Министерство образования и науки Российской Федерации
НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

51 № 3304
К637

Компьютерные технологии анализа данных и исследования статистических закономерностей

Методические указания
к выполнению лабораторных работ
для студентов V-го курса ФПМИ по направлению 010500
и специальности 010501 дневного отделения

НОВОСИБИРСК
2007

УДК 519.23:004(076.5)
К 637

Методические указания предназначены для студентов, выполняющих лабораторные работы по курсу "Компьютерные технологии анализа данных и исследования статистических закономерностей" (направление 010500 и специальность 010501). Указания содержат необходимые сведения для выполнения лабораторных работ, порядок выполнения, варианты заданий и контрольные вопросы, по которым осуществляется их защита.

Составители: доктор техн. наук., проф. *Б.Ю. Лемешко*,
канд. техн. наук, доц. *С.Н. Постовалов*,
аспирант *С.Б. Лемешко*

Рецензент: канд. техн. наук, доц. *Е.В. Чимитова*

Работа подготовлена на кафедре
прикладной математики

© Новосибирский государственный
технический университет, 2007

Содержание

ВВЕДЕНИЕ....4
Лабораторная работа № 1. Изучение методики компьютерного моделирования статистических закономерностей 5
Лабораторная работа № 2. Исследование свойств оценок параметров распределений вероятностей по эмпирическим данным....8
Лабораторная работа № 3. Экспериментальное исследование робастности оценок 11
Лабораторная работа № 4. Непараметрическое оценивание плотности распределения вероятностей....15
Лабораторная работа № 5. Исследование оценок максимального правдоподобия параметров распределений вероятностей по цензурированным данным....17
Лабораторная работа № 6. Экспериментальное исследование свойств критерия согласия χ^2 Пирсона. 20
Лабораторная работа № 7. Исследование распределения статистики и мощности критерия Рао-Робсона-Никулина....23
Лабораторная работа № 8. Экспериментальное исследование предельных распределений статистик непараметрических критериев согласия. 25
Лабораторная работа № 9. Исследование критериев проверки отклонения от нормального закона....27
Лабораторная работа № 10. Исследование распределений статистик в регрессионном анализе при ошибках измерений, отличных от нормального....30
Лабораторная работа № 11. Исследование статистик многомерного анализа....35
Лабораторная работа № 12. Исследование распределений статистик критериев проверки однородности выборок 42
Лабораторная работа № 13. Исследование распределений статистик критериев проверки гипотез о математических ожиданиях и дисперсиях....44
Лабораторная работа № 14. Исследование распределений статистик критериев проверки однородности средних....47
Лабораторная работа № 15. Исследование распределений статистик критериев проверки однородности дисперсий....51
Литература 54
Приложение. Законы распределения наблюдаемых случайных величин 55

ВВЕДЕНИЕ

В лабораторных работах по компьютерным технологиям анализа данных и исследования статистических закономерностей применяется методика компьютерного моделирования фундаментальных статистических закономерностей. В основе методики лежит использование метода Монте-Карло для моделирования эмпирических распределений некоторых функций от случайных величин. Преимуществом метода Монте-Карло является несложность реализации процедур моделирования сложных статистических закономерностей, обычно не поддающихся определению аналитическими методами.

Выполнение работ носит исследовательский характер и в большинстве случаев опирается на разработанное программное обеспечение.

При выполнении лабораторных работ необходимо учитывать точность полученных результатов. Поскольку в большинстве работ требуется сравнивать эмпирические распределения, то малые объемы моделируемых выборок статистик или оценок могут нивелировать разницу в статистических методах, а иногда и давать некорректные (противоположные аналитическим выводам) результаты.

Согласие эмпирических распределений с теоретическими следует проверять, опираясь на статистические критерии, так как сравнение графиков распределений "на глаз" является субъективной процедурой и не учитывает возможной статистической погрешности, которая зависит от объемов выборок.

Отчет по лабораторной работе должен содержать цель, задание на выполнение, результаты исследований в соответствии с заданием, необходимые таблицы, графики, иллюстрирующие результаты, анализ результатов, общие выводы по работе.

Основной теоретический материал, необходимый для выполнения работ, изложен в учебном пособии [1], дополнительные материалы и программное обеспечение размещены на персональных сайтах авторов http://ami.nstu.ru/~headrd и http://ami.nstu.ru/~post.

[Содержание]

Лабораторная работа № 1. Изучение методики компьютерного моделирования статистических закономерностей

Цель работы. Изучение методики компьютерного моделирования статистических закономерностей. Оценка точности моделирования и определение необходимого объема выборки.

Методические указания

1. Метод Монте-Карло. Идея метода заключается в следующем. Вместо того чтобы описывать исследуемый случайный процесс аналитически, составляется алгоритм, *имитирующий* этот процесс. В алгоритм включаются специальные процедуры для моделирования случайности. Конкретное осуществление алгоритма складывается каждый раз по-иному, со своими результатами. Множество реализаций алгоритма используется как некий искусственно полученный статистический материал, обработав который методами *математической статистики*, можно получить любые характеристики: вероятности событий, математические ожидания, дисперсии случайных величин.

2. Методика компьютерного моделирования статистических закономерностей. Несмотря на то, что впервые метод Монте-Карло был использован для решения задач математической физики, весьма плодотворным он оказывается для выявления и проверки *фундаментальных закономерностей теории вероятностей и математической статистики*.

Для того чтобы найти модель закона распределения некоторой статистики проводится многократное моделирование случайной величины [2] по некоторому известному (заданному) закону. Далее, по полученным выборкам вычисляются значения статистики, и, наконец, по эмпирическому распределению статистики подбирается аналитическая модель.

Реализация описанной процедуры компьютерного моделирования распределения статистики в настоящий момент не содержит ни принципиальных, ни практических трудностей, а современный уровень вычислительной техники позволяет быстро получить результаты моделирования.

3. Точность и количество реализаций. Точность вычисления значения $F(x)$ с помощью эмпирической функции распределения $F_N(x)$ по выборке объема N можно определить по формуле

$$\varepsilon = t_\gamma \frac{\sqrt{F(x)(1-F(x))}}{\sqrt{N}}, \qquad (1.1)$$

где $t_\gamma = \Phi^{-1}\left(\frac{\gamma+1}{2}\right)$ - квантиль стандартного нормального распределения, и $P\{|F(x)-F_N(x)|<\varepsilon\}=\gamma$.

Отсюда можно определить необходимый объем выборки для вычисления $F(x)$ с заданной точностью

$$N = t_\gamma^2 \frac{F(x)(1-F(x))}{\varepsilon^2}. \qquad (1.2)$$

Порядок выполнения

1. Выполнить постановку задачи компьютерного моделирования закона распределения по варианту задания (определить по какому закону моделируется выборка, какая статистика вычисляется).
2. Написать программу для моделирования закона распределения (или некоторой его числовой характеристики).
3. Вычислить необходимый объем выборки N для заданной точности $\varepsilon = 0.01$.
4. Выполнить моделирование распределения статистики, провести исследования согласно варианту задания.
5. Идентифицировать закон распределения с использованием программной системы ISW. Сравнить полученные результаты с теоретическими.

