M_k^* = \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^k;$$

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i;\ S^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^2 .$$

В связи с этим можно ставить вопрос о нахождении закона распределения выборочных числовых характеристик и их числовых характеристиках.

Располагая только ***сгруппированными*** данными, можно определить аналог эмпирической функции распределения следующим образом:

$$\tilde{F}_n^*(x) = \sum_{k:\,\tilde{u}_k < x} \frac{\nu_k}{n}.$$

Для вычисления выборочных моментов k-го порядка по сгруппированным данным используются формулы:

$$\tilde{\alpha}_k^* = \frac{1}{n}\sum_{i=1}^{N} \tilde{u}_i^k \nu_i\,, \qquad \tilde{\mu}_k^* = \frac{1}{n}\sum_{i=1}^{N}\left(\tilde{u}_i - \tilde{\alpha}_1^*\right)^k \nu_i\,.$$

В частности, выборочное среднее и выборочная дисперсия по сгруппированным данным определяются с помощью формул:

$$\tilde{x} = \frac{1}{n}\sum_{i=1}^{N} \tilde{u}_i \nu_i\,, \qquad \tilde{s}^2 = \frac{1}{n}\sum_{i=1}^{N}\left(\tilde{u}_i - \tilde{x}\right)^2 \nu_i\,.$$

1.2 Оценивание неизвестных параметров распределений

Пусть имеется выборка $(x_1,...,x_n)$, представляющая собой результат n независимых наблюдений над некоторой случайной величиной X, и предположим, что тип распределения генеральной совокупности известен, но зависит от неизвестного параметра: $F_X(x) = F(x;\theta),\ \ \theta \in \Theta$. В общем случае задача оценивания формулируется так: используя информацию, доставляемую выборкой, сделать статистические выводы об истинном значении неизвестного параметра θ, т.е. оценить параметр θ.

Различают точечные и интервальные оценки неизвестных параметров.

1.2.1 Точечные оценки. Методы нахождения точечных оценок

При точечном оценивании ищут ***статистику*** $\theta^* = \theta^*(x_1,...,x_n)$, (т.е. функцию, зависящую только от выборки $(x_1,...,x_n)$), значение которой при заданной выборке принимают за приближенное значение параметра θ. В этом случае статистику $\theta^* = \theta^*(x_1,...,x_n)$ называют ***оценкой*** параметра θ.

Обосновать качество оценки θ^* можно лишь исходя из ее свойств, не зависящих от конкретной выборки. Для изучения таких свойств (естественно, вероятностного характера) в соответствии с замечанием из п. 1.1. под оценкой следует понимать ***случайную величину*** $\theta^* = \theta^*(X_1,...,X_n)$. Выбор из множества оценок одного и того же параметра наилучшей основан на критерии сравнения качества оценок, предложенном Р.А.Фишером. Согласно этому критерию оценка $\theta^* = \theta^*(X_1,...,X_n)$ должна быть:

1) ***состоятельной***, т. е. с возрастанием объема выборки n должна сходиться по вероятности к истинному неизвестному значению параметра θ: $\theta^* = \theta^*(X_1,...,X_n) \xrightarrow{\mathrm{P}} \theta$;
2) ***несмещенной***, т. е. математическое ожидание θ^* должно быть равно оцениваемому параметру θ: $\mathrm{M}\theta^* = \theta$;
3) ***эффективной***, т. е. должна обладать минимальной дисперсией в рассматриваемом классе оценок.

Величина $b(\theta^*) = \mathrm{M}\theta^* - \theta$ называется ***смещением*** оценки θ^*. Таким образом, оценка θ^* является несмещенной тогда и только тогда, когда ее смещение $b(\theta^*) = 0$. Оценка θ^*, у которой $b(\theta^*) \to 0$ при $n \to \infty$, называется ***асимптотически несмещенной***.

Достаточным условием состоятельности несмещенной оценки в силу неравенства Чебышева является стремление к нулю ее дисперсии:

$$\mathrm{D}\theta^* \to 0 \text{ при } n \to \infty.$$

Эффективность оценки θ^* позволяет исследовать следующее неравенство Рао-Крамера: для широкого класса непрерывных распределений и для любой несмещенной оценки θ^*, имеющей конечную дисперсию, справедливо неравенство:

$$\mathrm{D}\theta^* \geq \frac{1}{n \cdot I(\theta)} = \frac{1}{n \cdot \mathrm{M}\left(\frac{\partial \ln f(X,\theta)}{\partial \theta}\right)^2},$$

где $f(x;\theta)$ - плотность вероятностей наблюдаемой случайной величины X, $I(\theta) = \mathrm{M}\left(\frac{\partial \ln f(X,\theta)}{\partial \theta}\right)^2$ - информация Фишера о параметре θ, содержащаяся в одном наблюдении над случайной величиной X.

Таким образом, оценка θ^* является эффективной, если она обращает неравенство Рао-Крамера в равенство, т.е. $\mathrm{D}\theta^* = \dfrac{1}{n \cdot I(\theta)}$.

Наиболее распространенными методами получения точечных оценок неизвестных параметров распределений, удовлетворяющих требованиям 1 – 3 (хотя бы частично), являются метод моментов и метод максимального правдоподобия.

Метод моментов. Пусть $(x_1,...,x_n)$ - выборка из генеральной совокупности, имеющей функцию распределения $F_X(x)=F(x;\theta)$, зависящую от векторного параметра $\theta=(\theta_1,...,\theta_r)$. Предположим, что у наблюдаемой случайной величины X существуют первые r моментов $\alpha_k=\mathrm{M}X^k$, $k=\overline{1,r}$, которые являются функциями от θ: $\alpha_k=\alpha_k(\theta_1,...,\theta_r)$. Метод моментов состоит в нахождении решения $\tilde{\theta}=(\tilde{\theta}_1,...,\tilde{\theta}_r)$ системы уравнений, получаемой приравниванием теоретических моментов соответствующим выборочным моментам:

$$\alpha_k(\theta_1,...,\theta_r)=\alpha_k^*,\ \ k=\overline{1,r}.$$

Для нахождения оценки $\tilde{\theta}=(\tilde{\theta}_1,...,\tilde{\theta}_r)$ может быть использована также система уравнений, основанных на приравнивании центральных теоретических и выборочных моментов:

$$\mu_k(\theta_1,...,\theta_r)=\mu_k^*,\ \ k=\overline{1,r}.$$

Использование именно первых r моментов является необязательным.

В случае двумерного неизвестного параметра $\theta=(\theta_1,\theta_2)$ его оценка по методу моментов $\tilde{\theta}=(\tilde{\theta}_1,\tilde{\theta}_2)$ обычно определяется как решение системы уравнений: $\begin{cases}\mathrm{M}X=\bar{x},\\ \mathrm{D}X=s^2.\end{cases}$

Оценки, получаемые по методу моментов являются:
- состоятельными (при весьма общих предположениях);
- несмещенными не всегда;
- вообще говоря, неэффективными.

На практике оценки, получаемые по методу моментов, часто используются как первое приближение, на основе которого находятся более «хорошие» оценки.

Достоинство метода моментов заключается в том, что системы уравнений для нахождения оценок решаются довольно просто. Однако имеет место произвол в выборе уравнений для нахождения оценок и метод вообще неприменим, когда моментов необходимого порядка не существует (пример, - закон распределения Коши).

Метод максимального правдоподобия. Пусть $(x_1,...,x_n)$ - выборка из генеральной совокупности, имеющей функцию распределения $F_X(x) = F(x;\theta)$, зависящую от неизвестного скалярного параметра θ.

Если закон распределения наблюдаемой случайной величины X является непрерывным, т.е. существует плотность вероятностей $f(x;\theta)$, то функция

$$L(x_1,...,x_n;\theta) = f(x_1;\theta)\cdot ... \cdot f(x_n;\theta),$$

рассматриваемая при фиксированной выборке $x_1,...,x_n$ как функция параметра θ, называется ***функцией правдоподобия***.

Если наблюдаемая случайная величина X имеет дискретный закон распределения, задаваемый вероятностями $\mathrm{P}(X = x_i) = p(x_i;\theta),\ i = \overline{1,n}$, то функция правдоподобия определяется равенством:

$$L(x_1,...,x_n;\theta) = p(x_1;\theta)\cdot ... \cdot p(x_n;\theta).$$

Оценкой максимального правдоподобия $\hat{\theta}$ параметра θ называется такое значение параметра, при котором функция правдоподобия при заданной выборке $(x_1,...,x_n)$ достигает максимума:

$$L(x_1,...,x_n;\hat{\theta}) = \max_{\theta} L(x_1,..,x_n;\theta).$$

Если функция правдоподобия дифференцируема по θ, то оценку максимального правдоподобия $\hat{\theta}$ можно найти, решив относительно θ ***уравнение правдоподобия***

$$\frac{\partial L(x_1,...,x_n;\theta)}{\partial \theta} = 0$$

или равносильное уравнение

$$\frac{\partial \ln L(x_1,...,x_n;\theta)}{\partial \theta} = 0.$$

Если $\theta = (\theta_1,...,\theta_r)$ - векторный параметр, то для отыскания оценки максимального правдоподобия $\hat{\theta} = (\hat{\theta}_1,...,\hat{\theta}_r)$ следует решить ***систему уравнений правдоподобия***

$$\frac{\partial \ln L(x_1,...,x_n;\theta_1,...,\theta_r)}{\partial \theta_i} = 0,\quad i = \overline{1,r}.$$

Все изложенные результаты остаются в силе и при оценивании не самого параметра θ, а некоторой параметрической функции $\tau(\theta)$.

Оценки максимального правдоподобия являются:
- состоятельными;
- асимптотически эффективными;
- несмещенными не всегда;
- асимптотически нормальными, т.е. при соответствующей нормировке закон распределения оценки максимального правдоподобия является нормальным (что очень важно для нахождения вероятностей отклонения их от истинных значений параметров).

Однако уравнения (системы уравнений) для нахождения оценок максимального правдоподобия могут решаться довольно сложно.

1.2.2 Интервальные оценки

На практике ограничиться нахождением «хороших» точечных оценок бывает обычно недостаточно. Приближенное равенство $\theta^* \approx \theta$ лишь указывает на то, что вместо неизвестного параметра θ можно использовать известное значение оценки θ^*. Однако важно знать (хотя бы в вероятностном смысле) величину совершаемой при этом ошибки. Для этого прибегают к построению интервальных оценок неизвестных параметров.

Пусть наблюдаемая величина X имеет функцию распределения $F(x;\theta)$, зависящую от неизвестного параметра θ. При интервальном оценивании параметра θ ищут две такие статистики $T_1 = T_1(X_1,...,X_n)$ и $T_2 = T_2(X_1,...,X_n)$ (T_1 и T_2 - случайные величины!), для которых при заданном $\gamma \in (0,1)$ выполняется соотношение

$$\mathrm{P}(T_1 < \theta < T_2) = \gamma .$$

В этом случае интервал $\Delta_\gamma(\theta) = (T_1, T_2)$ называют ***γ-доверительньм интервалом*** для параметра θ, число γ - ***доверительной вероятностью*** (надежностью, коэффициентом доверия), T_1 и T_2 - нижней и верхней ***доверительными границами*** соответственно.

Таким образом, γ-доверительный интервал — это ***случайный*** интервал, зависящий от выборки (но не от θ), который содержит (накрывает) истинное значение неизвестного параметра θ с вероятностью γ. На практике обычно используют значения доверительной вероятности γ из небольшого набора близких к 1 значений (0,9; 0,95; 0,98; 0,99 и т. д.) и строят соответствующие им доверительные интервалы.

Построение доверительных интервалов для отдельных параметров распределения генеральной совокупности зависит как от вида закона распределения, так и от того, являются известными значения остальных параметров распределения или нет.

- Если наблюдаемая случайная величина X имеет нормальный закон распределения $N(\theta,\sigma^2)$ с неизвестным математическим ожиданием θ и ***известной*** дисперсией σ^2, то ***доверительный интервал для математического ожидания*** θ имеет вид:

$$\Delta_\gamma(\theta)=\left(\bar{X}-c_{(1+\gamma)/2}\cdot\frac{\sigma}{\sqrt{n}}\ ,\ \bar{X}+c_{(1+\gamma)/2}\cdot\frac{\sigma}{\sqrt{n}}\right),$$

где $\bar{X}$ - выборочное среднее; n - объем выборки; число $c_{(1+\gamma)/2}$ - такое значение аргумента функции Лапласа $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x}e^{-u^2/2}\cdot du,$ при котором $\Phi(c_{(1+\gamma)/2})=(1+\gamma)/2$. Находят число $c_{(1+\gamma)/2}$ по заданной доверительной вероятности γ из таблицы Б.2.

Квантилью, соответствующей вероятности p, называется такое значение x_p, при котором выполняется соотношение

$$F(x_p)=P\left(X<x_p\right)=\int\limits_{-\infty}^{x_p}f(x)dx=p,$$

где $f\left(x\right)$ – плотность вероятностей соответствующего закона распределения (слово квантиль – женского рода). Геометрическое пояснение смысла квантили, отвечающей вероятности p, приведено на рисунке 2.

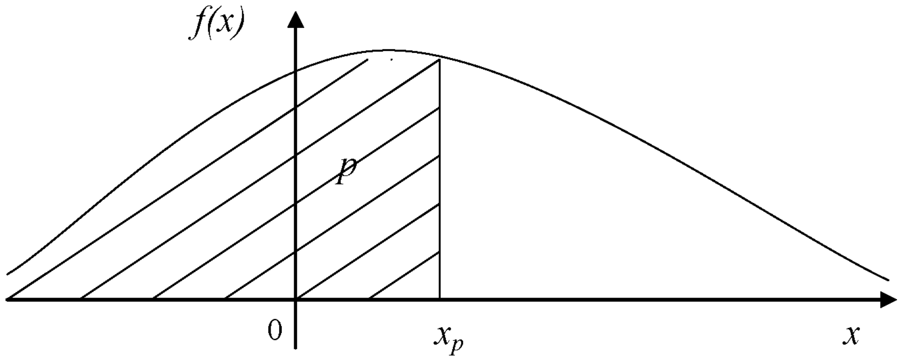

Рисунок 2 - Геометрическое пояснение смысла квантили x_p, отвечающей вероятности p

В этой терминологии число $c_{(1+\gamma)/2}$ есть $(1+\gamma)/2$ - квантиль стандартного нормального $N(0,1)$ закона распределения.

- Если наблюдаемая случайная величина X имеет нормальный закон распределения $N(\theta_1,\theta_2^2)$ с неизвестным математическим ожиданием θ_1 и ***неизвестной*** дисперсией θ_2^2, то ***доверительный интервал для математического ожидания*** θ_1 имеет вид:

$$\Delta_\gamma(\theta_1)=\left(\bar{X}-t_{(1+\gamma)/2;n-1}\cdot\frac{S}{\sqrt{n-1}}\ ;\ \bar{X}+t_{(1+\gamma)/2;n-1}\cdot\frac{S}{\sqrt{n-1}}\right),$$

где S^2 - выборочная дисперсия; $S=\sqrt{S^2}$; n - объем выборки; число $t_{(1+\gamma)/2;n-1}$ – $(1+\gamma)/2$ - квантиль распределения Стьюдента $S(n-1)$ с $(n—1)$ степенью свободы. Находят квантиль $t_{(1+\gamma)/2;n-1}$ по заданным γ и n из таблицы В.1.

При больших n (практически при $n\geq 30$) распределение Стьюдента приближается (в смысле слабой сходимости) к стандартному нормальному закону распределения, поэтому в этом случае $t_{(1+\gamma)/2;n-1}\approx c_{(1+\gamma)/2}$.

- ***Доверительный интервал для дисперсии*** θ^2 наблюдаемой случайной величины X, распределенной по нормальному закону $N(a,\theta^2)$, при ***известном*** математическом ожидании $\mathrm{M}X=a$ имеет вид:

$$\Delta_\gamma(\theta^2)=\left(\frac{\sum_{i=1}^{n}(X_i-a)^2}{\chi^2_{(1+\gamma)/2;n}}\ ;\ \frac{\sum_{i=1}^{n}(X_i-a)^2}{\chi^2_{(1-\gamma)/2;n}}\right),$$

где числа $\chi^2_{(1-\gamma)/2;n}$ и $\chi^2_{(1+\gamma)/2;n}$ есть $(1-\gamma)/2$- и $(1+\gamma)/2$- квантили распределения хи - квадрат $\chi^2(n)$ с n степенями свободы соответственно. Квантили распределения хи - квадрат находят по заданным γ и n из таблицы Г.1.