Варианты заданий

1. Электричка в метро состоит из трех вагонов по k мест каждый. Три студента *равновероятно* садятся на любое место в любом вагоне. Найти распределение случайной величины – «количество вагонов, в которых окажутся студенты». Исследовать это распределение при увеличении числа k от 1 до бесконечности.
2. В городе проживает $n+1$ человек. Один из них, узнав новость, сообщает ее другому, тот – третьему, и т.д., причем каждый человек передает новость *наугад* выбранному жителю, за исключением того от, которого он ее услышал. Пусть r – случайная величина, равная числу передач новости от одного человека к другому до момента возвращения к тому человеку, который узнал ее первым. Найти ее распределение, вычислить среднее значение и дисперсию, исследовать зависимость от n.
3. Для оплаты проезда пассажир пользуется двумя электронными карточками, оплачивая проезд наугад выбранной карточкой. Через некоторое время он обнаружил, что на одной из карточек осталось ноль поездок. Найти распределение случайной величины – «количество поездок, оставшихся на второй карточке», если считать, что на каждой карточке изначально было n поездок. Вычислить среднее значение и дисперсию случайной величины, исследовать зависимость от n.
4. Окружность единичного радиуса с центром в нуле имеет северный полюс на положительной полуоси абсцисс. Из полюса случайным образом направлен луч, причем его угол с осью абсцисс распределен равномерно на отрезке $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. Найти функцию распределения длины хорды внутри окружности.

5. Равнобедренный треугольник образован единичным вектором в направлении оси абсцисс и единичным вектором в случайном направлении. Найти функцию распределения третьей стороны в $\mathbf{R}^2$.
6. Равнобедренный треугольник образован единичным вектором в направлении оси абсцисс и единичным вектором в случайном направлении. Найти функцию распределения третьей стороны в $\mathbf{R}^3$.
8. Найти распределение выборочного квантиля порядка α по выборке объема n наблюдений *на примере экспоненциального распределения*, и исследовать его изменение с ростом объема выборки и величины α.

Примечание: В вариантах 9-19, $X_1, X_2, \ldots, X_n$ – выборка случайной величины ξ, подчиненной закону распределения с функцией распределения $F(x)$. Требуется найти распределение следующих статистик и исследовать его изменение с ростом объема выборки. Проверить зависит ли полученное распределение статистики от закона распределения случайной величины, а также от значений параметров распределения.

9. $\sqrt{n}\dfrac{\bar{X} - M\xi}{\sqrt{D\xi}}$.

10. $\sqrt{n}\dfrac{\left|\bar{X} - M\xi\right|}{\sqrt{D\xi}}$.

11. $\dfrac{1}{n\sqrt{D\xi}}\sum_{i=1}^{n} X_i - M\xi^{\ 2}$.

12. $\dfrac{1}{n\sqrt{D\xi}}\sum_{i=1}^{n} X_i - \bar{X}^{\ 2}$.

13. $\sqrt{n-1}\dfrac{\bar{X}}{S^2}$.

14. $\sqrt{n}D_n$, где $D_n = \max(D_n^+, D_n^-)$,
$D_n^+ = \max\limits_{1\le i\le n}\left\{\dfrac{i}{n} - F(X_{(i)},\theta)\right\}, D_n^- = \max\limits_{1\le i\le n}\left\{F(X_{(i)},\theta) - \dfrac{i-1}{n}\right\}$.

15. $\dfrac{6nD_n + 1}{6\sqrt{n}}$.

16. $\dfrac{(6nD_n^+ + 1)^2}{9n}$.

17. $\dfrac{(6nD_n^- + 1)^2}{9n}$.

18. $\dfrac{1}{12n} + \sum_{i=1}^{n}\left\{F(X_{(i)},\theta) - \dfrac{2i-1}{2n}\right\}^2$.

19. $-n - 2\sum_{i=1}^{n}\left\{\dfrac{2i-1}{2n}\ln F(X_{(i)},\theta) + \left(1 - \dfrac{2i-1}{2n}\right)\ln(1 - F(X_{(i)},\theta))\right\}$.

Контрольные вопросы

1. Предложите алгоритм для вычисления числа e методом Монте-Карло.
2. Сформулируйте достоинства и недостатки метода Монте-Карло.
3. В чем заключается методика компьютерного моделирования статистических закономерностей?

[Назад][Вперед][Содержание]

Лабораторная работа № 2. Исследование свойств оценок параметров распределений вероятностей по эмпирическим данным

Цель работы. Вычисление оценок параметров распределений вероятностей по эмпирическим данным различными методами. Исследование асимптотических свойств оценок методом Монте-Карло.

Методические указания

1. Постановка задачи. Пусть в эксперименте наблюдается непрерывная случайная величина ξ с функцией распределения $F(x,\theta)$, где θ – вектор параметров. По выборке $\mathbf{X}_n = X_1, X_2, \ldots, X_n$ требуется оценить неизвестные параметры распределения.

2. Методы оценивания.

Метод максимального правдоподобия. Оценки максимального правдоподобия (ОМП) находятся в результате максимизации функции правдоподобия:

$$\hat{\theta}_{ОМП} = \arg\max_{\theta} \prod_{j=1}^{n} f(X_j, \theta).$$

Метод минимального расстояния. *MD*-оценки находятся в результате минимизации расстояния $\rho(F(x,\theta), F_n(x))$ между теоретической и эмпирической функциями распределения:

$$\hat{\theta}_{MD} = \arg\min_{\theta} \rho(F(x,\theta), F_n(x)),$$

где $F_n(x)$ – эмпирическая функция распределения. В качестве меры близости можно использовать следующие статистики:

а) Статистика Колмогорова:

$$D_n = \sup_x \left| F(x,\theta) - F_n(x) \right|$$

б) Статистика ω^2 Мизеса:

$$\omega_n^2 = \frac{1}{12n^2} + \frac{1}{n}\sum_{i=1}^{n}\left\{ F(X_{(i)},\theta) - \frac{2i-1}{2n} \right\}^2$$

в) Статистика Ω^2 Мизеса:

$$\Omega_n^2 = -1 - \frac{2}{n}\sum_{i=1}^{n}\left\{ \frac{2i-1}{2n}\ln F(X_{(i)},\theta) + \left(1 - \frac{2i-1}{2n}\right)\ln(1 - F(X_{(i)},\theta)) \right\}$$

Вычисление оценок по порядковым статистикам. Для нахождения оценок часто используются линейные комбинации порядковых статистик или выборочных квантилей. Такие оценки называются L*-оценками*. L-оценки обладают двумя важными для практического применения качествами: простотой вычислений и хорошими свойствами робастности.

При построении L-оценок по выборочным квантилям $z_1 < z_2 < .... < z_k$ рассматриваемого закона оценки находят в виде:

$$\hat{\theta} = \sum_{i=1}^{k} \alpha_i z(p_i), \;\; z(p) = (X_{([np])} + X_{([np]+1)})/2,$$

где $X_{(i)}$ – i-я порядковая статистика, α_i и p_i – набор коэффициентов и вероятностей, которыми определяется конкретная оценка, n – объем выборки.

3. Свойства оценок.

Несмещенность. Оценка $\hat{\theta}$ называется *несмещенной*, если $\mathrm{M}[\hat{\theta}(X_n)] = \theta$.

Состоятельность. Оценка $\hat{\theta}$ называется *состоятельной*, если $\hat{\theta}(X_n) \xrightarrow{p} \theta$, т.е. $\forall \varepsilon > 0 \;\; \lim\limits_{n\to\infty} P\{\left|\hat{\theta}(X_n) - \theta\right| > \varepsilon\} = 0$.

Эффективность. Несмещенная оценка $\hat{\theta}$ называется *эффективной*, если $\mathrm{D}[\hat{\theta}\;(X_n)] = J_n^{-1}(\theta)$, где $J(\theta) = M\left(\dfrac{\partial \ln L(\mathbf{X}_n,\theta)}{\partial\theta}\right)^2 = -M\dfrac{\partial^2 \ln L(\mathbf{X}_n,\theta)}{\partial\theta^2}$ – информационное количество Фишера.

Асимптотическая нормальность. Оценка $\hat{\theta}$ называется асимптотически нормальной, если ее распределение с ростом объема выборки стремится к нормальному закону распределения.

4. Исследование асимптотических свойств оценок методом Монте-Карло.

1. Моделируется N выборок по n наблюдений в каждой в соответствии с заданным законом распределения и фиксированными значениями параметров θ.
2. По каждой выборке вычисляются оценки параметров, в результате получается выборка $T_N = \{\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_N\}$.
3. Исследуется (идентифицируется) распределение случайной величины $\hat{\theta}(\mathbf{X}_n)$.
4. Пункты 1-3 повторяются при большем значении n.

По серии машинных экспериментов делаются выводы об изменении свойств оценок с ростом объема выборки.

Порядок выполнения

1. Определить количество экспериментов N, при котором с вероятностью 0.99 отклонение эмпирической функции распределения от теоретической не превышает величину 0.01.

2. Используя программу **ISW с**моделировать выборки, состоящие из оценок:

- максимального правдоподобия $-\hat{\theta}_{ОМП}(\mathbf{X}_n)$,

- минимального расстояния Колмогорова – $\hat{\theta}_K(\mathbf{X}_n)$,

- минимального расстояния ω^2 Мизеса $-\hat{\theta}_{\omega^2}(\mathbf{X}_n)$,

- минимального расстояния $\hat{\theta}_{\Omega^2}(\mathbf{X}_n)$ Мизеса,

- по порядковым статистикам (L-оценки)

при $n = 20, 50, 100, 1000$.

3. Построить графики эмпирических функций распределения смоделированных выборок, сравнить функции распределения оценок параметров при разных объемах и методах оценивания.

4. Исследовать полученные выборки на нормальность.

5. Заполнить таблицу (для каждого метода оценивания).

	n =20	n =50	n =100	n =1000
$\mathrm{M}[\hat{\theta}(\mathbf{X}_N)] - \theta$				
$P\{\lvert\hat{\theta}(\mathbf{X}_N) - \theta\rvert > 0.1\}$				
$\mathrm{D}[\hat{\theta}(\mathbf{X}_N)] - J_n^{-1}(\theta)$				

На основании таблицы сделать выводы о свойствах различных методов оценивания.

Варианты заданий

№ варианта	Распределение, параметр
1.	Экспоненциальное распределение, параметр масштаба.
2.	Распределение Лапласа, параметр масштаба.
3.	Нормальное распределение, параметр сдвига.
4.	Нормальное распределение, параметр масштаба.
5.	Распределение Коши, параметр сдвига.
6.	Распределение Коши, параметр масштаба.
7.	Логистическое распределение, параметр сдвига.
8.	Логистическое распределение, параметр масштаба.
9.	Распределение экстремальных значений (минимум), параметр сдвига.
10.	Распределение экстремальных значений (минимум), параметр масштаба.
11.	Распределение экстремальных значений (максимум), параметр сдвига.
12.	Распределение экстремальных значений (максимум), параметр

	масштаба.

Контрольные вопросы

1. Метод максимального правдоподобия.
2. Методы минимального расстояния.
3. Оценивание параметров по порядковым статистикам.
4. Свойства оценок по выборкам фиксированного объема.
5. Асимптотические свойства оценок.
6. Экспериментальное исследование свойств оценок.

[Назад][Вперед] [Содержание]

Лабораторная работа № 3. Экспериментальное исследование робастности оценок

Цель работы. Исследование устойчивости оценок на наличие в выборке аномальных наблюдений. Исследование эффективности параметрической процедуры исключения аномальных наблюдений при использовании робастных оценок. Построение функций влияния Хампеля для ОМП. Исследование распределений статистик типа Граббса, предназначенных для анализа на аномальность сразу нескольких наблюдений.

Методические указания

1. Модель засорения выборки [3]. Пусть в эксперименте наблюдается непрерывная случайная величина ξ с функцией распределения

$$F_\xi(x) = (1-\nu)F(x,\theta) + \nu F_1(x,\theta_1), \qquad (3.1)$$

где ν – доля засорения выборки аномальными (с точки зрения закона $F(x,\theta)$) наблюдениями, подчиняющимися закону $F_1(x,\theta_1)$. По выборке $X_n = X_1, X_2, \ldots, X_n$ требуется оценить неизвестные параметры закона распределения $F(x,\theta)$.

2. Робастность. Под *робастностью* в статистике понимают нечувствительность к малым отклонениям от предположений (т.е. когда $\nu \to 0$).

3. Повышение робастности с помощью процедуры группирования. При группировании выборки теряется информация об индивидуальных наблюдениях, а фиксируется только количество наблюдений, попавших в интервалы группирования. В результате, небольшие отклонения от предполагаемого закона и аномальные выбросы не оказывают существенного влияния на оценки.

4. Параметрическая процедура отбраковки аномальных наблюдений состоит из двух этапов.

I этап. Одним из *робастных* методов находим оценки $\hat{\theta}$ параметров распределения $F(x,\hat{\theta})$.

II этап. Отбрасываем все наблюдения X_i: $X_i < \underline{d} \vee X_i > \overline{d}$. Пороговые значения определяются по формулам:

$$\underline{d} = F^{-1}(1-\sqrt[n]{1-\alpha},\hat{\theta}), \quad \overline{d} = F^{-1}(\sqrt[n]{1-\alpha},\hat{\theta}),$$

где n – объем выборки, α – уровень значимости критерия (вероятность ошибки первого рода, т.е. вероятность признать неаномальное наблюдение аномальным).

5. Функция влияния Хампеля. Для анализа робастности может использоваться функция влияния Хампеля:

$$IF(x;F,T) = \lim_{s\to 0}\frac{T((1-s)F + s\delta_x) - T(F)}{s},$$

где δ_x – единичная масса в точке x, $T(F)$ – статистика.