- ***Доверительный интервал для дисперсии*** θ_2^2 наблюдаемой случайной величины X, распределенной по нормальному закону $N(\theta_1,\theta_2^2)$, при ***неизвестном*** математическом ожидании $\mathrm{M}X=\theta_1$ имеет вид:

$$\Delta_\gamma(\theta_2^2) = \left(\frac{n \cdot S^2}{\chi^2{}_{(1+\gamma)/2;\,n-1}} ; \frac{n \cdot S^2}{\chi^2{}_{(1-\gamma)/2;\,n-1}} \right),$$

где S^2 - выборочная дисперсия, а $\chi^2{}_{(1-\gamma)/2}(n); \chi^2{}_{(1+\gamma)/2}(n)$ – соответствующие квантили распределения $\chi^2(n-1)$.

При больших n (практически при $n \geq 30$) с использованием центральной предельной теоремы можно показать, что ***приближенным (асимптотическим) доверительным интервалом для дисперсии*** θ_2^2 нормально распределенной $N(\theta_1, \theta_2^2)$ случайной величины X с неизвестным математическим ожиданием $\mathrm{M}X = \theta_1$ является интервал

$$\Delta_\gamma(\theta_2^2) = \left(\frac{n \cdot S^2}{n - 1 + c_{(1+\gamma)/2}\sqrt{2(n-1)}} \; ; \; \frac{n \cdot S^2}{n - 1 - c_{(1+\gamma)/2}\sqrt{2(n-1)}} \right).$$

Фактически это означает, что для квантилей распределения хи - квадрат $\chi^2{}_{(1-\gamma)/2}(n-1)$ и $\chi^2{}_{(1+\gamma)/2}(n-1)$ при $n \geq 30$ имеют место приближенные формулы:

$$\chi^2{}_{(1-\gamma)/2;\,n-1} \approx n - 1 - c_{(1+\gamma)/2} \cdot \sqrt{2(n-1)} \; ;$$

$$\chi^2{}_{(1+\gamma)/2;\,n-1} \approx n - 1 + c_{(1+\gamma)/2} \cdot \sqrt{2(n-1)} \; .$$

Если распределение наблюдаемой случайной величины X произвольное (не обязательно нормальное), то, используя асимптотическую нормальность выборочных моментов, можно показать, что при больших объемах выборки ***приближенными (асимптотическими) доверительными интервалами для математического ожидания*** $\mathrm{M}X = a$ ***и дисперсии*** $\mathrm{D}X = \sigma^2$ являются:

$$\Delta_\gamma(a) = \left(\bar{X} - c_{(1+\gamma)/2} \cdot \frac{S}{\sqrt{n}} \; ; \; \bar{X} + c_{(1+\gamma)/2} \cdot \frac{S}{\sqrt{n}} \right),$$

$$\Delta_\gamma(\sigma^2) = \left(S^2 - c_{(1+\gamma)/2} \frac{\sqrt{M_4^* - S^4}}{\sqrt{n}} \; ; \; S^2 + c_{(1+\gamma)/2} \frac{\sqrt{M_4^* - S^4}}{\sqrt{n}} \right),$$

где $\bar{X}$ - выборочное среднее; S^2 - выборочная дисперсия; $S = \sqrt{S^2}$; $M_4^* = \dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i - \bar{X})^4$ - выборочный центральный момент четвертого порядка.

Замечание: Все приведенные доверительные интервалы, рассчитанные для заданной выборки $(x_1,...,x_n)$, являются обычными числовыми интервалами, внутри которых неизвестный параметр находится в $\gamma \cdot 100\%$ случаев.

1.3 Проверка статистических гипотез

Статистической гипотезой называют любое утверждение о виде или свойствах наблюдаемых в эксперименте случайных величин. Правило, позволяющее по имеющимся статистическим данным (выборке) принять или отклонить выдвинутую гипотезу, называется ***статистическим критерием***. Если формулируется только одна гипотеза H_0 и требуется проверить, согласуются ли статистические данные с этой гипотезой или же они ее опровергают, то критерии, используемые для этого, называются ***критериями согласия***. Если гипотеза H_0 однозначно фиксирует закон распределения наблюдаемой случайной величины, то она называется простой, в противном случае — сложной. Пусть относительно наблюдаемой случайной величины X сформулирована некоторая гипотеза H_0; $(x_1,...,x_n)$ - выборка объема n, являющаяся реализацией случайного вектора $(X_1,...,X_n)$, координаты которого X_i, $i=\overline{1,n}$ независимы и распределены так же, как X.

Общий метод построения критерия согласия для проверки гипотезы H_0 состоит в следующем. Вначале ищут статистику $T = T(X_1,...,X_n)$ (случайную величину!), характеризующую отклонение эмпирического распределения от теоретического, закон распределения которой в случае справедливости H_0 можно определить (точно или приближенно). Далее задают некоторое положительное малое число α, так что событие с вероятностью α можно считать практически невозможным в данном эксперименте. Затем для заданного α определяют в множестве $K = \{t : t = T(x_1,...,x_n)\}$ возможных значений статистики T подмножество K_α, так чтобы $\mathrm{P}(T(X_1,...,X_n) \in K_\alpha / H_0) \le \alpha$.

Критерий согласия имеет следующий вид:

- если $t = T(x_1,...,x_n)$ значение статистики $T(X_1,...,X_n)$, соответствующее данной выборке $(x_1,...,x_n)$ и $t \in K_\alpha$, то гипотеза H_0 отвергается;

- если $t \notin K_\alpha$, то гипотеза H_0 принимается.

Статистика $T = T(X_1,...,X_n)$ называется ***статистикой критерия***; множество K_α - ***критической областью*** для гипотезы H_0, число α - ***уровнем значимости*** критерия.

1.3.1 Проверка гипотезы о виде распределения

Пусть $(x_1,...,x_n)$ - выборка объема n, представляющая собой результат n независимых наблюдений над случайной величиной X, относительно распределения которой выдвинута простая гипотеза $H_0 : F_X(x) = F(x)$ ($F(x)$ - теоретическая функция распределения, соответствующая гипотезе H_0). Наиболее распространенным критерием проверки этой гипотезы H_0 является критерии χ^2 Пирсона.

Чтобы воспользоваться критерием χ^2 Пирсона, выборочные данные $(x_1,...,x_n)$ следует предварительно сгруппировать, представив их в виде интервального статистического ряда. Пусть $J_k = [u_{k-1}, u_k)$, $k = \overline{1,N}$ - интервалы группировки; $\nu_1,...,\nu_N$ - частоты попадания выборочных значений в интервалы $J_1,...,J_N$ соответственно ($\nu_1 + ... + \nu_N = n$).

Обозначим p_k теоретическую (соответствующую H_0) вероятность попадания случайной величины X в интервал J_k :

$$p_k = \mathrm{P}(u_{k-1} \le X < u_k) = F(u_k) - F(u_{k-1}),\ k = \overline{1,N}.$$

Статистикой критерия χ^2 является величина:

$$\chi_n^2 = \sum_{k=1}^{N} \frac{(\nu_k - n p_k)^2}{n p_k},$$

которая характеризует отклонение эмпирической функции распределения $\tilde{F}_n^*(x)$ от теоретической функции распределения $F(x)$ (значение $\dfrac{\nu_k}{n}$ является приращением эмпирической функции $\tilde{F}_n^*(x)$ на интервале J_k, а p_k - приращением теоретической функции $F(x)$ на том же интервале). Поскольку относительные частоты $\frac{\nu_k}{n}$ сближаются с вероятностями p_k при $n \to \infty$, $k = \overline{1, N}$, то в случае справедливости гипотезы H_0 значение величины χ_n^2 не должно существенно отличаться от нуля. Поэтому критическая область критерия χ^2 задается в виде $K_\alpha = \{t \geq t_\alpha\}$, где $t = \chi_n^2(x_1, ..., x_n)$ – значение величины χ_n^2, полученное для заданной выборки, а порог t_α определяется по заданному уровню значимости α так, чтобы $\mathrm{P}(\chi_n^2 \in K_\alpha / H_0) = \alpha$. Нахождение t_α основано на том факте (известном как теорема Пирсона), что случайная величина χ_n^2 имеет при $n \to \infty$ предельное распределение хи-квадрат $\chi^2(N-1)$ с $N-1$ степенью свободы.

На практике предельное распределение $\chi^2(N-1)$ можно использовать с хорошим приближением при $n \geq 50$ и $\nu_k \geq 5$, $k = \overline{1, N}$. При выполнении этих условий для заданного уровня значимости α можно положить $t_\alpha = \chi^2_{1-\alpha; N-1}$, где $\chi^2_{1-\alpha; N-1}$ является (1—α)-квантилью распределения $\chi^2(N-1)$.

Таким образом, критерий согласия χ^2 Пирсона состоит в следующем:

1. По заданному уровню значимости α находится по таблице Г.1 порог $\chi^2_{1-\alpha; N-1}$.
2. По заданной выборке $(x_1, ..., x_n)$ объема $n \geq 50$ определяется число N интервалов группировки так, чтобы $\nu_k \geq 5$, $k = \overline{1, N}$. Вычисляется значение статистики $\chi_n^2(x_1, ..., x_n) = t$.
3. Если $t \geq \chi^2_{1-\alpha; N-1}$, то гипотезу H_0 отвергают.
4. Если $t < \chi^2_{1-\alpha; N-1}$, то гипотезу H_0 принимают.

Если случайная величина X дискретная, x_k, $k = \overline{1, N}$ - различные выборочные значения, а $\mathrm{P}(X = x_k) = p_k$ в случае справедливости H_0, то всегда можно определить N интервалов, содержащих ровно по одному выборочному значению. Поэтому в данном случае можно сразу считать, что $\nu_k = m_k$, $k = \overline{1,N}$, где m_k – частота выборочного значения x_k.

На практике теоретическое распределение полностью бывает определено редко. Чаще известен предположительно только тип распределения, но неизвестны параметры его определяющие. В этом случае гипотеза о виде распределения, подлежащая проверке, имеет вид $H_0: F_X(x) = F(x;\theta)$, $\theta = (\theta_1, ..., \theta_r) \in \Theta$ и является сложной параметрической гипотезой.

Критерий согласия χ^2 Пирсона применим для проверки такой гипотезы H_0 со следующими изменениями:

а) вероятности $p_k = \mathrm{P}(X \in J_k)$, $k = \overline{1,N}$ вычисляют, заменяя неизвестные параметры $\theta = (\theta_1, ..., \theta_r)$ их оценками максимального правдоподобия $\hat{\theta} = (\hat{\theta}_1, ..., \hat{\theta}_r)$: $p_k = F(u_k;\hat{\theta}) - F(u_{k-1};\hat{\theta})$, $k = \overline{1,N}$;

б) число степеней свободы предельного распределения хи - квадрат должно быть уменьшено на число неизвестных параметров и считаться равным $(N-1-r)$.

1.4 Изучение зависимости между случайными величинами

Предположим, что случайный эксперимент состоит в проведении повторных независимых наблюдений над случайным вектором (X,Y), имеющим неизвестное распределение вероятностей, т.е. неизвестную двумерную функцию распределения $F_{XY}(x,y) = F(x,y)$. В этом случае выборка объема n представляет собой множество пар чисел $((x_1, y_1), ..., (x_n, y_n))$ и она является исходной информацией для получения статистических выводов о вероятностных характеристиках случайного вектора.

При рассмотрении одновременно двух случайных величин X и Y наряду с одномерными статистическими задачами дополнительно возникают статистические задачи, связанные с изучением зависимости между случайными величинами. К ним, в частности, относятся: определение корреляционной зависимости между случайными величинами X и Y; проверка гипотезы о независимости случайных величин; построение регрессионной зависимости между случайными величинами X и Y.

1.4.1 Оценка коэффициента корреляции.

Точечными оценками математических ожиданий и дисперсий координат вектора X и Y в соответствии с разделом 1.1 являются выборочные средние и выборочные дисперсии:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i , \qquad \overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i ,$$

$$s_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 , \; s_Y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2 .$$

Аналогичным усреднением находится и оценка корреляционного момента $\mathrm{R}_{XY} = \mathrm{M}(X - \mathrm{M}X)(Y - \mathrm{M}Y) = \mathrm{M}(X \cdot Y) - \mathrm{M}X \cdot \mathrm{M}Y$, называемая ***выборочным корреляционным моментом***:

$$\hat{R}_{XY} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{x}\cdot\overline{y} .$$

С учетом этого оценка коэффициента корреляции $\mathrm{r}_{XY} = \dfrac{\mathrm{R}_{XY}}{\sqrt{\mathrm{D}X \cdot \mathrm{D}Y}}$, называемая ***выборочным коэффициентом корреляции***, определяется по формуле:

$$\hat{r}_{XY} = \frac{\hat{R}_{XY}}{\sqrt{s_X^2 \cdot s_Y^2}} = \frac{\sum\limits_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum\limits_{i=1}^{n}(x_i - \overline{x})^2 \cdot \sum\limits_{i=1}^{n}(y_i - \overline{y})^2}} =$$

$$= \frac{\sum\limits_{i=1}^{n} x_i y_i - n \cdot \overline{x} \cdot \overline{y}}{\sqrt{\left(\sum\limits_{i=1}^{n} x_i^2 - n \cdot \overline{x}^2\right)\left(\sum\limits_{i=1}^{n} y_i^2 - n \cdot \overline{y}^2\right)}} .$$

1.4.2 Проверка гипотезы о независимости

В общем случае для проверки гипотезы о независимости случайных величин X и Y можно воспользоваться критерием независимости χ^2 проверки гипотезы H_0, заключающейся в том, что функция распределения случайного вектора (X, Y)

$$F(x,y) = F_X(x) F_Y(y),$$

где $F_X(x)$ и $F_Y(y)$ - одномерные функции распределения координат вектора.

Статистика критерия независимости χ^2 имеет вид (см. [3, разд. 3.5]):

$$\chi_n^2 = n \cdot \sum_{i=1}^{K} \sum_{j=1}^{L} \frac{\left(\nu_{ij} - (\nu_i^X \nu_j^Y)/n\right)^2}{\nu_i^X \nu_j^Y} = n \cdot \left(\sum_{i=1}^{K} \sum_{j=1}^{L} \frac{\nu_{ij}^2}{\nu_i^X \nu_j^Y} - 1 \right),$$

где K и L – число интервалов группировки выборочных значений случайных величин X и Y соответственно; ν_i^X и ν_j^Y - частоты интервалов группировки выборочных значений случайных величин X и Y соответственно; ν_{ij} - частота прямоугольника, сторонами которого являются i-й интервал группировки выборочных значений случайной величины X и j-й интервал группировки выборочных значений случайной величины Y.

Гипотезу H_0 отвергают тогда, когда вычисленное по заданной выборке $((x_1, y_1), \ldots, (x_n, y_n))$ значение статистики χ_n^2 удовлетворяет неравенству $\chi_n^2 \geq \chi^2_{1-\alpha;(K-1)(L-1)}$, где $\chi^2_{1-\alpha;(K-1)(L-1)}$ является $(1-\alpha)$-квантилью распределения $\chi^2((K-1)(L-1))$ с $(K-1)(L-1)$ степенями свободы. В противном случае гипотезу H_0 принимают.

В случае ***нормального*** распределения случайного вектора (X, Y) равенство коэффициента корреляции нулю означает одновременно и независимость координат вектора. Поэтому гипотеза о независимости случайных величин X и Y в этом случае может быть сформулирована как гипотеза $Í_0 : \mathrm{r}_{XY} = 0$.

Статистикой критерия для проверки данной гипотезы H_0 является величина: $T = \dfrac{\hat{r}_{XY} \cdot \sqrt{n-2}}{\sqrt{1 - \hat{r}_{XY}^2}}$, где $\hat{r}_{XY}$ - выборочный коэффициент корреляции.

В случае справедливости H_0 значение величины T не должно существенно отличаться по модулю от нуля. Поэтому критическая область критерия для проверки H_0 является двусторонней (в отличие от критерия χ^2) и задается в виде $K_\alpha = \{ |t| \geq t_\alpha \}$, где $t = T((x_1, y_1), \ldots, (x_n, y_n))$ - значение величины T, полученное для заданной выборки, а порог t_α определяется по заданному уровню значимости α так, чтобы $\mathrm{P}(T \in K_\alpha / H_0) = \alpha$.

Поскольку (см. [3, разд. 3.5] и [4, разд. 4.8.1]) при больших n (практически при $n \geq 30$) в случае справедливости гипотезы H_0 случайная величина T имеет распределение Стьюдента $S(n-2)$ с $(n-2)$ степенями свободы, то для заданного уровня значимости α можно положить $t_\alpha = t_{1-\alpha/2;n-2}$, где $t_{1-\alpha/2;n-2}$ является 1—α/2 - квантилью распределения $S(n-2)$.

Таким образом, критерий для проверки гипотезы H_0 о равенстве нулю коэффициента корреляции состоит в следующем:

1. По заданному уровню значимости α находится по таблице В.1 порог $t_{1-\alpha/2;n-2}$.
2. По заданной выборке $((x_1, y_1),...,(x_n, y_n))$ вычисляется значение статистики $T((x_1, y_1),...,(x_n, y_n)) = t$.
3. Если $|t| \geq t_{1-\alpha/2;n-2}$, то гипотезу H_0 отвергают и делают вывод о том, что случайные величины X и Y являются зависимыми.
4. Если $|t| < t_{1-\alpha/2;n-2}$, то гипотезу H_0 принимают и считают, что случайные величины X и Y являются независимыми.

Замечание. В общем случае (отличном от нормального) гипотеза $H_0 : \mathrm{r}_{XY} = 0$ является гипотезой о некоррелированности случайных величин X и Y и известна также как гипотеза о значимости коэффициента корреляции.

1.4.3 Эмпирические уравнения регрессии.

Функцией регрессии $\psi(x)$ случайной величины Y на случайную величину X называется условное математическое ожидание $\mathrm{M}(Y / X = x)$. Эта функция наилучшим (в среднеквадратическом смысле) образом описывает зависимость случайной величины Y от случайной величины X.

Известно, что если случайный вектор (X, Y) имеет двумерный нормальный закон распределения $N(a_X, a_Y, \sigma_X^2, \sigma_Y^2, \mathrm{r}_{XY})$, то функция регрессии случайной величины Y на случайную величину X является линейной и имеет вид (случай нормальной регрессии):

$$y = a_Y + \mathrm{r}_{XY}\frac{\sigma_Y}{\sigma_X}(x - a_X).$$

Заменяя в этом уравнении $a_X, a_Y, \sigma_X, \sigma_Y, \mathrm{r}_{XY}$ на их точечные оценки $\overline{x}, \overline{y}, s_X, s_Y, \hat{r}_{XY}$ соответственно, получаем эмпирическое уравнение регрессии случайной величины Y на случайную величину X вида (подробнее см. [4, разд. 4.8.2]):

$$y - \overline{y} = \hat{r}_{XY}\frac{s_Y}{s_X}(x - \overline{x}).$$

Аналогично определяется функция регрессии $\varphi(y) = \mathrm{M}(X / Y = y)$ случайной величины X на случайную величину Y. При этом эмпирическое уравнение регрессии случайной величины X на случайную величину Y в нормальном случае имеет вид:

$$x - \overline{x} = \hat{r}_{XY}\frac{s_X}{s_Y}(y - \overline{y}).$$

Геометрически уравнение регрессии представляет собой прямую, около которой группируются значения случайного вектора (X, Y). Чем ближе значение выборочного коэффициента корреляции к 1, тем плотнее значения вектора (X, Y) располагаются вдоль прямой регрессии.

1.5 Моделирование случайных величин и векторов

1.5.1 Моделирование непрерывных случайных величин

Случайные величины обычно моделируют с помощью преобразований одного или нескольких независимых значений случайной величины U, равномерно распределенной на отрезке $[0,1]$, которые называют равномерными случайными числами. Неограниченный источник, вырабатывающий последовательность $u_1, u_2, \ldots$ равномерных случайных чисел, называется датчиком случайных чисел. В ряде ЭВМ имеются встроенные физические датчики случайных чисел, принцип работы которых основан на использовании в качестве источников случайности быстро флуктуирующих шумовых процессов, вырабатываемых радиоэлементами. Однако чаще всего в качестве равномерных случайных чисел используют так называемые псевдослучайные числа, т.е. детерминированные числа, получаемые по некоторому алгоритму и обладающие в той или иной мере свойствами случайных чисел. «Качество» псевдослучайных чисел проверяется при помощи статистических критериев проверки их случайности и равномерной распределенности. Известны различные методы получения псевдослучайных чисел, и соответствующие программы входят в математическое обеспечение современных ЭВМ.

Стандартный метод моделирования непрерывных случайных величин

Пусть U - равномерно распределенная на отрезке $\left[0,1\right]$ случайная величина, то есть

$$f_U\left(u\right)=\begin{cases}1, & u\in\left[0,1\right],\\ 0, & u\notin\left[0,1\right],\end{cases}\qquad F_U\left(u\right)=\begin{cases}0, & u<0,\\ u, & u\in\left[0,1\right],\\ 1, & u>1,\end{cases}$$

X - моделируемая случайная величина с непрерывной и строго монотонной функцией распределения $F\left(x\right)$; $F^{-1}\left(y\right)$ - функция, обратная к $F\left(x\right)$. Тогда случайная величина $F^{-1}(U)$ распределена также как X, поскольку

$$\mathrm{P}\left(F^{-1}(U)<x\right)=\mathrm{P}\left(U<F\left(x\right)\right)=F_U\left(F\left(x\right)\right)=F\left(x\right).$$

Таким образом, если $F^{-1}\left(y\right)$ может быть явно вычислена, а $u_1,u_2,\ldots u_n$ последовательность равномерных случайных чисел, то последовательность чисел $x_1,x_2,\ldots x_n$ с $x_k=F^{-1}\left(u_k\right),\ k=\overline{1,n}$ будет последовательностью значений случайной величины X, имеющей функцию распределения $F\left(x\right)$.

Пример: Моделирование случайной величины с экспоненциальным законом распределения.

Пусть случайная величина X имеет плотность вероятностей $f(x)=\lambda e^{-\lambda x},\quad x\ge 0,\lambda>0$. Функция распределения этой случайной величины $F(x)=1-e^{-\lambda x},\quad x\ge 0,\lambda>0$ является строго возрастающей и потому существует обратная к ней: $x=F^{-1}(y)=-\dfrac{1}{\lambda}\ln\left(1-y\right),\quad y\in\left(0,1\right)$. Для получения значений случайной величины X следует положить $x_k=-\dfrac{1}{\lambda}\ln\left(1-u_k\right),\quad k=1,2,\ldots$. Поскольку случайные величины U и $1-U$ одинаково распределены, то можно также положить

$$x_k = -\frac{1}{\lambda}\ln(u_k), \quad k = 1,2,\ldots \; .$$

Замечание: Если смоделирована случайная величина X с функцией распределения $F(x)$ и $\{x_k\}$ – последовательность ее значений, то случайные числа $\{a + bx_k\}$, $b > 0$ будут являться значениями случайной величины с функцией распределения $F\left(\frac{x-a}{b}\right)$. В частности, линейное преобразование $Y = a + (b-a)U$, где U - равномерно распределенная на отрезке $[0,1]$ случайная величина, позволяет получать случайные числа, равномерно распределенные на отрезке $[a,b]$.

Следует отметить, что стандартный метод моделирования применим на практике, если только функции $F(x)$ и $F^{-1}(y)$ выражаются через элементарные. Для некоторых распределений это не так, но вместе с тем, как показывают примеры ниже, эффективные методы моделирования существуют и в этих случаях.

Моделирование случайной величины, имеющей однопараметрическое гамма-распределение

Плотность вероятностей однопараметрического гамма-распределения с параметром a имеет вид

$$f_a(x) = \frac{x^{a-1} e^{-x}}{\Gamma(a)},$$

где $x > 0$, а $\Gamma(a) = \int_0^{\infty} x^{a-1} e^{-x}\, dx$ - гамма функция.

Известно, что сумма двух независимых случайных величин X_α и X_β, имеющих однопараметрическое гамма-распределение с параметрами α и β, снова имеет гамма-распределение с параметром $\alpha + \beta$. Из вида плотности $f_1(x)$ следует, что X_1 есть экспоненциально распределенная случайная величина, моделирование которой осуществляется в соответствии с рассмотренным выше примером. Поэтому моделирование случайной величины X_n, имеющей гамма-распределение с натуральным параметром n можно производить с использованием формулы:

$$x_n = x_1^{(1)} + \ldots x_1^{(n)} = -\sum_{k=1}^{n} \ln(u_k) = -\ln\left(\prod_{k=1}^{n} u_k\right), k = 1,2,\ldots$$

где $x_1^{(k)}$ - независимые экспоненциально распределенные случайные числа.

Моделирование случайной величины, имеющей бета-распределение

Для моделирования случайной величины, имеющей бета-распределение на интервале (0,1) с плотностью вероятностей

$$f_{p,q}(x) = \frac{x^{p-1}(1-x)^{q-1}}{B(p,q)}, 0 < x < 1,$$

где параметры $p > 0,\ q > 0$, а $B(p,q) = \Gamma(p)\Gamma(q)/\Gamma(p+q)$ – бета функция, может быть использован метод Йонка:

1. моделируются два независимых значения u_1 и u_2 случайной величины U, равномерно распределенной на отрезке $[0,1]$;
2. если $u_1^{1/p} + u_2^{1/p} < 1$, то $x_{p,q} = u_1^{1/p}\left(u_1^{1/p} + u_2^{1/p}\right)$, в противном случае возвращаемся к пункту 1.

Моделирование нормально распределенной случайной величины

Большое число статистических задач связано с анализом последовательностей значений нормально распределенных случайных величин, называемых нормальными случайными числами. Последовательность нормальных случайных чисел не может быть получена стандартным методом моделирования, поскольку функция распределения нормального закона распределения (функция Лапласа) не выражается через элементарные функции. Наиболее распространенный способ получения нормальных случайных чисел из последовательности равномерных основан на использовании центральной предельной теоремы теории вероятностей.

Рассмотрим сумму $S_N = \sum_{i=1}^{N} U_i$, где U_i - независимые равномерно распределенные на отрезке $[0,1]$ случайные величины, имеющие математические ожидания $\mathrm{M}U_i = \frac{1}{2}$ и дисперсии $\mathrm{D}U_i = \frac{1}{12}$, $i = 1,\ldots,N$.

В соответствии с центральной предельной теоремой случайная величина

$$X_N = \frac{S_N - \mathrm{M}S_N}{\sqrt{\mathrm{D}S_N}} = \frac{S_N - 0.5N}{\sqrt{N}}\sqrt{12}$$

имеет приближенно нормальное распределение с параметрами $(0,1)$ (приближение тем лучше, чем больше N). На практике удовлетворительное приближение к нормальному распределению $N(0,1)$ получается уже при $N = 12$ и это значение параметра N обычно используют для конкретных вычислений. Таким образом, из последовательности равномерных случайных чисел $u_1, u_2, \ldots$ можно получить последовательность нормальных случайных чисел $x_1, x_2, \ldots$ с параметрами $(0,1)$ по формуле:

$$x_i = \sum_{k=1}^{12} u_{ik} - 6, \quad i = 1,2,\ldots$$

Если есть необходимость в получении нормальных случайных чисел $y_1, y_2, \ldots$ с параметрами $\left(a, \sigma^2\right)$, то следует положить $y_i = \sigma x_i + a$.

Приведенный алгоритм получения нормальных случайных чисел прост в реализации, однако он приводит лишь к приближенно нормальным случайным числам. Кроме того, он имеет существенный недостаток в смысле быстродействия процесса моделирования, так как для получения одного нормального случайного числа используется $N = 12$ равномерных случайных чисел.

Другой метод получения нормальных случайных чисел из равномерных, часто используемый на практике, состоит в построении соответствующего явно вычисляемого функционального преобразования. Пример такого преобразования дает следующее утверждение.

Утверждение (см. [5, раздел 2.12]). Пусть случайные величины U_1 и U_2 являются независимыми и распределенными равномерно на отрезке $[0,1]$. Тогда случайные величины

$$X_1 = \sqrt{-2\ln(U_2)}\cos(2\pi U_1),\quad X_2 = \sqrt{-2\ln(U_2)}\sin(2\pi U_1)$$

являются независимыми и имеют стандартное нормальное распределение $N(0,1)$.

Доказательство этого утверждения представляет собой простое упражнение в вычислении плотности вероятностей при взаимнооднозначном преобразовании случайных векторов. Якобиан указанного преобразования равен

$$J(u_1, u_2) = \begin{vmatrix} \dfrac{\partial x_1}{\partial u_1} & \dfrac{\partial x_2}{\partial u_1} \\ \dfrac{\partial x_1}{\partial u_2} & \dfrac{\partial x_2}{\partial u_2} \end{vmatrix} = \frac{2\pi}{u_2} = 2\pi\exp\left(\frac{x_1^2 + x_2^2}{2}\right).$$

Поэтому совместная плотность вероятностей величин X_1 и X_2 равна

$$f_{X_1X_2}(x_1, x_2) = \frac{1}{2\pi}\exp\left(-\frac{x_1^2 + x_2^2}{2}\right).$$

Таким образом, каждая пара $(u_{2k-1}, u_{2k}),\ k = 1,2,\ldots$ равномерных случайных чисел с помощью указанного в утверждении преобразования порождает пару нормальных $N(0,1)$ случайных чисел (x_{2k-1}, x_{2k}).

Приведенным алгоритмом предпочтительно пользоваться, когда необходимо получить достаточно много нормальных чисел, так как он требует существенно меньше равномерных случайных чисел. Следует отметить также, что нормальные случайные числа, получаемые с его помощью, являются более точными, нежели полученные с помощью центральной предельной теоремы.

1.5.2 Моделирование нормально распределенного случайного вектора

Закон распределения нормального (гауссовского) случайного вектора $\mathbf{X} = (X_1, \ldots, X_n)$, как известно, вполне определяется его вектором математических ожиданий $\mathbf{a} = (a_1, \ldots, a_n)$ и корреляционной матрицей $\mathbf{R} = \left\| \mathrm{R}_{ij} \right\|_{i,j=1}^{n}$, где $\mathrm{R}_{ij} = \mathrm{M}(X_i - a_i)(X_j - a_j)$. В предположении, что распределение вектора $\mathbf{X}$ является невырожденным ($\det \mathbf{R} > 0$) плотность вероятностей многомерного нормального распределения имеет вид:

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det \mathbf{R}}} \exp\left\{-\frac{1}{2}(\mathbf{R}^{-1}(\mathbf{x}-\mathbf{a}),(\mathbf{x}-\mathbf{a}))\right\},$$

где $\mathbf{x} = (x_1,...,x_n) \in R^n$, $\mathbf{R}^{-1}$ - матрица, обратная к $\mathbf{R}$, а символ $(\cdot,\cdot)$ обозначает скалярное произведение в евклидовом пространстве R^n: $(\mathbf{x},\mathbf{y}) = \sum\limits_{i=1}^{n} x_i y_i$. для любых $\mathbf{x},\mathbf{y} \in R^n$.

Известно, что вектор $\mathbf{X}$ с многомерным нормальным распределением можно получить специальным линейным преобразованием вектора $\mathbf{Y} = (Y_1,...,Y_n)$ с независимыми, одинаково распределенными по закону $N(0,1)$ координатами. Обычно предполагают, что матрица $\mathbf{B}$ преобразования

$$\mathbf{X} = \mathbf{BY} + \mathbf{a}$$

является треугольной, т.е. k-я строка матрицы $\mathbf{B}$ имеет вид:

$$(b_{k1},...,b_{kk},0,...,0), \qquad k = 1,...,n.$$

Коэффициенты b_{ij} при этом легко определяются рекуррентной процедурой. Поскольку $X_1 = b_{11}Y_1 + a_1$, то $b_{11} = \sqrt{\mathrm{R}_{11}} = \sqrt{\mathrm{D}X_1}$. Далее имеем:

$$X_2 = b_{21}Y_1 + b_{22}Y_2 + a_2,$$

$$\mathrm{M}[b_{11}Y_1(b_{21}Y_1 + b_{22}Y_2)] = \mathrm{R}_{12},$$

$$\mathrm{M}(b_{21}Y_1 + b_{22}Y_2)^2 = \mathrm{R}_{22}.$$

Следовательно,

$$b_{21} = \frac{\mathrm{R}_{12}}{b_{11}} = \frac{\mathrm{R}_{12}}{\sqrt{\mathrm{R}_{11}}}, \qquad b_{22} = \sqrt{\mathrm{R}_{22} - \frac{\mathrm{R}_{21}^2}{\mathrm{R}_{11}}}.$$

Общая рекуррентная формула выглядит следующим образом:

$$b_{ij} = \frac{\mathrm{R}_{ij} - \sum\limits_{k=1}^{j-1} b_{ik}b_{jk}}{\sqrt{\mathrm{R}_{jj} - \sum\limits_{k=1}^{j-1} b_{jk}^2}}, \qquad 1 \le j \le i \le n,$$

где полагается, что $\sum\limits_{k=1}^{0} b_{ik}b_{jk} = 0$.