Если функция влияния неограничена, то резко выделяющиеся наблюдения могут приводить к существенным изменениям оценок или статистик. Чувствительность к большой ошибке характеризуется величиной:

$$\gamma^* = \sup_x \left| IF(x,F,T) \right|.$$

<u>Функция влияния для асимптотически эффективных оценок</u>, к которым относятся и ОМП, имеет вид:

$$IF(x,F,T) = J^{-1}(F(x,\theta))\frac{\partial \ln f(x,\theta)}{\partial\theta},$$

где $J^{-1}(F(x,\theta)) = \int_{-\infty}^{\infty}\left(\frac{\partial \ln f(x,\theta)}{\partial\theta}\right)^2 f(x,\theta)dx$ – информационное количество Фишера.

В случае группированных наблюдений для k-го интервала группирования x_{k-1}, x_k функция влияния постоянна и равна

$$IF(x,F,T) = J^{-1}(F(x,\theta))\frac{\partial \ln P_k(\theta)}{\partial\theta},$$

где $P_k(\theta) = F(x_k,\theta) - F(x_{k-1},\theta)$.

6. Критерии типа Граббса. В случае принадлежности выборок нормальному закону для отбраковки аномальных наблюдений могут использоваться критерии типа Граббса. Наблюдаемая выборка упорядочивается (строится вариационный ряд) $X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$.

Проверка на аномальность одного значения. При проверке на выброс $X_{(n)}$ статистика критерия Граббса имеет вид

$$G_n = (X_{(n)} - \overline{X})/S, \tag{3.2}$$

где $\overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j, S^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X})^2$.

При проверке на выброс наименьшего выборочного значения конкурирующая гипотеза H_1 предполагает, что $X_{(1)}$ принадлежит некоторому

другому закону, существенно сдвинутому влево. В данном случае вычисляемая статистика принимает вид

$$G_1 = (\overline{X} - X_{(1)}) / S. \qquad (3.3)$$

Максимальный или минимальный элемент выборки считается выбросом, если значение соответствующей статистики превысит критическое: $G_n \geq G_{n,1-\alpha}$ или $G_1 \geq G_{1,1-\alpha}$, где α – задаваемый уровень значимости.

Проверка на два выброса. В этом случае конкурирующая гипотеза H_1 может быть связана с предположением, что, например, некоторому другому закону принадлежат $X_{(n-1)}$ и $X_{(n)}$ (либо $X_{(1)}$ и $X_{(2)}$). При проверке на выброс одновременно двух наибольших значений статистика критерия Граббса имеет вид

$$G = S^2_{n-1,n} / S^2_0, \qquad (3.4)$$

где

$$S^2_0 = \sum_{j=1}^{n} (X_j - \overline{X})^2, \quad S^2_{n-1,n} = \sum_{j=1}^{n-2} (X_j - \overline{X}_{n-1,n})^2, \quad \overline{X}_{n-1,n} = \frac{1}{n-2} \sum_{j=1}^{n-2} X_j .$$

Для проверки на выброс одновременно двух наименьших величин $X_{(1)}$ и $X_{(2)}$ статистика критерия принимает вид

$$G = S^2_{1,2} / S^2_0, \qquad (3.5)$$

где

$$S^2_{1,2} = \sum_{j=3}^{n} (X_j - \overline{X}_{1,2})^2, \quad \overline{X}_{1,2} = \frac{1}{n-2} \sum_{j=3}^{n} X_j .$$

Оба значения ($X_{(n-1)}$, $X_{(n)}$ или $X_{(1)}$, $X_{(2)}$) считаются выбросами, если значение соответствующей статистики окажется ниже критического: $G < G_\alpha$.

Очевидно, как можно сформировать статистику критерия для проверки на наличие любой комбинации аномальных измерений.

Дополнительная информация размещена по адресу:
http://www.ami.nstu.ru/~headrd/seminar/publik_html/Z_lab_1.htm
http://www.ami.nstu.ru/~headrd/seminar/publik_html/NADEG_1.htm
http://www.ami.nstu.ru/~headrd/seminar/publik_html/Sibgim_1.htm
http://www.ami.nstu.ru/~headrd/seminar/publik_html/Z_lab_10.htm
http://www.ami.nstu.ru/~headrd/seminar/publik_html/Izm_T_6.htm

Порядок выполнения. Используя программу **ISW:**

1. Исследовать робастность оценок максимального правдоподобия (ОМП); ОМП по группированным данным; MD-оценок, минимизирующих расстояния, задаваемые статистиками Колмогорова, ω^2 и Ω^2 Мизеса; L-оптимальных оценок по выборочным квантилям. Для этого:
   - моделируются выборки с засорением (3.1), где параметр $\nu = 0, 0.1, 0.2$;
   - оцениваются параметры распределения $F(x,\theta)$ всеми методами;
   - сравниваются значения оценок при разных значениях ν.

2. Отбраковка аномальных наблюдений. Используя параметрическую процедуру отбраковки аномальных наблюдений очистить полученные в п.1 выборки от аномальных наблюдений и проверить, как изменились результаты применения критериев для проверки согласия эмпирического распределения с распределением $F(x,\theta)$ после удаления части наблюдений.
3. Построить функции влияния Хампеля для ОМП и ОМП по группированным данным.
4. Исследовать распределения статистик критериев типа Граббса, предназначенных для анализа на аномальность сразу нескольких наблюдений (*по заданию преподавателя*) в предположении о принадлежности выборки нормальному закону. Построить эмпирические распределения для статистик критериев типа Граббса, найти приближенные значения процентных точек. Для вариантов 1-5 применить критерий Граббса для отбраковки аномальных наблюдений по выборкам с засорением, полученным в п.1.

Варианты заданий.

№	$F(x,\theta)$	$F_1(x,\theta_1)$
1	Нормальное	Коши
2	Нормальное	Нормальное с масштабом 5
3	Нормальное	Нормальное со сдвигом 5
4	Нормальное	Нормальное с масштабом 10
5	Нормальное	Нормальное со сдвигом 10
6	Логистическое	Логистическое с масштабом 5
7	Логистическое	Логистическое со сдвигом 5
8	Логистическое	Логистическое с масштабом 10
9	Логистическое	Логистическое со сдвигом 10
10	Экспоненциальное	Экспоненциальное с масштабом 3
11	Экспоненциальное	Экспоненциальное с масштабом 5
12	Экспоненциальное	Экспоненциальное с масштабом 10

Контрольные вопросы.

1. Какие статистические процедуры называются робастными?
2. Как исследовать робастность процедуры оценивания параметров?
3. Как вычислить функцию влияния Хампеля?

4. Что представляет собой функция влияния Хампеля по группированным наблюдениям.
5. Как влияют объём выборки и уровень значимости α на параметрическую процедуру отбраковки аномальных наблюдений?
6. Можно ли применять критерии типа Граббса в случае нарушения предположений о нормальности исходной выборки? Если нет, то почему? Если да, то, в каком случае?

[Назад][Вперед] [Содержание]

Лабораторная работа № 4. Непараметрическое оценивание плотности распределения вероятностей

Цель работы. Восстановление плотности распределения с помощью ядерных функций (оценок Розенблатта-Парзена). Экспериментальный подбор параметра размытости.

Методические указания

1. Модель. Пусть в эксперименте наблюдается непрерывная случайная величина ξ с плотностью распределения вероятностей $f(x)$. По выборке $\mathbf{X}_n = X_1, X_2, \ldots, X_n$ требуется найти оценку функции плотности $\tilde{f}_n(x)$.