Таким образом, моделирование произвольного невырожденного нормального случайного вектора $\mathbf{X}$ сводится к моделированию n независимых случайных величин $Y_1,...,Y_n$, каждая из которых имеет стандартный нормальный закон распределения $N(0,1)$. Алгоритмы моделирования случайных величин Y_k, $k=1,...,n$ приведены в разделе 1.5.1.

Подробнее о методах моделирования случайных величин и векторов см. [2], [5], [7].

2 ПРАКТИЧЕСКИЕ ЗАНЯТИЯ

2.1 Первичная обработка статистических данных

2.1.1 Записать выборку (5, 3, 7, 10, 5, 5, 2, 10, 7, 2, 7, 7, 4, 2, 4) в виде вариационного ряда и статистического ряда, обозначив $(y_1,...,y_r)$ различные среди выборочных значений $(x_1,...,x_n)$.

2.1.2 В эксперименте наблюдалась целочисленная случайная величина X. Полученные выборочные значения оказались равными (3, 0, 4, 3, 6, 0, 3, 1). Записать их в виде статистического ряда, найти соответствующую эмпирическую функцию распределения $F_8^*(x)$ и изобразить ее графически.

2.1.3 Пусть (0,8; 2,9; 4,4; -5,6; 1,1; -3.2) – наблюдавшиеся значения некоторой случайной величины. Построить эмпирическую функцию распределения $F_6^*(x)$ и проверить, что $F_6^*(-5)=1/6$, $F_6^*(0)=1/2$, $F_6^*(4)=5/6$.

2.1.4 Найти и изобразить графически эмпирические функции распределения, соответствующие следующим выборкам, представленным в виде статистических рядов:

а)

y_i	1	4	6
m_i	10	15	25

б)

y_i	2	5	7	8
m_i	1	3	2	4

2.1.5 Построить эмпирические функции распределения и вычислить выборочные средние и выборочные дисперсии, соответствующие следующим выборкам, представленным в виде статистических рядов:

а)

y_i	-1	0	1	4
m_i	5	2	9	4

б)

y_i	1	4	5	9	10
m_i	1	3	4	2	5

2.1.6 В результате эксперимента непрерывная случайная величина X приняла следующие значения (округленные до целых): (6, 17, 9, 13, 21, 11, 7, 7, 19, 5, 17, 5, 20, 18, 11, 4, 6, 22, 21, 15, 15, 23, 19, 25, 1) Построить интервальный статистический ряд, взяв 5 интервалов одинаковой длины; построить гистограмму и полигон частот.

2.1.7 Построить гистограмму и полигон частот по следующим статистическим данным, представленным в виде интервального статистического ряда:

Номер интервала k	Границы интервала $[u_{k-1}, u_k)$	Частота интервала ν_k
1	0 – 2	2
2	2 – 4	4
3	4 – 6	8
4	6 – 8	12
5	8 – 10	16
6	10 – 12	10
7	12 - 14	3

Найти и построить эмпирическую функцию распределения $\tilde{F}_n^*(x)$, соответствующую этим сгруппированным данным; вычислить выборочное среднее и выборочную дисперсию.

2.1.8 Проводились опыты с бросанием 12 игральных костей. Наблюдаемую случайную величину X брали равной числу костей, на которых выпало 4, 5 или 6 очков. Пусть m_i - число опытов, в которых наблюдалось значение $X = i \quad (i = 0,1,...,12)$. Ниже приведены результаты $n = 4096$ опытов:

i	0	1	2	3	4	5	6	7	8	9	10	11	12
m_i	0	7	60	198	430	731	918	847	536	257	71	11	0

Взяв в качестве интервалов $J_i = \left[\frac{i-1}{2}, \frac{i+1}{2}\right)$, $i = 0,1,...,12$, построить гистограмму и полигон частот. Вычислить выборочное среднее и выборочную дисперсию, соответствующие этим данным.

2.1.9 Ниже приведены результаты измерения роста случайно отобранных 100 студентов:

Рост, см	154-158	158-162	162-166	166-170	170-174	174-178	178-182
Число студентов	10	14	26	28	12	8	2

Найти выборочное среднее и выборочную дисперсию роста обследованных студентов, построить гистограмму и полигон частот.

2.1.10 На телефонной станции производились наблюдения за числом неправильных соединений в минуту. Наблюдения в течение часа дали следующие результаты:

(3, 1, 3, 4, 2, 1, 1, 3, 2, 7, 2, 0, 2, 4, 0, 3, 0, 2, 0, 1, 3, 3, 1, 2, 2, 0, 2, 1, 4, 3, 4, 2, 0, 2, 3, 1, 3, 1, 4, 2, 2, 1, 2, 5, 1, 1, 0, 1, 1, 2, 1, 0, 3, 4, 1, 2, 2, 1, 1, 5).

Найти выборочное среднее и выборочную дисперсию числа неправильных соединений в минуту и сравнить эмпирическое распределение с распределением Пуассона.

2.1.11 Измерительным прибором, практически не имеющим систематической погрешности, было сделано пять независимых измерений некоторой величины. Получены следующие результаты:

Номер измерения	1	2	3	4	5
Результат измерения	2781	2836	2807	2763	2858

Найти: а) выборочную дисперсию погрешности измерения, если измеряемая величина точно известна и равна 2800; б) выборочное среднее и выборочную дисперсию, если точное значение измеряемой величины неизвестно.

2.1.12 Доказать, что выборочные начальные и выборочные центральные моменты k-го порядка связаны соотношением:

$$\mu_k^* = \sum_{j=1}^{k} (-1)^{k-j} C_k^j \alpha_j^* (\overline{x})^{k-j}.$$

2.2 Точечные оценки неизвестных параметров

2.2.1 Показать, что выборочное среднее $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ является несмещенной и состоятельной оценкой математического ожидания наблюдаемой случайной величины X.

2.2.2 Показать, что выборочная дисперсия $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ не является несмещенной оценкой дисперсии наблюдаемой случайной величины X. Определить смещение оценки S^2. Является ли она асимптотически несмещенной? Показать, что несмещенной оценкой дисперсии является величина $S'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{n}{n-1} S^2$, называемая ***исправленной выборочной дисперсией.***

2.2.3 Показать, что справедлива формула:

$$\mathrm{D}S^2 = \frac{(n-1)^2}{n^3}\left(\mu_4 - \frac{n-3}{n-1}\mu_2^2\right),$$

где $\mu_4 = \mathrm{M}(X - \mathrm{M}X)^4$ – четвертый центральный момент случайной величины X. Являются ли выборочная дисперсия S^2 и исправленная выборочная дисперсия S'^2 состоятельными оценками наблюдаемой случайной величины X?

2.2.4 Показать, что эмпирическая функция распределения $F_n^*(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i < x)$ является несмещенной и состоятельной оценкой теоретической функции распределения $F(x)$.

2.2.5 Вычислить информацию Фишера $I(\theta)$ о параметре θ, содержащуюся в одном наблюдении над случайной величиной X, имеющей:

а) нормальное распределение $N(\theta, \sigma^2)$;

б) нормальное распределение $N(a, \theta^2)$;

в) нормальное распределение $N(\theta_1, \theta_2^2)$;

г) гамма-распределение $\Gamma(\theta, \lambda)$;

д) распределение Коши $K(\theta)$;

е) биномиальное распределение $Bi(n, \theta)$;

ж) распределение Пуассона $\Pi(\theta)$.

2.2.6 Доказать, что выборочное среднее $\overline{X}$ является эффективной оценкой математического ожидания наблюдаемой случайной величины X, имеющей распределение: а) $N(\theta, \sigma^2)$; б) $N(\theta_1, \theta_2^2)$.

2.2.7 Доказать, что статистика $\theta^* = \frac{1}{n}\sum_{i=1}^{n}(X_i - a)^2$ является несмещенной и эффективной оценкой дисперсии наблюдаемой случайной величины X, имеющей нормальное распределение $N(a, \theta^2)$.

2.2.8 Показать, что исправленная выборочная дисперсия S'^2 является асимптотически эффективной оценкой дисперсии наблюдаемой случайной величины X, имеющей нормальное распределение $N(\theta_1, \theta_2^2)$.

2.2.9 Пусть наблюдаемая случайная величина X имеет распределение Коши $K(\theta)$. Показать, что выборочное среднее $\overline{X}$ не является состоятельной оценкой параметра θ.

2.2.10 Показать, что в случае логистического распределения, плотность вероятностей которого при $-\infty < x < \infty$ имеет вид:

$$f(x;\theta) = e^{-x+\theta}(1+e^{-x+\theta})^{-2},$$

справедливы утверждения:

а) выборочное среднее $\overline{X}$ – несмещенная оценка θ и $\mathrm{D}\overline{X} = \dfrac{\pi^2}{3n}$;

б) информация Фишера $I(\theta) = 1/3$ и поэтому $\overline{X}$ не является эффективной оценкой θ.

2.2.11 Пусть по выборке $(x_1, ..., x_n)$ из генеральной совокупности, имеющей биномиальное распределение $Bi(1,\theta)$, требуется оценить функцию $\tau(\theta) = 1/\theta$. Показать, что в данном случае несмещенных оценок не существует.

2.2.12 Пусть по одному наблюдению над случайной величиной X, имеющей отрицательное биномиальное распределение $\overline{Bi(1,\theta)}$, требуется оценить параметр θ. Найти оценку, удовлетворяющую условию несмещенности, и показать, что она практически бесполезна.

2.2.13 Пусть производится одно наблюдение над случайной величиной X, имеющей распределение Пуассона $\Pi(\theta)$ с неизвестным параметром θ:

$$\mathrm{P}(X = x) = (\theta^x / x!)e^{-\theta}, \ \ x = 0,1,2,\dots.$$

Показать, что в этом случае не существует несмещенных оценок параметрической функции $\tau(\theta) = 1/\theta$.

2.2.14 Случайная величина X имеет распределение Пуассона $\Pi(\theta)$ с неизвестным параметром θ. По выборке $(x_1, ..., x_n)$ найти точечные оценки параметра θ по методу моментов, используя первый и второй моменты. Исследовать полученные оценки на несмещенность.

2.2.15 Случайная величина X (число нестандартных изделий в партии изделий) имеет распределение Пуассона $\Pi(\theta)$ с неизвестным параметром θ. Найти методом моментов оценку параметра θ, если обследование $n = 200$ партий на наличие нестандартных изделий дало следующие результаты:

x_i	0	1	2	3	4
m_i	132	43	20	3	2

где x_i - число нестандартных изделий в одной партии; m_i - число партий, содержащих x_i нестандартных изделий.

2.2.16 По выборке $(x_1,...,x_n)$ из генеральной совокупности, имеющей биномиальное распределение $Bi(m,\theta)$, т.е. $\mathrm{P}(X = x) = C_m^x \theta^x (1-\theta)^{m-x}$, $x = 0,1,2,...,m$, найти оценку неизвестной вероятности успеха θ по методу моментов, используя первый момент.

2.2.17 Случайная величина X (число появлений события A в m независимых испытаниях) подчинена биномиальному закону распределения $Bi(m,\theta)$ с неизвестным параметром θ. Ниже приведены результаты числа появлений события в 10 опытах по 5 испытаний в каждом:

x_i	0	1	2	3	4
m_i	5	2	1	1	1

где x_i - число появлений события A в одном опыте; m_i - количество опытов, в которых наблюдалось x_i появлений события A. Найти методом моментов оценку параметра θ по этим статистическим данным.

2.2.18 Случайная величина X имеет геометрическое распределение $G(\theta)$ с неизвестным параметром θ, т.е. $\mathrm{P}(X = x) = \theta(1-\theta)^{x-1}$, $x = 1,2,...$, где x – число испытаний, произведенных до первого появления события; θ - вероятность появления события в одном испытании. По выборке $(x_1,...,x_n)$ найти методом моментов оценку параметра θ.

2.2.19 Найти методом моментов оценку параметра θ геометрического распределения $G(\theta)$, если в четырех опытах событие появилось соответственно после двух, четырех, шести и восьми испытаний.

2.2.20 По выборке $(x_1,...,x_n)$ найти методом моментов оценку неизвестного параметра θ показательного распределения $E(\theta)$, плотность вероятностей которого $f(x;\theta)=\theta e^{-\theta x}$, $x\ge 0$. Исследовать полученную оценку на несмещенность.

2.2.21 Найти методом моментов по выборке $(x_1,...,x_n)$ точечные оценки неизвестных параметров α и β гамма-распределения $\Gamma(\alpha,\beta)$, плотность вероятностей которого

$$f(x;\alpha,\beta)=\frac{1}{\Gamma(\alpha+1)\beta^{\alpha+1}}x^{\alpha}e^{-x/\beta},\ (\alpha>-1,\ \beta>0,\ x\ge 0).$$

2.2.22 Случайная величина X имеет равномерное распределение на отрезке $[\theta_1,\theta_2]$ с неизвестными границами. По выборке $(x_1,...,x_n)$ найти методом моментов оценки параметров θ_1 и θ_2.

2.2.23 По выборке $(x_1,...,x_n)$ из генеральной совокупности, имеющей биномиальное распределение $Bi(m,\theta)$, найти оценку максимального правдоподобия $\hat{\theta}$ неизвестной вероятности успеха θ. Сравнить ее с оценкой, полученной по методу моментов (см. задачу 2.2.16).

2.2.24 Случайная величина X имеет распределение Пуассона $\Pi(\theta)$ с неизвестным параметром θ. По выборке $(x_1,...,x_n)$ найти оценку максимального правдоподобия $\hat{\theta}$ параметра θ и исследовать ее на несмещенность.

2.2.25 По выборке $(x_1,...,x_n)$ найти оценку максимального правдоподобия $\hat{\theta}$ неизвестного параметра θ показательного распределения $E(\theta)$. Сравнить ее с оценкой, полученной по методу моментов (см. задачу 2.2.20).

2.2.26 Случайная величина X (время безотказной работы элемента) имеет показательный закон распределения $E(\theta)$ с неизвестным параметром θ. В результате проверки 1000 элементов были получены следующие значения среднего времени их работы:

x_i	5	15	25	35	45	55	65
m_i	365	245	150	100	70	45	25

Здесь x_i – среднее время безотказной работы одного элемента, часов; m_i - количество элементов, проработавших в среднем x_i часов. Найти на основании этих данных оценку параметра θ по методу максимального правдоподобия.

2.2.27 По выборке $(x_1,...,x_n)$ из генеральной совокупности, имеющей гамма-распределение $\Gamma(1,\theta)$ с плотностью вероятностей $f(x;\theta)=(x/\theta^2)e^{-x/\theta}$, $x \ge 0$, найти оценку максимального правдоподобия $\hat{\theta}$ неизвестного параметра θ. Сравнить ее с оценкой, полученной по методу моментов (см. задачу 2.2.21).

2.2.28 Устройство состоит из элементов, время безотказной работы которых подчинено гамма-распределению $\Gamma(1,\beta)$. Испытания пяти элементов дали следующие результаты (время работы элемента в часах до отказа): (50, 60, 100, 200, 250). Найти по этим выборочным значениям оценку максимального правдоподобия неизвестного параметра β.

2.2.29 Случайная величина X, характеризующая срок службы элементов электронной аппаратуры, имеет плотность вероятностей $f(x;\theta)=(2x/\theta)e^{-x^2/\theta}$, $x \ge 0$ (закон распределения Релея). По выборке $(x_1,...,x_n)$ построить оценку максимального правдоподобия $\hat{\theta}$ неизвестного параметра θ. Исследовать ее на несмещенность и сравнить с оценкой, которую дает метод моментов.