2. Ядерные оценки. В качестве оценки $\tilde{f}_n(x)$ в непараметрической статистике используется смесь из n ядерных функций:

$$\tilde{f}_n(x) = \frac{1}{n\lambda_n}\sum_{i=1}^{n} g\left(\frac{x - X_i}{\lambda_n}\right), \tag{4.1}$$

где n – объем выборки, $g(x)$ – "ядро" – функция, удовлетворяющая условиям регулярности: $g(x) = g(-x)$, $0 \le g(x) < \infty$, $\int_{-\infty}^{\infty} g(x)dx = 1$, $\int_{-\infty}^{\infty} x^2 g(x)dx = 1$, $\int_{-\infty}^{\infty} x^m g(x)dx < \infty$, а параметр масштаба ("размытости") такой, что $\lim_{n\to\infty} \lambda_n = 0$, $\lim_{n\to\infty} n\lambda_n = \infty$.

3. Определение параметра размытости. Среднеквадратическая ошибка аппроксимации (4.2) равна:

$$M\left[\int f(x) - \tilde{f}_n(x)\ ^2 dx\right] \approx \frac{1}{n\lambda_n}\int_{-\infty}^{\infty} g(x)\ ^2 dx + \frac{\lambda_n^4}{4}\int_{-\infty}^{\infty} f''(x)\ ^2 dx. \tag{4.2}$$

Из минимизации (4.2) по λ_n можно найти оптимальное значение параметра размытости

$$\lambda_n^* = \left[\frac{\int_{-\infty}^{\infty} g(x)^2 dx}{n \int_{-\infty}^{\infty} f''(x)^2 dx} \right]^{1/5}. \qquad (4.3)$$

и квадратичную ядерную функцию, обладающую наилучшими свойствами,

$$g^*(x) = \begin{cases} \frac{3}{4\sqrt{5}} - \frac{3}{20\sqrt{5}} x^2, \forall |x| < \sqrt{5}, \\ 0, \ \forall |x| < \sqrt{5}. \end{cases} \qquad (4.4)$$

Однако в (4.3) присутствует вторая производная функции плотности, поэтому вычислить оптимальное значение параметра размытости можно лишь при априорно известной функции плотности распределения.

Оценку параметра λ_n можно найти также <u>по методу максимального правдоподобия</u>:

$$\hat{\lambda} = \arg\min_{\lambda} \prod_{j=1}^{n} \frac{1}{(n-1)\lambda} \sum_{\substack{i=1 \\ i \neq j}}^{n} g\left(\frac{X_j - X_i}{\lambda} \right). \qquad (4.5)$$

Если случайная величина ξ определена на интервале $0, \infty$, тогда при построении непараметрических оценок можно перейти к случайной величине $\eta = \ln \xi$, определенной на всем множестве R. Если случайная величина ξ определена на интервале a, b, тог можно перейти к случайной величине $\eta = \ln \frac{x-a}{b-x}$, также определенной на всем множестве R.

Порядок выполнения

В процессе выполнения использовать программу **ISW.**

1. Сгенерировать выборку объема n.

2. Построить ядерную оценку плотности при разных значениях параметра размытости.

а) Оптимальное значении параметра размытости.

б) Оценка максимального правдоподобия параметра размытости.

в) Значение параметра размытости *меньше* оптимального.

г) Значение параметра размытости *больше* оптимального.

д) $\lambda = n^{-1/5}$.

Опишите вид графика ядерной оценки, сравните с плотностью истинного закона распределения случайной величины.

3. Повторить пункты 1-2 при *n = 20, 50, 100, 1000*. Сделать выводы об изменении ядерной оценки и параметра размытости с ростом объема выборки.

Варианты заданий.

1. Нормальное распределение.

2. Распределение Лапласа.
3. Распределение Коши.
4. Логистическое распределение.
5. Распределение экстремальных значений (минимум).
6. Распределение экстремальных значений (максимум) .
7. Экспоненциальное распределение.
8. Гамма-распределение.
9. Распределение Вейбулла-Гнеденко
10. Логнормальное распределение
11. Равномерное распределение.
12. Бета-распределение I-го рода.

Контрольные вопросы.

1. Если считать ядерную функцию плотностью некоторой случайной величины, то какими свойствами она обладает?
2. Является ли функция плотности нормального закона ядерной функцией? Ответ обоснуйте.
3. Является ли функция плотности логистического закона ядерной функцией?
4. Является ли функция плотности распределения Коши ядерной функцией?
5. Является ли функция плотности распределения Лапласа ядерной функцией?
6. Можно ли с помощью ядерных функций аппроксимировать функцию распределения?
7. Используя методы *вариационного исчисления*, найти оптимальное ядро (4.4), минимизируя (4.2).
8. Найти оптимальный параметр размытости, минимизируя (4.2).

[Назад][Вперед] [Содержание]

Лабораторная работа № 5. Исследование оценок максимального правдоподобия параметров распределений вероятностей по цензурированным данным

Цель работы. Вычисление оценок максимального правдоподобия параметров распределений вероятностей по цензурированным данным. Исследование асимптотических свойств оценок методом Монте-Карло.

Методические указания

1. В исследованиях на надежность используется следующая схема эксперимента. Имеется партия из n изделий. В эксперименте фиксируются моменты выхода изделий из строя. Очевидно, что если проводить испытания бесконечно долго, то все изделия рано или поздно выйдут из строя. Однако реально время проведения эксперимента ограничено интервалом $[0,t]$, где t – момент прекращения испытаний.

Для определения параметров надежности по незавершенным испытаниям характерна ситуация, когда к моменту прекращения испытаний большой

партии изделий наблюдается выход из строя лишь части из них, обычно достаточно малой по сравнению с объемом всей партии. В этом случае анализ приходится проводить по выборке, сильно цензурированной справа. Особенно часто приходится сталкиваться с задачей обработки цензурированных выборок, когда наблюдению оказывается доступной только часть области определения случайной величины, а для выборочных значений, попавших левее или правее этой области, фиксируется лишь сам факт этого попадания.

Очевидно, что в такой неполной (цензурированной) выборке содержится меньше информации, чем в полной и это, естественно, отражается на точности оценивания параметров аппроксимирующего закона распределения. В этой связи оказывается интересным, насколько точно можно оценить параметры наблюдаемого закона в зависимости от объема всей выборки n (объема партии) и величины наблюдаемой ее части.

2. Наиболее универсальным методом по отношению к форме представления выборочных данных является метод максимального правдоподобия. В отличие от других метод позволяет находить оценки максимального правдоподобия (ОМП) параметров по негруппированным, частично группированным и группированным данным.

ОМП неизвестного параметра по частично группированным наблюдениям называется такое значение параметра, при котором функция правдоподобия

$$L(\theta) = \gamma \prod_{(1)} P_i^{n_i}(\theta) \prod_{(2)} \prod_{j=1}^{n_i} f(x_{ij}, \theta), \tag{5.1}$$

где γ - некоторая константа; $f(x,\theta)$ - функция плотности случайной величины; $P_i(\theta) = \int\limits_{x_{(i-1)}}^{x_{(i)}} f(x,\theta)dx$ - вероятность попадания наблюдения в $\boldsymbol{i}$-й интервал значений, достигает максимума на множестве возможных значений параметра. При вычислении ОМП максимизируют (1) или решают систему уравнений правдоподобия

$$\sum_{(1)} n_i \frac{\partial \ln P_i(\theta)}{\partial \theta_l} + \sum_{(2)} \sum_{j=1}^{n_i} \frac{\partial \ln f(x_{ij}, \theta)}{\partial \theta_l} = 0, \; l = \overline{1,m}, \tag{5.2}$$

где m - размерность вектора параметров $\theta = (\theta_1, \theta_2, \ldots, \theta_m)^{\mathrm{T}}$.

В случае группированных или частично группированных данных система (5.2), за редким исключением, решается только численно.

При выборке, цензурированной с двух сторон, являющейся частным случаем частично группированной выборки, область определения случайной величины разбита на 3 интервала граничными точками $x_{(1)} < x_{(2)}$ так, что значения левее $x_{(1)}$ и правее $x_{(2)}$ не наблюдаются. И система (5.2) принимает вид

$$n_1 \frac{\partial \ln P_1(\theta)}{\partial \theta_l} + \sum_{j=1}^{n_2} \frac{\partial \ln f(x_{2j},\theta)}{\partial \theta_l} + n_3 \frac{\partial \ln P_3(\theta)}{\partial \theta_l} = 0,\ l = \overline{1,m}. \qquad (5.3)$$

Если оценивается скалярный параметр, то асимптотическая дисперсия его ОМП определяется соотношением

$$D(\hat{\theta}) = n^{-1} J_c^{-1}(\hat{\theta}), \qquad (5.4)$$

где информационное количество Фишера определяется соотношением

$$J_c(\theta) = \frac{1}{P_1(\theta)} \left[\frac{\partial P_1(\theta)}{\partial \theta} \right]^2 + \int_{x_{(1)}}^{x_{(2)}} \left[\frac{\partial \ln f(x,\theta)}{\partial \theta} \right]^2 f(x,\theta) dx + \frac{1}{P_3(\theta)} \left[\frac{\partial P_3(\theta)}{\partial \theta} \right]^2. \qquad (5.5)$$

Если выборка цензурирована только справа, то в выражении исчезает левое слагаемое, только слева - правое слагаемое. Это соотношение позволяет судить о потерях информации о параметре распределения в зависимости от степени цензурирования слева или справа.

Об эффективности оценивания параметров по цензурированной выборке по отношению к оцениванию по полной выборке можно судить по величине $J_c(\theta)/J(\theta)$, где $J(\theta)$ – количество информации Фишера в полной выборке.

Порядок выполнения.

В процессе выполнения использовать программу **ISW.**

1. Оценить количество экспериментов N, при котором с вероятностью 0.99 отклонение $D[\hat{\theta}_n]/D[\hat{\theta}_n^c]$ по эмпирической функции распределения от $D[\hat{\theta}_n]/D[\hat{\theta}_n^c]$ по теоретической не превышает величину 0.01.
2. Смоделировать выборки состоящие из оценок максимального правдоподобия – $\hat{\theta}_{ОМП}(X_n)$ при $n = 100, 200, 300$. Уровень цензурирования (вероятность попадания в область цензурирования слева или справа) выбирается равным 0.1, 0.5, 0.9 (соответственно, вероятность попадания в наблюдаемую область 0.9, 0.5, 0.1). Повторив эксперимент N раз, оценить дисперсии оценок и величину $D[\hat{\theta}_n]/D[\hat{\theta}_n^c]$, сравнить ее с величиной асимптотической эффективности $J_c(\theta)/J(\theta)$
3. Сравнить графики эмпирических функций распределения смоделированных выборок при разных объемах выборок.
4. Исследовать полученные выборки на нормальность.
5. Заполнить таблицу (для каждого α)

	n =100	*n =200*	*n =300*
$\mathrm{M}[\hat{\theta}(X_n)] - \theta$			

$P\{\left\|\hat{\theta}(X_n)-\theta\right\|>0{,}1\}$			
$\mathrm{D}[\hat{\theta}(\mathbf{X}_N)]-J_n^{-1}(\theta)$			
$J_C(\theta)/J(\theta)$			

На основании таблицы сделать выводы о свойствах ОМП по цензурированным данным.

Варианты заданий совпадают с заданиями из лабораторной работы № 2.

Контрольные вопросы

1. Какие бывают типы цензурирования? Привести примеры цензурированных выборок.
2. Оценивание параметров по цензурированным выборкам.
3. Вычислить потери информации при 10% цензурировании выборки, распределенной по нормальному закону.
4. Вычислить потери информации при 10% цензурировании слева выборки, распределенной по экспоненциальному закону.
5. Вычислить потери информации при 10% цензурировании справа выборки, распределенной по экспоненциальному закону.

[Назад][Вперед]

Лабораторная работа № 6. Экспериментальное исследование свойств критерия согласия χ^2 Пирсона.

Цель работы. Исследование влияния способов группирования на предельные распределения статистики критерия согласия χ^2 Пирсона при простых и сложных гипотезах (при использовании для вычисления оценок по негруппированным данным метода максимального правдоподобия).

Методические указания

В практике статистического анализа с необходимостью использования критериев согласия приходится сталкиваться при проверке простой гипотезы H_0: $f(x)=f(x,\theta_u)$, где $f(\cdot)$ - плотность распределения наблюдаемого закона, θ_u - известное истинное значение параметра (вектора параметров) закона, или при проверке сложной гипотезы H_0: $f(x)\in\{f(x)=f(x,\theta),\theta\in\Theta\}$, когда оценка $\hat{\theta}$ параметра предполагаемого закона распределения оценивается по этой же выборке.

Статистика критерия χ^2 Пирсона вычисляется в соответствии с соотношением

$$X_n^2 = n\sum_{i=1}^{k}\frac{(n_i/n - P_i(\theta))^2}{P_i(\theta)}, \qquad (6.1)$$

где n_i – количество наблюдений, попавших в интервал, $n = \sum_{i=1}^{k} n_i$ $P_i(\theta)$ - вероятность попадания наблюдения в i-й интервал. При справедливости (простой) гипотезы H_0 предельным распределением $G(S|H_0)$ статистики является χ_r^2-распределение с числом степеней свободы $r = k - 1$. Если по выборке оценивалось p параметров закона в результате минимизации статистики X_n^2, статистика подчиняется χ_r^2-распределению с $r = k - p - 1$ степеней свободы. При справедливости конкурирующей гипотезы H_1 предельное распределение $G(S|H_1)$ представляет собой нецентральное χ_r^2-распределение с тем же числом степеней свободы и параметром нецентральности

$$\lambda = n\sum_{i=1}^{k}\frac{(P_i(\theta_1) - P_i(\theta))^2}{P_i(\theta)},$$

где $P_i(\theta_1)$ - вероятности попадания наблюдения в i-й интервал при конкурирующей гипотезе.