2.2.30 По выборке $(x_1,...,x_n)$ найти оценки максимального правдоподобия неизвестных параметров распределений: а) $N(\theta,\sigma^2)$; б) $N(a,\theta^2)$; в) $N(\theta_1,\theta_2^2)$. Проанализировать качество полученных оценок.

2.2.31 Наблюдаемая случайная величина X имеет геометрическое распределение $G(\theta)$ с неизвестным параметром θ. По выборке $(x_1,...,x_n)$ построить оценку максимального правдоподобия неизвестной вероятности θ и сравнить ее с оценкой, полученной по методу моментов (см. задачу 2.2.18).

2.2.32 Случайная величина X имеет равномерное распределение на отрезке $[0,\theta]$. По выборке $(x_1,...,x_n)$ построить оценку максимального правдоподобия $\hat{\theta}$ неизвестного параметра θ.

2.2.33 Случайная величина X имеет равномерное распределение на отрезке единичной длины $[\theta,\theta+1]$. По выборке $(x_1,...,x_n)$ построить оценку максимального правдоподобия $\hat{\theta}$ неизвестного параметра θ.

2.2.34 Случайная величина X имеет равномерное распределение на отрезке $[\theta_1, \theta_2]$. По выборке $(x_1, \ldots, x_n)$ построить оценки максимального правдоподобия неизвестных концов отрезка θ_1 и θ_2.

2.3 Интервальные оценки неизвестных параметров

2.3.1 Найти доверительный интервал для математического ожидания нормально распределенной случайной величины X, если известны дисперсия $\mathrm{D}X = \sigma^2$, выборочное среднее $\overline{x}$, объем выборки n и доверительная вероятность γ:

а) $\sigma^2 = 16; \quad \overline{x} = 10{,}2; \quad n = 16; \quad \gamma = 0{,}95;$

б) $\sigma^2 = 25; \quad \overline{x} = 16{,}8; \quad n = 25; \quad \gamma = 0{,}98;$

в) $\sigma^2 = 4; \quad \overline{x} = 7{,}0; \quad n = 100; \quad \gamma = 0{,}99.$

2.3.2 Одним и тем же прибором со средним квадратическим отклонением случайных ошибок измерений $\sigma = 40$ м произведено пять равноточных измерений высоты полета аппарата. Найти доверительный интервал для истинной высоты h полета с надежностью $\gamma = 0{,}999$, зная среднее арифметическое результатов измерений $\overline{x} = 20000$ м, и предполагая, что результаты измерений распределены нормально.

2.3.3 Найти минимальный объем выборки, при котором с надежностью 0,975 точность оценки математического ожидания a случайной величины X по выборочному среднему равна 0,3, если известна дисперсия $\mathrm{D}X = 1{,}44$.

2.3.4 Построить доверительный интервал для математического ожидания нормально распределенной случайной величины X, если известны выборочное среднее $\overline{x}$, выборочная дисперсия s^2, объем выборки n и доверительная вероятность γ:

а) $\overline{x} = 3{,}2; \quad s^2 = 6{,}25; \quad n = 16; \quad \gamma = 0{,}95;$

б) $\overline{x} = 1{,}1; \quad s^2 = 1{,}69; \quad n = 25; \quad \gamma = 0{,}98;$

в) $\overline{x} = 2{,}0; \quad s^2 = 4{,}0; \quad n = 100; \quad \gamma = 0{,}99.$

2.3.5 По данным девяти независимых равноточных измерений некоторой физической величины найдены среднее арифметическое результатов измерений $\overline{x} = 30{,}1$ и выборочное среднее квадратическое отклонение $s = 6$. Оценить истинное значение измеряемой величины с помощью доверительного интервала с надежностью $\gamma = 0{,}999$. Предполагается, что результаты измерений распределены нормально.

2.3.6 Построить доверительный интервал для дисперсии нормально распределенной случайной величины X, если известны выборочная дисперсия s^2, объем выборки n и доверительная вероятность γ:

а) $s^2 = 1{,}21;\quad n = 10;\quad \gamma = 0{,}95;$ б) $s^2 = 10{,}0;\quad n = 30;\quad \gamma = 0{,}98;$

в) $s^2 = 4{,}5;\quad n = 50;\quad \gamma = 0{,}99.$

2.3.7 Зная выборочную дисперсию $s^2 = 2$ и объем выборки $n = 100$ из нормальной генеральной совокупности, построить точный и приближенный доверительные интервалы для неизвестной дисперсии σ^2, отвечающие доверительной вероятности $\gamma = 0{,}95$. Сравнить полученные результаты.

2.3.8 Построить доверительный интервал для математического ожидания случайной величины X, имеющей биномиальное распределение $Bi(1,\theta)$, если выборочное среднее $\bar{x} = 5{,}0$, а объем выборки $n = 100$. Доверительную вероятность принять равной $\gamma = 0{,}95$.

2.3.9 Зная объем выборки $n = 100$, выборочное среднее $\bar{x} = 1{,}0$, выборочную дисперсию $s^2 = 9$ и выборочный центральный момент $\mu_4^* = 90$, найти доверительные интервалы для математического ожидания и дисперсии случайной величины X, имеющей показательное распределение с плотностью вероятностей $f(x;\theta) = \theta e^{-\theta x}$, $x \geq 0$. Доверительную вероятность принять равной $\gamma = 0{,}975$.

2.3.10 Показать, что в случае экспоненциального распределения с плотностью вероятностей $f(x;\theta) = e^{-(x-\theta)}$, $x \geq 0$, γ-доверительный интервал для θ имеет вид: $\left(X_{(1)} + \dfrac{\ln(1-\gamma)}{n}, X_{(1)} \right)$, где $X_{(1)} = \min\limits_{1 \leq i \leq n} X_i$.

2.3.11 Пусть n, $\bar{X}$ и S^2 - соответственно объем, выборочное среднее и дисперсия выборки из генеральной совокупности, имеющей нормальное распределение с неизвестными параметрами. Показать, что с вероятностью γ результат следующего, $(n+1)$-го, наблюдения находится в интервале:

$$(\bar{X} \pm t_{\gamma;n-1} \cdot S\sqrt{(n+1)(n-1)}).$$

2.3.12 В результате пяти независимых взвешиваний одного и того же тела получены следующие результаты (в граммах): (4,12; 3,92; 4,55; 4,04; 4,35). Считая погрешности измерений нормальными $N(0,\sigma^2)$ случайными величинами, указать доверительные границы для результатов предстоящего шестого взвешивания (доверительную вероятность принять равной $\gamma = 0{,}95$).

2.4 Проверка статистических гипотез

2.4.1 В таблице даны результаты $n = 200$ наблюдений над случайной величиной X:

x_i	0,3	0,5	0,7	0,9	1,1	1,3	1,5	1,7	1,9	2,1	2,3
m_i	6	9	26	25	30	26	21	24	20	8	5

Проверить, используя критерий χ^2 Пирсона, согласуются ли эти данные с гипотезой о нормальном $N(1,4)$ распределении генеральной совокупности. Уровень значимости принять равным $\alpha = 0{,}05$.

2.4.2 Используя критерий χ^2 при уровне значимости $\alpha = 0{,}02$ проверить, согласуется ли следующая выборка, представленная в виде статистического ряда, с гипотезой о нормальном распределении генеральной совокупности.

x_i	5	7	9	11	13	15	17	19	21
m_i	15	26	25	30	26	21	24	20	13

2.4.3 Используя критерий χ^2, проверить при уровне значимости $\alpha = 0{,}05$, согласуются ли с гипотезой о нормальном распределении генеральной совокупности следующие данные, представленные в виде интервального статистического ряда:

Номер интервала k	Границы интервала $[u_{k-1}, u_k)$	Частота интервала ν_k
1	3 – 8	6
2	8 – 13	88
3	13 – 18	15
4	18 – 23	40
5	23 – 28	16
6	28 – 33	8
7	33 – 38	7

2.4.4 В итоге регистрации времени прихода 800 посетителей выставки (в качестве начала отсчета времени принят момент открытия работы выставки) получены данные представленные в следующей таблице:

$[u_{k-1}, u_k)$	ν_k	$[u_{k-1}, u_k)$	ν_k
0 – 1	259	4 – 5	70
1 – 2	167	5 – 6	47
2 – 3	109	6 – 7	40
3 – 4	74	7 – 8	34

где $[u_{k-1}, u_k)$ - интервалы времени; ν_k - частоты интервалов, т.е. количество посетителей, пришедших в течение соответствующего интервала. Требуется при уровне значимости $\alpha = 0,01$ проверить гипотезу о том, что время прихода посетителей выставки распределено по показательному закону с параметром $\lambda = 2$.

2.4.5 Испытания 200 элементов на продолжительность работы дали результаты, приведенные в следующей таблице:

$[u_{k-1}, u_k)$	ν_k	$[u_{k-1}, u_k)$	ν_k
0 – 5	133	15 – 20	4
5 – 10	45	20 – 25	2
10 – 15	15	25 – 30	1

где $[u_{k-1}, u_k)$ - интервалы времени в часах; ν_k - частоты интервалов, т.е. количество элементов, проработавших время в пределах соответствующего интервала. Требуется, используя критерий χ^2 при уровне значимости $\alpha = 0,05$ проверить гипотезу о том, что время работы элементов распределено по показательному закону с плотностью вероятностей $f(x) = \lambda e^{-\lambda x}$, $x \geq 0$.

2.4.6 В опытах наблюдалась неотрицательная непрерывная случайная величина X. Ее значения (упорядоченные по величине и округленные до двух знаков) для $n = 50$ опытов оказались равными: (0,01; 0,01; 0,04; 0,17; 0,18; 0,22; 0,22; 0,25; 0,25; 0,29; 0,42; 0,46; 0,47; 0,56; 0,59; 0,67; 0,68; 0,70; 0,72; 0,76; 0,78; 0,83; 0,85; 0,87; 0,93; 1,00; 1,01; 1,01; 1,02; 1,03; 1,05; 1,32; 1,34; 1,37; 1,47; 1,50; 1,52; 1,54; 1,59; 1,71; 1,90; 2,10; 2,35; 2,46; 2,46; 2,50; 3,73; 4,07; 5, 55; 6,03).

Проверить на основании этих статистических данных гипотезу $H_0 : F_X(x) = 1 - e^{-x}$, $x \geq 0$, применяя критерий χ^2 и используя группировку с четырьмя равновероятными интервалами. Уровень значимости принять равным $\alpha = 0,1$.

2.4.7 Наблюдались показания 500 наугад выбранных часов, выставленных в витринах часовщиков. Пусть i - номер промежутка от i-го часа до $(i+1)$-го, $i = 0,1,...,11$, а ν_i - число часов, показания которых принадлежали i-му промежутку. Результаты наблюдений оказались следующими:

i	0	1	2	3	4	5	6	7	8	9	10	11
ν_i	41	34	54	39	49	45	41	33	37	41	47	39

Согласуются ли эти данные с гипотезой о том, что показания часов равномерно распределены на интервале $(0,12)$ при уровне значимости $\alpha = 0,05$?

2.4.8 В течение 10 часов регистрировали прибытие автомашин к бензоколонке и получили данные представленные в таблице:

$[u_{k-1}, u_k)$	ν_k	$[u_{k-1}, u_k)$	ν_k
8 – 9	12	13 – 14	6
9 – 10	40	14 – 15	11
10 – 11	22	15 – 16	33
11 – 12	16	16 – 17	18
12 – 13	28	17 – 18	14

Требуется при уровне значимости $\alpha = 0,01$ проверить гипотезу о том, что время прибытия автомашин распределено равномерно.

2.4.9 Четыре монеты подбрасывались одновременно 100 раз, причем каждый раз отмечалось число выпавших «гербов». Ниже приведены частоты m_i случаев, когда число выпавших «гербов» было равно x_i:

x_i	0	1	2	3	4
m_i	8	20	42	22	8

Пользуясь критерием χ^2, проверить согласие этих данных с гипотезой о биномиальном распределении числа выпавших «гербов», если вероятность выпадения «герба» при бросании каждой из монет равна 0,5. Уровень значимости принять равным $\alpha = 0,05$.

2.4.10 Из 2020 семей, имеющих двух детей, в 527 семьях по два мальчика и в 476 – по две девочки (в остальных 1017 семьях дети разного пола). Можно ли с уровнем значимости $\alpha = 0,05$ считать, что количество мальчиков в семье с двумя детьми – биномиальная случайная величина?

2.4.11 Выборка представляет собой целочисленный случайный вектор (47, 46, 49, 53, 50). Можно ли с уровнем значимости $\alpha = 0{,}1$ считать распределение наблюдавшейся случайной величины пуассоновским?

2.4.12 При $n = 4040$ бросаниях монеты было получено $m_1 = 2048$ выпадений «герба» и $m_2 = n - m_1 = 1992$ выпадений «решки». Проверить, используя критерий χ^2, совместимы ли эти данные с гипотезой о том, что монета была симметричной.

2.4.13 При $n = 4000$ независимых испытаниях события A_1, A_2 и A_3, образующие полную группу событий, осуществились 1905, 1015 и 1080 раз соответственно. Проверить, согласуются ли эти данные на уровне значимости $\alpha = 0{,}05$ с гипотезой $H_0 : p_1 = 1/2;\ p_2 = p_3 = 1/4$, где $p_i = \mathrm{P}(A_i)$, $i = 1,2,3$.

2.4.14 Таблица «случайных чисел» содержит реализации 10 000 независимых и одинаково распределенных случайных величин, принимающих значения 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. Корректно ли предположение о равновероятности этих значений, если в упомянутой таблице числа, не превосходящие 4, встречаются 4806 раз? При каком уровне значимости гипотеза равновероятности отвергается?

3 ИНДИВИДУАЛЬНОЕ ЗАДАНИЕ «МОДЕЛИРОВАНИЕ И СТАТИСТИЧЕСКИЙ АНАЛИЗ СЛУЧАЙНЫХ ДАННЫХ»

3.1 Содержание задания

1. Смоделировать случайную величину X, имеющую нормальный закон распределения с параметрами (a_X, σ_X^2). На основе выборки объема n исследовать статистические характеристики случайной величины X, решив следующие задачи.
   1.1. Построить гистограмму распределения и изобразить ее графически одновременно с теоретической плотностью вероятностей.
   1.2. Вычислить выборочное среднее и выборочную дисперсию.
   1.3. Найти оценки математического ожидания и дисперсии методом максимального правдоподобия. Указать несмещенную оценку дисперсии.
   1.4. Построить доверительные интервалы для математического ожидания и дисперсии, соответствующие доверительной вероятности γ.
   1.5. Проверить гипотезу о нормальном распределении случайной величины X, используя критерий χ^2 Пирсона при уровне значимости α.
2. Смоделировать случайную величину Y, имеющую заданный непрерывный закон распределения (отличный от нормального) с заданными параметрами. На основе выборки объема n исследовать статистические характеристики случайной величины Y, решив следующие задачи.
   2.1. Построить гистограмму распределения и изобразить ее графически одновременно с теоретической плотностью вероятностей.
   2.2. Определить точечные оценки математического ожидания и дисперсии.
   2.3. При заданном виде распределения построить оценки входящих в него неизвестных параметров методом моментов.
   2.4. Построить доверительные интервалы для математического ожидания и дисперсии, соответствующие доверительной вероятности γ.
   2.5. Проверить гипотезу о виде распределении случайной величины Y, используя критерий χ^2 Пирсона при уровне значимости α.
3. Смоделировать случайный вектор (X, Y), имеющий двумерный нормальный закон распределения с параметрами $(a_X, a_Y, \sigma_X^2, \sigma_Y^2, r_{XY})$. На

основе выборки объема n исследовать статистические характеристики случайного вектора (X,Y), решив следующие задачи.

3.1. Найти точечные оценки параметров, входящих в распределение.

3.2. Проверить гипотезу о независимости случайных величин X и Y при уровне значимости α.

3.3. Найти эмпирические уравнения регрессии Y на X и X на Y и изобразить их графически одновременно с выборочными значениями.

Примечание: В зависимости от действующего учебного плана по изучаемому курсу на основе данного задания может быть сформировано либо индивидуальное задание для расчетно-графической работы (по усмотрению преподавателя некоторые разделы могут быть исключены), либо индивидуальное задание для курсового проекта.