В случае проверки **сложных** гипотез и оценивании по выборке параметров распределений использование в качестве предельных χ_{k-p-1}^2-распределений справедливо лишь при определении оценок параметров по сгруппированным данным и использовании для оценивания статистики X_n^2:

$$\hat{\theta} = \arg\min_{\theta} X_n^2.$$

Это же справедливо в случае применения для сгруппированных данных метода максимального правдоподобия, оптимальных L-оценок по выборочным квантилям и некоторых других методов.

При использовании критериев согласия конкурирующая гипотеза H_1 (альтернатива) обычно не задается. Задавая конкретную альтернативу и имея возможность построить распределения статистик при истинности гипотезы H_0 ($G(S|H_0)$) и истинности гипотезы H_1 ($G(S|H_1)$), можно при заданном уровне значимости α (α - вероятность ошибки первого рода) вычислить мощность критерия $1-\beta$, которая определяет способность различения этих гипотез (β - вероятность ошибки второго рода).

Порядок выполнения

Исследовать распределения статистики критерия для простых и сложных гипотез (при использовании оценок максимального правдоподобия по негруппированным данным) при справедливой нулевой и конкурирующей гипотезах.

1. В соответствии с заданным наблюдаемым распределением (гипотеза H_0) смоделировать эмпирические распределения статистики критерия при простой гипотезе (а) (по выборке не оцениваются параметры), для сложной гипотезы (б) (по выборке оцениваются все параметры).
2. Идентифицировать построенные законы распределения (найти аналитические модели, наиболее хорошо описывающие эмпирические распределения).
3. Повторить п.1, моделируя выборку по закону, соответствующему гипотезе H_1, а оценивая по этой выборке параметры закона, соответствующего гипотезе H_0. *Для того, чтобы гипотеза H_1 была наиболее близка к гипотезе H_0, следует подобрать параметры распределения, соответствующего гипотезе H_1, из условия минимизации расстояния до распределения, соответствующего гипотезе H_1.*

Для моделирования распределения статистики при справедливой гипотезе H_0 ($G(S|H_0)$) следует генерировать псевдослучайные величины, соответствующие наблюдаемому закону, и оценивать его параметры (в случае сложной гипотезы). Для моделирования распределения этой же статистики при проверке той же самой гипотезы H_0, но при справедливой гипотезе H_1 ($G(S|H_1)$), следует генерировать псевдослучайные величины по закону, соответствующему гипотезе H_1, а оценивать параметры закона, соответствующего гипотезе H_0.

4. В результате такого моделирования будут получены пары эмпирических распределений $G(S|H_0)$ и $G(S|H_1)$ для простой и сложной гипотез, по которым, задаваясь значением α, можно вычислить мощность критерия $1-\beta$ относительно конкурирующей гипотезы:

$$\int_{-\infty}^{S_\alpha} g(s|H_0)ds = 1-\alpha, \quad \int_{S_\alpha}^{\infty} g(s|H_1)ds = 1-\beta. \qquad (6.2)$$

Опираясь на эти распределения, построить оперативные характеристики критерия для проверки простой и сложной гипотез как функции вида $(1-\beta)(\alpha)$.

5. Повторяя пункты 3-4, исследовать влияние количества интервалов, способа группирования (равновероятное и асимптотически оптимальное) на мощность критерия.

Варианты заданий

№ п/п	Гипотеза H_0	Гипотеза H_1
1.	Нормальное	Логистическое
2.	Логистическое	Нормальное
3.	Нормальное	Лапласа
4.	Лапласа	Нормальное
5.	Логистическое	Лапласа
6.	Лапласа	Логистическое
7.	Гамма	Вейбулла-Гнеденко
8.	Вейбулла-Гнеденко	Гамма
9.	Экстремальных значений (минимум)	Экстремальных значений (максимум)
10.	Экстремальных значений (максимум)	Экстремальных значений (минимум)
11.	Нормальное	Коши
12.	Коши	Нормальное

Контрольные вопросы

1. Распределена статистика критерия χ^2 Пирсона при справедливости проверяемой гипотезы? При справедливости конкурирующей гипотезы?
2. Почему при асимптотически оптимальном группировании мощность критерия χ^2 Пирсона максимальна относительно близких конкурирующих гипотез?
3. Как идентифицировать закон распределения по заданной выборке?

[Назад][Вперед] [Содержание]

Лабораторная работа № 7. Исследование распределения статистики и мощности критерия Рао-Робсона-Никулина

Цель работы: Исследовать влияние способов группирования на предельные распределения статистики критерия согласия типа χ^2 Никулина при простых и сложных гипотезах (при использовании для вычисления оценок метода максимального правдоподобия по негруппированным данным). Сравнение мощности критерия Никулина с мощностью критерия χ^2 Пирсона. Исследовать мощности критерия Никулина в зависимости от числа интервалов.

Методические указания

Никулиным М.С. [6] предложена модификация стандартной статистики (6.1), для которой предельным распределением является обычное распределение χ^2_{k-1} (количество степеней свободы не зависит от числа

оцениваемых параметров). Неизвестные параметры распределения $F(x,\theta)$ в этом случае должны оцениваться по негруппированным данным методом максимального правдоподобия. При этом вектор вероятностей попадания в интервалы $p=(p_1,\ldots,p_k)^{\mathrm{T}}$ предполагается заданным и граничные точки интервалов определяются соотношениями $x_i(\theta)=F^{-1}(p_1+\ldots+p_i)$, $i=\overline{1,(k-1)}$.

Предложенная статистика отличается от X_n^2 только при проверке сложных гипотез и имеет вид

$$Y_n^2(\theta)=X_n^2+n^{-1}a^{\mathrm{T}}(\theta)\Lambda(\theta)a(\theta),$$

где X_n^2 вычисляется в соответствии с (1). Элементы и размерность матрицы

$$\Lambda(\theta)=\left[J(\theta_l,\theta_j)-\sum_{i=1}^{k}\frac{w_{\theta_l i}w_{\theta_j i}}{p_i}\right]_{m\times m}^{-1}$$

определяются оцениваемыми компонентами вектора параметров θ, $J(\theta_l,\theta_j)$ - элементы информационной матрицы $\mathbf{J}(\theta)$, $a(\theta_l)=w_{\theta_l 1}n_1/p_1+\ldots+w_{\theta_l k}n_k/p_k$ - элементы вектора $a(\theta)$, величины $w_{\theta_l i}$ определяются соотношением

$$w_{\theta_l i}=-f[x_i(\theta),\theta]\frac{\partial x_i(\theta)}{\partial\theta_l}+f[x_{i-1}(\theta),\theta]\frac{\partial x_{i-1}(\theta)}{\partial\theta_l}.$$

Порядок выполнения.

1. Проверить, насколько хорошо распределение статистики Y_n^2 при справедливой нулевой гипотезе $G(Y_n^2|H_0)$ соответствует распределению χ^2_{k-1}
2. Исследовать влияние типа группирования на распределение статистики Никулина.
3. Построить оперативные характеристики критерия для простой и сложных гипотез как функции вида $(1-\beta)(\alpha)$. Сравнить мощность предложенного Никулиным критерия с мощностью критерия χ^2 Пирсона (результаты по мощности критерия χ^2 Пирсона взять из отчета по лабораторной работе № 6). *Для того чтобы распределение, соответствующее гипотезе H_1 было наиболее близким к распределению, соответствующему гипотезе H_0, следует подобрать параметры распределения, соответствующего гипотезе H_1, из условия минимизации расстояния до распределения, соответствующего основной гипотезе.*
4. Исследовать мощность критерия Никулина в зависимости от числа интервалов и выполнить сравнение с аналогичными исследованиями по критерию χ^2 Пирсона.