3.2 Исходные данные к заданию

Каждый студент получает индивидуальный вариант задания, в котором указаны:

для раздела 1: значения объема выборки n, математического ожидания a_X и дисперсии σ_X^2, доверительной вероятности γ и уровня значимости α.

для раздела 2: закон распределения случайной величины (отличный от нормального) и значения его параметров, значения объема выборки n, доверительной вероятности γ и уровня значимости α, задания для аналитических расчетов.

для раздела 3: значение объема выборки n, значения математических ожиданий a_X и a_Y, дисперсий σ_X^2 и σ_Y^2 и коэффициента корреляции r_{XY}, значение уровня значимости α.

Варианты индивидуальных заданий приведены в Приложении А.

3.3 Методические указания по выполнению задания

Задание должно выполняться на компьютере с использованием универсальных математических пакетов MCAD или MATLAB.

1. При моделировании нормальной случайной величины руководствоваться теоретическими сведениями из раздела 1.5.1. Для генерирования равномерно распределенных на отрезке $[0,1]$ случайных чисел использовать стандартные функции (в MCAD – функцию rnd, в MATLAB – функцию rand).

1.1. При выполнении задания руководствоваться теоретическими сведениями из раздела 1.1. Результаты расчетов представить в виде таблицы, содержащей: границы интервалов группировки; частоты и относительные частоты интервалов; суммы частот; значения высот столбцов гистограммы; значения теоретической плотности вероятностей в серединах интервалов.

1.2. При выполнении задания руководствоваться теоретическими сведениями из раздела 1.1. Результаты сравнить с истинными значениями математического ожидания и дисперсии.

1.3. При выполнении задания руководствоваться теоретическими сведениями из раздела 1.2.1. Результаты сравнить с истинными значениями математического ожидания и дисперсии.

1.4. При выполнении задания руководствоваться теоретическими сведениями из раздела 1.2.2. Рассмотреть отдельно случаи известных и неизвестных параметров распределения. Проверить попадание истинных значений математического ожидания и дисперсии в построенные доверительные интервалы.

1.5. При выполнении задания руководствоваться теоретическими сведениями из раздела 1.3.1. Для выполнения требования, состоящего в том, чтобы частоты всех интервалов были не менее 5, произвести, при необходимости, объединение соседних интервалов. Результаты расчетов представить в виде таблицы, содержащей: границы интервалов группировки; частоты и относительные частоты интервалов; суммы частот; теоретические вероятности попадания в интервалы (до и после объединения). Привести значения статистики критерия и порога и сделать вывод о том, принимается выдвинутая гипотеза о виде распределения или нет. Рассмотреть отдельно случаи известных и неизвестных параметров распределения.

2. При моделировании непрерывных случайных величин с законом распределения отличным от нормального руководствоваться теоретическими сведениями из раздела 1.5.1 и из [1, разд. 1, пп. 2.1-2.11, 2.13-2.27]. Привести подробный алгоритм моделирования для своего варианта задания и аналитический расчет теоретических значений математического ожидания и дисперсии.

2.1. При выполнении задания руководствоваться теоретическими сведениями из раздела 1.1. Результаты расчетов представить в виде таблицы, содержащей: границы интервалов группировки; частоты и относительные частоты интервалов; суммы частот; значения высот столбцов гистограммы; значения теоретической плотности вероятностей в серединах интервалов.

2.2. При выполнении задания руководствоваться теоретическими сведениями из раздела 1.1. Результаты сравнить с истинными значениями математического ожидания и дисперсии.

2.3. При выполнении задания руководствоваться теоретическими сведениями из раздела 1.2.1. Результаты сравнить с истинными значениями математического ожидания и дисперсии.

2.4. При выполнении задания руководствоваться теоретическими сведениями из раздела 1.2.2, считая параметры распределения неизвестными. Проверить попадание истинных значений математического ожидания и дисперсии в полученные доверительные интервалы.

2.5. При выполнении задания руководствоваться теоретическими сведениями из раздела 1.3.1, считая параметры распределения неизвестными. Для выполнения требования, состоящего в том, чтобы частоты всех интервалов были не менее 5, произвести, при необходимости, объединение соседних интервалов. Результаты расчетов представить в виде таблицы, содержащей: границы интервалов группировки; частоты и относительные частоты интервалов; суммы частот; теоретические вероятности попадания в интервалы (до и после объединения). Привести значения статистики критерия и порога и сделать вывод о том, принимается выдвинутая гипотеза о виде распределения или нет.

3. При моделировании случайного вектора, имеющего двумерное нормальное распределение руководствоваться разделом 1.5.2 и [1, разд. 2, п. 4].

3.1. При нахождении точечных оценок математических ожиданий и дисперсий координат вектора руководствоваться теоретическими сведениями из раздела 1.1. Оценку коэффициента корреляции находить по формулам раздела 1.4.1. Результаты оценивания сравнить с истинными значениями.

3.2. Проверку гипотезы независимости произвести и с использованием критерия независимости χ^2, и с использованием критерия проверки значимости коэффициента корреляции (см. раздел 1.4.2, [3, разд. 3.5], [4, разд. 4.8.1]).

При использовании критерия независимости χ^2 следует обеспечить выполнение требования, чтобы частоты всех прямоугольников группировки ν_{ij} были не менее 5. Это достигается объединением соседних интервалов группировки выборочных значений случайных величин X и Y соответственно

Результаты расчетов представить в виде двумерной таблицы, содержащей: границы прямоугольников группировки; частоты и относительные частоты интервалов и прямоугольников; суммы частот (до и после объединения интервалов). Привести значения статистики критерия и порога и сделать вывод о том, принимается выдвинутая гипотеза о независимости или нет.

При использовании критерия проверки значимости коэффициента корреляции привести значения статистики критерия и порога и сделать вывод о том, принимается выдвинутая гипотеза о независимости или нет.

3.3. При выполнении задания руководствоваться теоретическими сведениями из раздела 1.4.3. и результатами выполнения п. 3.1. Проанализировать изменение полученных результатов в зависимости от значения коэффициента корреляции.

В заключение сделать выводы об изменении качества всех статистических выводов при увеличении объема выборки n, качества построения доверительных интервалов в зависимости от доверительной вероятности γ, качества проверки гипотез в зависимости от уровня значимости α.

3.4 Требования к оформлению отчета

Отчет об индивидуальном задании должен быть оформлен в виде расчетно-пояснительной записки в соответствии с требованиями оформления учебных текстовых документов, изложенными в СТО СГАУ 02068410-004-2007 «Общие требования к учебным текстовым документам» [6].

Отчет представляется в печатном виде в обложке со стандартным титульным листом, форма которого приведена в Приложении Д.

Расчетно-пояснительная записка включает в себя следующие основные части.

1. ВВЕДЕНИЕ. В нем следует дать общую характеристику роли и значений практических приложений математической статистики.

2. ЗАДАНИЕ И ИСХОДНЫЕ ДАННЫЕ. Следует конкретизировать постановку задач разделов 1 - 3 задания в соответствии с полученными исходными данными.

3. РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ И ИССЛЕДОВАНИЕ СТАТИСТИЧЕСКИХ ХАРАКТЕРИСТИК. Эта основная часть работы содержит последовательное решение задач из разделов 1 - 3 задания. При решении каждой задачи следует приводить описание используемого метода, необходимые формулы и пояснения к ним. Если задача решается двумя методами, следует произвести сравнение этих методов и полученных результатов.

Расчеты и результаты представляются в виде таблиц или графиков функций и помещаются в соответствующих разделах записки.

В Приложениях А, Б, и В к отчету должны быть помещены полные тексты программ, результатов моделирования и статистического анализа для разделов задания 1, 2 и 3 соответственно.

Список использованных источников приводится в конце записки и оформляется в соответствии с требованиями ГОСТ 7.1. В тексте самой записки следует делать ссылки на используемые источники с указанием номера ее в списке (в квадратных скобках).

ЗАКЛЮЧЕНИЕ

Учебное пособие содержит полное методическое обеспечение всех видов учебных занятий по модулю «Математическая статистика» курса «Теория вероятностей и математическая статистика», читавшегося автором в течение ряда лет на факультете информатики СГАУ. В состав учебного пособия входят: краткие теоретические сведения, включающие основы статистического моделирования; методические указания по проведению практических занятий (72 задачи разного уровня сложности по всем разделам математической статистики); 30 вариантов индивидуального задания и методические указания по его выполнению с использованием универсальных пакетов MCAD и MATLAB.

Учебное пособие предназначено для подготовки бакалавров, обучающихся на факультете информатики по направлению 010400.62 «Прикладная математика и информатика», при изучении дисциплины «Теория вероятностей и математическая статистика» в 4 семестре. Пособие также может быть полезным для бакалавров и специалистов других естественнонаучных направлений подготовки для получения практических навыков при статистическом анализе случайных данных.

ЛИТЕРАТУРА

1. **Боровков, Александр Алексеевич.** Математическая статистика. Оценка параметров. Проверка гипотез [Текст] : [учеб. пособие для вузов по физ.-мат. направлениям и специальностям] / А. А. Боровков. – 2-е изд., перераб. - М.: Физматлит, 2003. – 472 с.
2. **Вентцель, Елена Сергеевна.** Теория вероятностей [Текст] : [учеб. для втузов] / Е. С. Вентцель. - 11-е изд., стер. - М.: КНОРУС, 2010. - 658 с. - (Technology).
3. **Вентцель, Елена Сергеевна.** Задачи и упражнения по теории вероятностей [Текст] : [учеб. пособие для втузов] / Е. С. Вентцель, Л. А. Овчаров. - 8-е изд., стер. - М.: КНОРУС, 2010. - 493 с. - (Mathematics).
4. **Гмурман, Владимир Ефимович.** Теория вероятностей и математическая статистика [Текст]: [учеб. пособие для вузов] / В. Е. Гмурман. - 12-е изд., перераб. - М.: Высш. образование: Юрайт-Издат, 2009. - 479 с. - (Основы наук).
5. **Гмурман, Владимир Ефимович.** Руководство к решению задач по теории вероятностей и математической статистике [Текст]: [учеб. пособие для вузов] / В.Е. Гмурман. - Изд. 8-е, доп. - М.: Высш. образование: Юрайт-Издат, 2009. - 404 с.
6. **Ермаков, Сергей Михайлович.** Статистическое моделирование [Текст]: [учеб. пособие для вузов по физ.-мат. направлениям и специальностям] / С.М. Ермаков, Михайлов Г.А. – 3-е изд. перераб. - М.: Физматлит, 2002. – 396 с.
7. **Ивченко, Григорий Иванович.** Математическая статистика [Текст]: [учеб. пособие для втузов] / Г.И. Ивченко, Ю.И. Медведев – 3-е изд. перераб. - М.: «Высшая школа», 2001.- 248 с.
8. **Ивченко, Григорий Иванович.** Сборник задач по математической статистике [Текст]: [учеб. пособие для втузов] / Г.И. Ивченко, Ю.И. Медведев, А.В. Чистяков – 2-е изд. перераб. - М.: Высшая школа, 2001.- 255 с.
9. **Сборник задач** по теории вероятностей, математической статистике и теории случайных функций [Текст]: [учеб. пособие] / Б. Г. Володин и др.; под общей ред. А. А. Свешникова. - Изд. 3-е, перераб. - СПб. [и др.]: Лань, 2007. - 445 с. - (Учебники для вузов. Специальная литература) (Лучшие классические учебники.Математика) (Классические задачники и практикумы).
10. **Тюрин Юрий Николаевич.** Статистический анализ данных на компьютере [Текст]: [учеб. пособие] / Ю.Н. Тюрин, А.А. Макаров; под ред. Фигурнова В.Э. – изд. 2-е, перераб. - М.: ИНФРА-М, 2004. - 528 с.
11. **Харин, Юрий Семенович.** Практикум на ЭВМ по математической статистике [Текст]: [учеб. пособие для мат. спец. ун-тов] / Ю.С. Харин, М.Д. Степанова – 2-е изд. доп. - Минск: Изд-во «Университетское», 2003. – 304 с.

ПРИЛОЖЕНИЕ А
ВАРИАНТЫ ИНДИВИДУАЛЬНЫХ ЗАДАНИЙ

А.1 Исходные данные к разделу 1

Исходные данные к разделу 1 индивидуального задания представлены в таблице А.1, в которой n - объем выборки, a_X и σ_X^2 - математическое ожидание и дисперсия случайной величины X соответственно, γ - доверительная вероятность, α - уровень значимости.

Таблица А.1 – Исходные данные к разделу 1

Вариант	n	a_X	σ_X^2	γ	α
1	400	4.1	2.25	0.95	0.05
2	350	2.6	7.29	0.96	0.01
3	275	0.1	14.44	0.9	0.01
4	425	3.3	4.7	0.95	0.001
5	450	1.15	11.56	0.92	0.02
6	325	4.36	3.4	0.98	0.005
7	425	0.7	10.0	0.9	0.005
8	350	2.2	6.76	0.99	0.001
9	475	1.8	9.5	0.999	0.001
10	450	4.6	2.56	0.98	0.02
11	375	1.2	11.7	0.99	0.005
12	400	3.15	4.6	0.95	0.025
13	275	2.8	6.9	0.92	0.005
14	325	0.65	15.21	0.96	0.025
15	500	3.9	4.3	0.9	0.025
16	475	5.2	1.7	0.95	0.001
17	425	4.75	8.5	0.9	0.02
18	375	0.4	11.1	0.98	0.005
19	450	2.35	6.4	0.95	0.005
20	325	6.0	1.1	0.99	0.001
21	300	5.1	1.8	0.999	0.001
22	350	3.45	4.2	0.96	0.02
23	275	5.15	1.25	0.99	0.005
24	300	5.7	1.6	0.95	0.025
25	475	1.35	10.7	0.9	0.05
26	400	2.1	8.41	0.999	0.05
27	375	1.12	12.0	0.99	0.01
28	500	4.95	3.7	0.95	0.01
29	300	3.7	5.4	0.98	0.001
30	500	0.45	9.9	0.95	0.05

А.2 Исходные данные к разделу 2

Вариант № 1. ***Однопараметрическое показательное распределение***:

$$f(x)=\begin{cases} ae^{-ax}, & x\ge 0,\\ 0, & x<0. \end{cases}$$

Исходные данные: $a=3.5,\ \alpha=\ 0.001,\ \gamma=\ 0.96,\ n=300.$

Рассчитать аналитически: $\mathrm{M}X, \mathrm{D}X, F(x), F^{-1}(u)$.

Найти точечную оценку параметра a методом моментов.

Вариант № 2. ***Общее показательное распределение***:

$$f(x)=\begin{cases} ae^{-a(x-b)}, & x\ge b,\\ 0, & x<b. \end{cases}$$

Исходные данные: $a=2.2,\ b=\text{-}1.5,\ \alpha=\ 0.025,\ \gamma=\ 0.92,\ n=250.$

Рассчитать аналитически: $\mathrm{M}X, \mathrm{D}X, F(x), F^{-1}(u)$.

Найти точечные оценки параметров a и b методом моментов.

Вариант № 3. ***Однопараметрическое распределение Лапла***са:

$$f(x)=\frac{a}{2}\exp(-a|x|)\ .$$

Исходные данные: $a=0.3,\ \alpha=0.005,\ \gamma=0.94,\ n=400.$

Рассчитать аналитически: $\mathrm{M}X, \mathrm{D}X, F(x), F^{-1}(u)$.

Найти точечную оценку параметра a методом моментов.

Вариант № 4. ***Общее распределение Лапласа***:

$$f(x)=\frac{a}{2}\exp(-a|x-b|)\ .$$

Исходные данные: $a = 1.5$, $b = \text{-}2$, $\alpha = 0.005$, $\gamma = 0.94$, $n = 350$.

Рассчитать аналитически: $\mathrm{M}X, \mathrm{D}X, F(x), F^{-1}(u)$.

Найти точечные оценки параметров a и b методом моментов.

Вариант № 5. ***Распределение Релея:***

$$f(x)=\begin{cases} \dfrac{x}{a}e^{-\frac{x^2}{2a}}, & x\ge 0,\\ 0, & x<0. \end{cases}$$

Исходные данные: $a = 4$, $\alpha = 0.025$, $\gamma = 0.999$, $n = 375$.

Рассчитать аналитически: $\mathrm{M}X, \mathrm{D}X, F(x), F^{-1}(u)$.

Найти точечную оценку параметра a методом моментов.

Вариант № 6. ***Однопараметрическое логистическое распределение:***

$$f(x) = \frac{ae^{-ax}}{(1+e^{-ax})^2}.$$

Исходные данные: $a = 1/2$, $\alpha = 0.025$, $\gamma = 0.9$, $n = 400$.

Рассчитать аналитически: $\mathrm{M}X, \mathrm{D}X, F(x), F^{-1}(u)$.

Найти точечную оценку параметра a методом моментов.

Вариант № 7. ***Общее логистическое распределение:***

$$f(x) = \frac{ae^{-a(x-b)}}{\left[1+e^{-a(x-b)}\right]^2}.$$

Исходные данные: $a = 1/2$, $b = 3$, $\alpha = 0.025$, $\gamma = 0.95$, $n = 350$.

Рассчитать аналитически: $\mathrm{M}X, \mathrm{D}X, F(x), F^{-1}(u)$.

Найти точечные оценки параметров a и b методом моментов.

Вариант № 8. ***Распределение Эрланга k-го порядка:***

$$f(x) = \begin{cases} \dfrac{(ax)^{k-1}}{(k-1)!} a e^{-ax}, & x \ge 0, \\ 0, & x < 0. \end{cases}$$

Исходные данные: $a = 2$, $k = 3$, $\alpha = 0.025$, $\gamma = 0.9$, $n = 400$.

Рассчитать аналитически: $\mathrm{M}X, \mathrm{D}X$.

Найти точечные оценки параметров a и k методом моментов.

Указание: Воспользоваться тем, что моделируемая случайная величина $X = Y_1 + Y_2 + \ldots + Y_k$, где Y_k - независимые, распределенные по показательному закону с параметром a случайные величины, то есть:

$$f_{Y_k}(y) = a e^{-ay}, \quad y \ge 0.$$

Вариант № 9. ***Степенное распределение:***

$$f(x) = \begin{cases} a x^{a-1}, & x \in [0,1] \\ 0, & x \notin [0,1]. \end{cases}$$

Исходные данные: $a = 4$, $\alpha = 0.05$, $\gamma = 0.92$, $\mathrm{N} = 300$.

Рассчитать аналитически: $\mathrm{M}X, \mathrm{D}X, F(x), F^{-1}(u)$.

Найти точечную оценку параметра a методом моментов.

Вариант № 10. ***Степенное распределение:***

$$f(x)=\begin{cases} b(1-x)^{b-1}, & x\in[0,1], \\ 0, & x\notin[0,1]. \end{cases}.$$

Исходные данные: $b = 5$, $\alpha = 0.05$, $\gamma = 0.92$, N = 300.

Рассчитать аналитически: $\mathrm{M}X, \mathrm{D}X, F(x), F^{-1}(u)$.

Найти точечную оценку параметра b методом моментов.

Вариант № 11. ***Однопараметрическое распределение прямоугольного треугольника:***

$$f(x)=\begin{cases} ax, & 0\le x\le\sqrt{\dfrac{2}{a}}, \\ 0, & x<0,\ x>\sqrt{\dfrac{2}{a}}. \end{cases}$$

Исходные данные: $a = 0.5$, $\alpha = 0.1$, $\gamma = 0.98$, $n = 250$.

Рассчитать аналитически: $\mathrm{M}X, \mathrm{D}X, F(x), F^{-1}(u)$.

Найти точечную оценку параметра a методом моментов.

Вариант № 12. ***Однопараметрическое распределение прямоугольного треугольника:***

$$f(x)=\begin{cases} \dfrac{1}{a}\left(1-\dfrac{x}{2a}\right), & 0\le x\le 2a, \\ 0, & x<0,\ x>2a. \end{cases}$$

Исходные данные: $a = 3$, $\alpha = 0.1$, $\gamma = 0.98$, $n = 250$.

Рассчитать аналитически: $\mathrm{M}X, \mathrm{D}X, F(x), F^{-1}(u)$.

Найти точечную оценку параметра a методом моментов.

Вариант № 13. ***Общее распределение прямоугольного треугольника:***

$$f(x)=\begin{cases} \dfrac{2}{(b-a)^2}(x-a), & a\le x\le b, \\ 0, & x<a,\ x>b. \end{cases}$$

Исходные данные: $a = -2$, $b= 3$, $\alpha = 0.01$, $\gamma = 0.95$, $n = 250$.

Рассчитать аналитически: $\mathrm{M}X, \mathrm{D}X, F(x), F^{-1}(u)$.

Найти точечные оценки параметров a и b методом моментов.

Вариант № 14. ***Однопараметрическое распределение Симпсона:***

$$f(x)=\begin{cases}\dfrac{x}{a^2}, & 0\le x\le a,\\[2mm] \dfrac{2a-x}{a^2}, & a\le x\le 2a,\\[2mm] 0, & x<0,\quad x>2a.\end{cases}$$

Исходные данные: $a = 2$, $\alpha = 0.05$, $\gamma = 0.98$, $n = 300$.

Рассчитать аналитически: $\mathrm{M}X, \mathrm{D}X, F(x), F^{-1}(u)$.

Найти точечную оценку параметра a методом моментов.

Вариант № 15. ***Симметричное однопараметрическое распределение Симпсона:***

$$f(x)=\begin{cases}\dfrac{1}{a}\left(1-\dfrac{|x|}{a}\right), & |x|\le a,\\ 0\ , & |x|>a.\end{cases}$$

Исходные данные: $a = 2$, $\alpha = 0.05$, $\gamma = 0.95$, $n = 300$.

Рассчитать аналитически: $\mathrm{M}X, \mathrm{D}X, F(x), F^{-1}(u)$.

Найти точечную оценку параметра a методом моментов.

Вариант № 16. ***Общее распределение Симпсона:***

$$f(x)=\begin{cases}\dfrac{x-2a}{(b-a)^2}, & 2a\le x\le a+b,\\[2mm] \dfrac{2b-x}{(b-a)^2}, & a+b\le x\le 2b,\\[2mm] 0, & x<2a,\quad x>2b.\end{cases}$$

Исходные данные: a = -0.8, $b = 0.5$, $\alpha = 0.05$, $\gamma = 0.98$, $n = 300$.

Рассчитать аналитически: $\mathrm{M}X, \mathrm{D}X, F(x), F^{-1}(u)$.

Найти точечные оценки параметров a и b методом моментов.

Вариант № 17.

$$f(x)=\begin{cases}\dfrac{3}{2a}\sqrt{1-\dfrac{x}{a}}, & 0\le x\le a,\\ 0, & x<0,\ x>a.\end{cases}$$

Исходные данные: $a = 3$, $\alpha = 0.1$, $\gamma = 0.98$, $n = 350$.

Рассчитать аналитически: $\mathrm{M}X, \mathrm{D}X, F(x), F^{-1}(u)$.

Найти точечную оценку параметра a методом моментов.

Вариант № 18.

$$f(x)=\begin{cases}\dfrac{3}{4a}\sqrt{1-\dfrac{|x|}{a}}, & |x|\le a,\\ 0, & |x|>a.\end{cases}$$

Исходные данные: $a = 2$, $\alpha = 0.1$, $\gamma = 0.95$, $n = 350$.

Рассчитать аналитически: $\mathrm{M}X, \mathrm{D}X, F(x), F^{-1}(u)$.

Найти точечную оценку параметра a методом моментов.

Вариант № 19.

$$f(x)=\begin{cases}\dfrac{a}{x^2}, & \dfrac{a}{2}\le x\le a,\\ 0, & x<\dfrac{a}{2},\ x>a.\end{cases}$$

Исходные данные: $a = 2$, $\alpha = 0.1$, $\gamma = 0.98$, $n = 250$.

Рассчитать аналитически: $\mathrm{M}X, \mathrm{D}X, F(x), F^{-1}(u)$.

Найти точечную оценку параметра a методом моментов.

Вариант № 20.

$$f(x)=\begin{cases}\dfrac{a}{b}\left(\dfrac{b}{x}\right)^{a+1}, & x\ge b,\\ 0, & x<b,\end{cases}\qquad a>2, b>0.$$

Исходные данные: $a = 2$, $b = 10$, $\alpha = 0.1$, $\gamma = 0.98$, $n = 450$.

Рассчитать аналитически: $\mathrm{M}X, \mathrm{D}X, F(x), F^{-1}(u)$.

Найти точечные оценки параметров a и b методом моментов.

Вариант № 21.

$$f(x)=\begin{cases}\dfrac{ab}{b-a}x^{a-1}, & x\in[0,1),\\ \dfrac{ab}{b-a}x^{b-1}, & x\in[1,+\infty),\end{cases}\qquad a>1,\ b<-2.$$

Исходные данные: $a = 2$, $b = -3$, $\alpha = 0.002$, $\gamma = 0.96$, $n = 350$.

Рассчитать аналитически: $\mathrm{M}X, \mathrm{D}X, F(x), F^{-1}(u)$, считая параметр a известным. Найти точечную оценку параметра b методом моментов.

Вариант № 22.

$$f(x)=\begin{cases}\dfrac{ab}{b-a}x^{a-1}, & x\in[0,1),\\ \dfrac{ab}{b-a}x^{b-1}, & x\in[1,+\infty),\end{cases}\qquad a>1,\ b<-2.$$

Исходные данные: $a = 3$, $b = -3$, $\alpha = 0.002$, $\gamma = 0.98$, $n = 350$.

Рассчитать аналитически: $\mathrm{M}X, \mathrm{D}X, F(x), F^{-1}(u)$, считая параметр b известным. Найти точечную оценку параметра a методом моментов.

Вариант № 23. ***Распределение арксинуса:***

$$f(x)=\begin{cases}\dfrac{1}{\pi\sqrt{a^2-x^2}}, & |x|<a,\\ 0, & |x|\ge a.\end{cases}$$

Исходные данные: $a = 1/2$, $\alpha = 0.025$, $\gamma = 0.94$, $n = 400$.

Рассчитать аналитически: $\mathrm{M}X, \mathrm{D}X, F(x), F^{-1}(u)$.

Найти точечную оценку параметра a методом моментов.

Вариант *№ 24.*

$$f(x)=\begin{cases}\dfrac{a}{2}\sin(ax+b), & x\in\left[-\dfrac{b}{a};\dfrac{\pi-b}{a}\right],\\ 0, & x\notin\left[-\dfrac{b}{a};\dfrac{\pi-b}{a}\right].\end{cases}$$

Исходные данные: $a = 2$, $b = 1$, $\alpha = 0.05$, $\gamma = 0.98$, $n = 400$.

Рассчитать аналитически: $\mathrm{M}X, \mathrm{D}X, F(x), F^{-1}(u)$.

Найти точечные оценки параметров a и b методом моментов.

Вариант *№ 25.*

$$f(x)=\begin{cases}\dfrac{a}{2}\cos(ax+b), & x\in\left[-\dfrac{\pi+2b}{2a};\dfrac{\pi-2b}{2a}\right],\\ 0, & x\notin\left[-\dfrac{\pi+2b}{2a};\dfrac{\pi-2b}{2a}\right].\end{cases}$$

Исходные данные: $a = 1.5$, $b = -1$, $\alpha = 0.01$, $\gamma = 0.98$, $n = 400$.

Рассчитать аналитически: $\mathrm{M}X, \mathrm{D}X, F(x), F^{-1}(u)$.

Найти точечные оценки параметров a и b методом моментов.

Вариант ***№ 26.***

$$f(x)=\begin{cases}\dfrac{a}{2}\cos\sqrt{ax}\,\dfrac{1}{\sqrt{ax}}, & x\in\left(0;\dfrac{\pi^2}{4a}\right),\\ 0, & x\notin\left(0;\dfrac{\pi^2}{4a}\right).\end{cases}$$

Исходные данные: $a = 1/2$, $\alpha = 0.01$, $\gamma = 0.96$, $n = 300$.

Рассчитать аналитически: $\mathrm{M}X, \mathrm{D}X, F(x), F^{-1}(u)$.

Найти точечную оценку параметра a методом моментов.

Вариант ***№ 27.***

$$f(x)=\begin{cases}\dfrac{a}{2}\sin\sqrt{ax}\,\dfrac{1}{\sqrt{ax}}, & x\in\left(0;\dfrac{\pi^2}{4a}\right),\\ 0, & x\notin\left(0;\dfrac{\pi^2}{4a}\right).\end{cases}$$

Исходные данные: $a = 1/2$, $\alpha = 0.001$, $\gamma = 0.95$, $n = 300$.

Рассчитать аналитически: $\mathrm{M}X, \mathrm{D}X, F(x), F^{-1}(u)$.

Найти точечную оценку параметра а методом моментов.

Вариант № 28. ***Гамма-распределение:***

$$f(x)=\begin{cases}\dfrac{x^{\lambda-1}}{\Gamma(\lambda)}a^{\lambda}e^{-ax}, & x\geq 0,\\ 0, & x<0,\end{cases}$$

где $\Gamma(\lambda)=\int\limits_0^{\infty}x^{\lambda-1}e^{-x}dx$ - гамма-функция.

Исходные данные: $a = 2$, $\lambda = 3/2$, $\alpha = 0.01$, $\gamma = 0.99$, $n = 400$.

Рассчитать аналитически: $\mathrm{M}X, \mathrm{D}X$.

Найти точечные оценки параметров a и λ методом моментов.

Указание: Воспользоваться алгоритмом моделирования для случая, когда $\lambda = m+\dfrac{1}{2}, m = 0,1,2,...$ или общим алгоритмом моделирования гамма-распределения.

Вариант № 29. ***Общее бета-распределение:***

$$f(x)=\begin{cases}\dfrac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}x^{a-1}(1-x)^{b-1}, & 0\le x\le 1,\\ 0, & x<0,\, x>1.\end{cases}$$

Исходные данные: a = 2, b = 2, α = 0.05, γ = 0.95, n = 300.

Рассчитать аналитически: $\mathrm{M}X, \mathrm{D}X$.

Найти точечные оценки параметров a и b методом моментов.

Указание: Воспользоваться алгоритмом моделирования для случая целых a и b или общим алгоритмом моделирования бета-распределения.

Вариант № 30. ***Распределение Парето:***

$$f(x)=\begin{cases}\dfrac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}\dfrac{x^{b-1}}{(1+x)^{a+b}}, & x\ge 0,\\ 0, & x<0.\end{cases}$$

Исходные данные: a = 2, b = 2, α = 0.05, γ = 0.92, n = 300.

Рассчитать аналитически: $\mathrm{M}X, \mathrm{D}X$.

Найти точечные оценки параметров a и b методом моментов.

Указание: Воспользоваться тем, что моделируемая случайная величина $X=Y^{-1}-1$, где Y - случайная величина, имеющая бета-распределение с параметрами a и b.

А.3 Исходные данные к разделу 3

Исходные данные к разделу 3 индивидуального задания представлены в таблице А.2, в которой a_Y и σ_Y^2 - математическое ожидание и дисперсия случайной величины Y соответственно, r_{XY} - коэффициент корреляции случайных величин X и Y. При выполнении задания значения параметров $n, a_X, \sigma_X^2, \alpha$ следует заимствовать из таблицы А.1.

Таблица А.2 – Исходные данные к разделу 3

Вариант	a_Y	σ_Y^2	r_{XY}	Вариант	a_Y	σ_Y^2	r_{XY}
1	1.12	11.56	-0.21	**16**	2.35	7.5	0.24
2	-5.6	1.69	-0.15	**17**	-1.4	10.2	0.3
3	-3.3	4.7	0.74	**18**	3.7	4.41	-0.425
4	0.45	14.0	0.25	**19**	5.9	1.21	0.49
5	4.25	3.24	-0.33	**20**	2.2	6.25	0.045
6	-1.5	9.9	0.41	**21**	-2.35	8.0	-0.235
7	3.35	4.6	-0.29	**22**	-0.1	15.21	0.415
8	-5.4	1.75	0.12	**23**	2.4	8.5	-0.45
9	-4.5	3.61	-0.18	**24**	2.95	9.0	0.07
10	1.3	9.3	-0.63	**25**	-4.21	2.1	-0.39
11	4.48	2.56	0.08	**26**	5.25	1.44	0.14
12	0.5	10.7	-0.85	**27**	4.3	3.8	-0.54
13	-5.44	1.36	0.5	**28**	1.7	12.96	0.315
14	3.8	5.45	-0.37	**29**	-0.55	14.44	-0.125
15	-0.7	12.96	-0.05	**30**	-3.7	4.3	0.09

ПРИЛОЖЕНИЕ Б

НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Таблица Б.1 - Значения функции Лапласа $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{u^2}{2}} du, \ x \geq 0.$

x	*Ф(x)*	*x*	*Ф(x)*	*x*	*Ф(x)*	*x*	*Ф(x)*
0,00	0,5000	0,32	0,6255	0,64	0,7389	0,96	0,8315
0,01	0,5040	0,33	0,6293	0,65	0,7422	0,97	0,8340
0,02	0,5080	0,34	0,6331	0,66	0,7454	0,98	0,8365
0,03	0,5120	0,35	0,6368	0,67	0,7486	0,99	0,8389
0,04	0,5160	0,36	0,6406	0,68	0,7516	1,00	0,8413
0,05	0,5199	0,37	0,6443	0,69	0,7549	1,01	0,8438
0,06	0,5239	0,38	0,6480	0,70	0,7580	1,02	0,8461
0,07	0,5279	0,39	0,6517	0,71	0,7611	1,03	0,8485
0,08	0,5319	0,40	0,6554	0,72	0,7642	1,04	0,8508
0,09	0,5359	0,41	0,6591	0,73	0,7673	1,05	0,8531
0,10	0,5398	0,42	0,6628	0,74	0,7703	1,06	0,8554
0,11	0,5438	0,43	0,6664	0,75	0,7734	1,07	0,8577
0,12	0,5478	0,44	0,6700	0,76	0,7764	1,08	0,8599
0,13	0,5517	0,45	0,6736	0,77	0,7794	1,09	0,8621
0,14	0,5557	0,46	0,6772	0,78	0,7823	1,10	0,8643
0,15	0,5596	0,47	0,6808	0,79	0,7852	1,11	0,8665
0,16	0,5636	0,48	0,6844	0,80	0,7881	1,12	0,8686
0,17	0,5675	0,49	0,6879	0,81	0,7910	1,13	0,8708
0,18	0,5714	0,50	0,6915	0,82	0,7939	1,14	0,8729
0,19	0,5753	0,51	0,6950	0,83	0,7967	1,15	0,8749
0,20	0,5793	0,52	0,6985	0,84	0,7995	1,16	0,8770
0,21	0,5832	0,53	0,7019	0,85	0,8023	1,17	0,8790
0,22	0,5871	0,54	0,7054	0,86	0,8051	1,18	0,8810
0,23	0,5910	0,55	0,7088	0,87	0,8078	1,19	0,8830
0,24	0,5948	0,56	0,7123	0,88	0,8106	1,20	0,8849
0,25	0,5987	0,57	0,7157	0,89	0,8133	1,21	0,8869
0,26	0,6026	0,58	0,7190	0,90	0,8159	1,22	0,8883
0,27	0,6064	0,59	0,7224	0,91	0,8186	1,23	0,8907
0,28	0,6103	0,60	0,7257	0,92	0,8212	1,24	0,8925
0,29	0,6141	0,61	0,7291	0,93	0,8228	1,25	0,8944
0,30	0,6179	0,62	0,7324	0,94	0,8264	1,26	0,8962
0,31	0,6217	0,63	0,7357	0,95	0,8289	1,27	0,8980

Продолжение таблицы Б.1

x	*Ф(x)*	*x*	*Ф(x)*	*x*	*Ф(x)*	*x*	*Ф(x)*
1,28	0,8997	1,61	0,9463	1,94	0,9738	2,54	0,9945
1,29	0,9015	1,62	0,9474	1,95	0,9744	2,56	0,9948
1,30	0,9032	1,63	0,9484	1,96	0,9750	2,58	0,9951
1,31	0,9049	1,64	0,9495	1,97	0,9756	2,60	0,9953
1,32	0,9066	1,65	0,9505	1,98	0,9761	2,62	0,9956
1,33	0,9082	1,66	0,9515	1,99	0,9767	2,64	0,9959
1,34	0,9099	1,67	0,9525	2,00	0,9772	2,66	0,9961
1,35	0,9115	1,68	0,9535	2,02	0,9783	2,68	0,9963
1,36	0,9131	1,69	0,9545	2,04	0,9793	2,70	0,9965
1,37	0,9147	1,70	0,9554	2,06	0,9803	2,72	0,9967
1,38	0,9162	1,71	0,9564	2,08	0,9812	2,74	0,9969
1,39	0,9177	1,72	0,9573	2,10	0,9821	2,76	0,9971
1,40	0,9192	1,73	0,9582	2,12	0,9830	2,78	0,9973
1,41	0,9207	1,74	0,9591	2,14	0,9838	2,80	0,9974
1,43	0,9236	1,76	0,9608	2,18	0,9854	2,84	0,9977
1,44	0,9251	1,77	0,9616	2,20	0,9861	2,86	0,9979
1,45	0,9265	1,78	0,9625	2,22	0,9868	2,88	0,9980
1,46	0,9279	1,79	0,9633	2,24	0,9875	2,90	0,9981
1,47	0,9292	1,80	0,9641	2,26	0,9881	2,92	0,9982
1,48	0,9306	1,81	0,9649	2,28	0,9887	2,94	0,9984
1,49	0,9319	1,82	0,9656	2,30	0,9893	2,96	0,9985
1,50	0,9332	1,83	0,9664	2,32	0,9898	2,98	0,9986
1,51	0,9345	1,84	0,9671	2,34	0,9904	3,00	0,99865
1,52	0,9357	1,85	0,9678	2,36	0,9909	3,20	0,99931
1,53	0,9370	1,86	0,9686	2,38	0,9913	3,40	0,99966
1,54	0,9382	1,87	0,9693	2,40	0,9918	3,60	0,999841
1,55	0,9394	1,88	0,9699	2,42	0,9922	3,80	0,999928
1,56	0,9406	1,89	0,9706	2,44	0,9927	4,00	0,999968
1,57	0,9418	1,90	0,9713	2,46	0,9931	4,50	0,999997
1,58	0,9429	1,91	0,9719	2,48	0,9934	5,00	0,999997
1,59	0,9441	1,92	0,9726	2,50	0,9938		
1,60	0,9452	1,93	0,9732	2,52	0,9941		

Примечание: $\Phi(-x) = 1 - \Phi(x), \quad x \geq 0$.

Таблица Б.2 - Квантили нормального распределения $\tilde{n}_p : \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{c_p} \exp(-\frac{x^2}{2})\, dx = p$.

p	c_p	p	c_p	p	c_p
0,50	0.000	0,860	1,080	0,9910	2,366
0,55	0,126	0,870	1,126	0,9920	2,409
0,60	0,253	0,880	1,175	0,9930	2,457
0,65	0,385	0,890	1,227	0,9940	2,512
0.70	0,524	0,900	1,282	0,9950	2,576
0,75	0,674	0,910	1,341	0,9955	2,612
0,76	0,706	0,920	1,405	0,9960	2,652
0,77	0,739	0,930	1.476	0,9965	2,697
0,78	0,772	0,940	1,555	0,9970	2,748
0,79	0,806	0,950	1,645	0,9975	2,807
0,80	0,842	0,960	1,751	0,9980	2,878
0,81	0,878	0,970	1,881	0,9985	2,968
0,82	0,915	0,975	1,960	0,9990	3,090
0,83	0,954	0,980	2,051	0,9995	3,291
0,84	0,994	0,985	2,170	0,9999	3,720
0,85	1,036	0,990	2,326	0,99999	4,265

Примечание: Если $0<p<0.5$, то $c_p = -c_{1-p}$.

ПРИЛОЖЕНИЕ В

РАСПРЕДЕЛЕНИЕ СТЬЮДЕНТА $S(n)$

Таблица В.1 - Квантили распределения Стьюдента $t_{p,n} : \int_{-\infty}^{t_{p,n}} s_n(x)\,dx = p$

Число степеней свободы n	Вероятность, p					
	0,9	0,95	0,975	0,99	0,995	0,9995
1	3,078	6,314	12,706	31,821	63,657	636,619
2	1,886	2,920	4,303	6,965	9,925	31,598
3	1,638	2,353	3,182	4,541	5,841	12,941
4	1,533	2,132	2,776	3,747	4,604	8,610
5	1,476	2,015	2,571	3,365	4,032	6,869
6	1,440	1,943	2,447	3,143	3,707	5,959
7	1,415	1.895	2,365	2,998	3,499	5,405
8	1,397	1,860	2,306	2,896	3,355	5,041
9	1,383	1,833	2,262	2,821	3,250	4,781
10	1,372	1,812	2,228	2,764	3,169	4,587
12	1,356	1,782	2,179	2,681	3,055	4,318
14	1,345	1,761	2,145	2,625	2,977	4,140
16	1,337	1,746	2,120	2,584	2,921	4,015
18	1,330	1,734	2,101	2,552	2,878	3,922
20	1,325	1,725	2,086	2,528	2,845	3,850
22	1,321	1,717	2,074	2,508	2,819	3,792
24	1,318	1,711	2,064	2,492	2,797	3,745
26	1,315	1,706	2,056	2,479	2,779	3,707
28	1,313	1,701	2,048	2,467	2,763	3,674
30	1,310	1,697	2,042	2,457	2,750	3,646
60	1,296	1,671	2,000	2,390	2,660	3,460
120	1,289	1,658	1,980	2,358	2,617	3,373
∞	1,282	1,645	1,960	2,326	2,576	3,291

Примечание: Плотность вероятностей распределения Стьюдента $S(n)$ имеет вид:

$$s_n(n) = \frac{\Gamma((n+1)/2)}{\Gamma(n/2)\sqrt{\pi n}}\left(1+\frac{x^2}{n}\right)^{-(n+1)/2}.$$

ПРИЛОЖЕНИЕ Г

РАСПРЕДЕЛЕНИЕ ХИ-КВАДРАТ $\chi^2(n)$

Таблица Г.1 - Квантили распределения хи-квадрат $\chi^2_{p,n}: \int_{-\infty}^{\chi^2_{p,n}} k_n(x)\,dx = p$

Число степеней свободы n	Вероятность, p						
	0,001	0,005	0,01	0,02	0,025	0,05	0,1
1	0,0000016	0,000039	0,00016	0,00063	0,00098	0,004	0,016
2	0,002	0,010	0,020	0,040	0,051	0,103	0,211
3	0,024	0,072	0,115	0,185	0,216	0,352	0,584
4	0,091	0,207	0,297	0,429	0,484	0,711	1,064
5	0,210	0,412	0,554	0,752	0,831	1,145	1,610
6	0,381	0,676	0,872	1,134	1,237	1,635	2,204
7	0,598	0,989	1,239	1,564	1,690	2,167	2,833
8	0,857	1,344	1,646	2,032	2,180	2,733	3,490
9	1,152	1,735	2,008	2,532	2,700	3,325	4,168
10	1,479	2,156	2,558	3,059	3,247	3,940	4,865
11	1,834	2,603	3,053	3,609	3,816	4,575	5,578
12	2,214	3,074	3,571	4.178	4,404	5,226	6,304
13	2,617	3,565	4,107	4,765	5,009	5,892	7,042
14	3,041	4,075	4,660	5,368	5,629	6,571	7,790
15	3,483	4,601	5,229	5,985	6,262	7,261	8,547
16	3,942	5,142	5,812	6,614	6,908	7,962	9,312
17	4.416	5.697	6.408	7,255	7,564	8,672	10,085
18	4,905	6,265	7,015	7,096	8,231	9,390	10,865
19	5,407	6,844	7,633	8,567	8,907	10,117	11,651
20	5,921	7,434	8,260	9,237	9,591	10,851	12,443
21	6,447	8,034	8,897	9,915	10,283	11,591	13,240
22	6,983	8,643	9,542	10,600	10,982	12,338	14,041
23	7.529	9,260	10,196	11,293	11,688	13,091	14,848
24	8,085	9,886	10,856	11,992	12,401	13,848	15,659
25	8,649	10,520	11,524	12,697	13,120	14,611	16,473
26	9,222	11,160	12,198	13,409	13,844	15,379	17,292
27	9,803	11,808	12,879	14,125	14,573	16,151	18,114
28	10,391	12,461	13,565	14,847	15,308	16,928	18.939
29	10,986	13,121	14,256	15,574	16,047	17,708	19,768
30	11,588	13,787	14,953	16,306	16,791	18,493	20,599

Продолжение таблицы Г.1

Число степеней свободы n	Вероятность, p						
	0,9	0,95	0,975	0,98	0,99	0,995	0,999
1	2,706	3,841	5,024	5,412	6,635	7,879	10,827
2	4,605	5,991	7,378	7,824	9,210	10,597	13,815
3	6,251	7,815	9,348	9,837	11,345	12,838	16,268'
4	7,779	9,488	11,143	11.668	13,277	14,860	18,465
5	9,236	11,070	12,832	13,388	15,086	16,750	20,517
6	10,645	12,592	14,449	15,033	16,812	18,548	22,457
7	12,017	14,067	16,013	16,622	18,475	20,278	24,322
8	13,362	15,507	17,535	18,168	20,090	21,955	26,125
9	14,684	16,919	19,023	19,679	21,666	23,589	27,877
10	15,987	18,307	20,483	21,161	23,209	25,188	29,588
11	17,275	19,675	21,920	22,618	24,725	26,757	31,264
12	18,549	21,026	23,337	24,054	26,217	28,300	32,909
13	19,812	22,362	24,736	25,472	27,688	29,819	34,528
14	21,064	23,685	26,119	26,873	29,141	31,319	36,123
15	22,307	24,996	27,488	28,259	30,578	32,801	37,697
16	23,542	26,296	28,845	29,633	32,000	34,267	39,252
17	24.769	27,587	30,191	30,995	33,409	35,718	40,790
18	25,989	28.809	31,526	32,346	34,805	37,156	42,312
19	27,204	30,144	32,852	33,687	36,191	38,582	43,820
20	28,412	31,410	34,170	35,020	37,566	39,997	45,315
21	29,615	32,671	35,479	36,343	38,932	41,401	46,797
22	30,813	33,924	36,781	37,659	40,289	42,796	48,268
23	32,007	35,172	38,076	38,968	41,638	44,181	49,728
24	33,196	36,415	39,364	40,270	42,980	45,558	51,179
25	34,382	37,652	40,646	41,566	44,314	46,928	52,620
26	35,563	38,885	41,923	42,856	45,642	48,290	54,052
27	36,741	40,113	43,194	44,140	46,963	49,645	55,476
28	37,916	41.337	44,461	45,419	48,278	50,993	56,893
29	39,087	42,557	45,722	46,693	49,588	52,336	58,302
30	40,256	43,773	46,9.79	47,962	50,892	53,672	59,703

Примечание 1. Плотность вероятностей распределения хи-квадрат $\chi^2(n)$ имеет вид:

$$k_n(n) = \frac{1}{\Gamma(n/2)2^{n/2}} x^{(n/2)-1} \exp(-x/2), \quad x > 0 .$$

Примечание 2: При $n \geq 30$ $\chi^2_{p,n} \approx \left(c_p + \sqrt{2n-1}\right)^2 \Big/ 2$, где c_p — квантиль нормального распределения (см. таблицу Б.2).

ПРИЛОЖЕНИЕ Д

ОБРАЗЕЦ ОФОРМЛЕНИЯ ТИТУЛЬНОГО ЛИСТА ОТЧЕТА

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО
ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«САМАРСКИЙ ГОСУДАРСТВЕННЫЙ АЭРОКОСМИЧЕСКИЙ
УНИВЕРСИТЕТ ИМЕНИ АКАДЕМИКА С.П. КОРОЛЕВА
(НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)» (СГАУ)

КАФЕДРА «ТЕХНИЧЕСКАЯ КИБЕРНЕТИКА»

ОТЧЕТ
ОБ ИНДИВИДУАЛЬНОМ ЗАДАНИИ ПО КУРСУ
«ТЕОРИЯ ВЕРОЯТНОСТЕЙ
И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА»

МОДЕЛИРОВАНИЕ И СТАТИСТИЧЕСКИЙ АНАЛИЗ
СЛУЧАЙНЫХ ДАННЫХ

Студент_____________________
Группа _____________________
Руководитель ________________
Оценка _____________________

САМАРА
2011