Варианты заданий совпадают с заданиями из лабораторной работы № 6

Контрольные вопросы

1. Сравнить критерии χ^2 Пирсона и Рао-Робсона-Никулина. Когда нельзя применять критерий Рао-Робсона-Никулина?
2. Вычислить поправку Никулина для экспоненциального распределения с оцениванием параметра масштаба.
3. Вычислить поправку Никулина для нормального распределения с с оцениванием параметра сдвига.
4. Вычислить поправку Никулина для нормального распределения с оцениванием параметра масштаба.

[Назад][Вперед] [Содержание]

Лабораторная работа № 8. Экспериментальное исследование предельных распределений статистик непараметрических критериев согласия.

Цель работы. Исследование распределений статистик непараметрических критериев согласия при проверке простых и различных сложных гипотез.

Методические указания

В практике статистического анализа с необходимостью использования критериев согласия приходится сталкиваться как при проверке простой гипотезы H_0: $F(x) = F(x, \theta_è)$, где $F(\cdot)$ - плотность распределения наблюдаемого закона, θ_u - известное истинное значение параметра (вектора параметров) закона, так и при проверке сложной гипотезы H_0: $F(x) \in \{F(x, \theta), \theta \in \Theta\}$. Проблемы имеются, когда оценка $\hat{\theta}$ параметра предполагаемого закона распределения вычисляется по той же самой выборке, по которой проверяется согласие. Если оценка $\hat{\theta}$ вычисляется по другой выборке, то ситуация не отличается от проверки простой гипотезы. В дальнейшем, если оценка параметра $\hat{\theta}$ вычисляется по этой же выборке, будем обозначать сложную гипотезу H_0: $F(x) = F(x, \hat{\theta})$.

В случае проверки сложных гипотез предельные распределения статистик непараметрических критериев согласия типа Колмогорова, Смирнова, ω^2 и Ω^2 Мизеса, при справедливости нулевой гипотезы H_0: $F(x) = F(x, \theta)$ отличаются от предельных распределений классических статистик (когда по выборке не оцениваются параметры) [7]. В случае сложной гипотезы предельные распределения статистик зависят: от вида наблюдаемого закона, от количества и типа оцениваемых параметров этого распределения, от используемого метода оценивания параметров. А при ограниченных объемах выборок распределения статистик существенно зависят и от объема выборки.

Знание распределения статистики при проверке одной и той же гипотезы, но при различных истинных гипотезах (H_0 или H_1) позволяет определить мощность критерия, т.е. его способность различать эти гипотезы. Задавая конкретную альтернативу и имея возможность построить распределения статистик при истинности нулевой гипотезы H_0 ($G(S|H_0)$) и истинности альтернативы H_1 ($G(S|H_1)$), можно при заданном уровне значимости α вычислить мощность критерия $1-\beta$, которая определяет способность различения этих гипотез.

Для построения распределения $G(S|H_0)$ (распределения статистики при справедливой гипотезе H_0) следует моделировать псевдослучайные величины, соответствующие наблюдаемому закону и оценивать его параметры, после чего вычислять значение требуемой статистики S. А для построения распределения $G(S|H_1)$ (распределения той же статистики при проверке той же самой сложной гипотезы H_0, но при справедливой гипотезе H_1) следует моделировать псевдослучайные величины по закону, соответствующему гипотезе H_1, а оценивать параметры закона, соответствующего гипотезе H_0.

Порядок выполнения

1. Смоделировать распределение статистики S для заданного критерия согласия при простой гипотезе H_0. Сравнить полученное эмпирическое распределение с предельным распределением классической статистики.
2. Смоделировать распределение статистики S для этого же критерия согласия при сложной гипотезе H_0. Попытаться идентифицировать полученное эмпирическое распределение, используя систему статистического анализа **ISW**.
3. Смоделировать распределения статистики S исследуемого критерия согласия при проверке простой и сложной гипотез H_0 при справедливой гипотезе H_1. *Для того чтобы распределение, соответствующее гипотезе H_1 было наиболее близким к распределению, соответствующему гипотезе H_0, следует подобрать параметры распределения, соответствующего гипотезе H_1, из условия минимизации расстояния до распределения, соответствующего основной гипотезе.*
4. Построить оперативные характеристики критерия для простой и сложных гипотез как функции вида $(1-\beta)(\alpha)$. Сравнить мощность непараметрических критериев с мощностью критериев χ^2 Пирсона и Рао-Робсона-Никулина (оперативные характеристики критериев χ^2 Пирсона и Рао-Робсона-Никулина взять из отчетов по лабораторным работам № 6-7)

Варианты заданий совпадают с заданиями из лабораторной работы № 6.

Контрольные вопросы

1. Критерий Колмогорова при проверке простых и сложных гипотез?
2. Критерий Крамера-Мизеса-Смирнова при проверке простых и сложных гипотез?
3. Критерий Андерсона-Дарлинга при проверке простых и сложных гипотез?
4. Влияет ли метод оценивания параметров на распределения статистик непараметрических критериев согласия при проверке сложной гипотезы?
5. Влияют ли значения параметров на распределения статистик непараметрических критериев согласия при проверке сложной гипотезы?

[Назад][Вперед] [Содержание]

Лабораторная работа № 9. Исследование критериев проверки отклонения от нормального закона

Цель работы. Исследование распределений статистик критериев, используемых при проверке отклонения эмпирических распределений наблюдаемых величин от нормального закона (в том числе критериев проверки гипотез о симметричности и о значении эксцесса при различных наблюдаемых законах). Исследование распределений статистик критериев Шапиро-Уилка, Эппса-Палли, Д'Агостино. Исследование и сравнение мощности критериев относительно близких конкурирующих гипотез.

Методические указания

1. **Критерий симметричности** предназначен для проверки гипотез о симметричности наблюдаемого закона (против наличия асимметрии) при объемах выборки $8 \le n \le 5000$. Статистика критерия имеет вид

$$\sqrt{\hat{\beta}_1} = \frac{\hat{\mu}_3}{\hat{\sigma}^3}. \qquad (9.1)$$

Проверяется гипотеза H_0: $\sqrt{\beta_1}=0$ против альтернативы $\sqrt{\beta_1}>0$ (положительная асимметрия) или $\sqrt{\beta_1}<0$ (отрицательная асимметрия).

2. **Критерий проверки на эксцесс** рассматривается при объемах выборок $8 \le n \le 5000$. Статистика критерия проверки на значение эксцесса имеет вид

$$\hat{\beta}_2 = \frac{\hat{\mu}_4}{\hat{\sigma}^4}. \qquad (9.2)$$

Проверяется гипотеза вида H_0: $\beta_2=3$ против альтернативы $\beta_2>3$ (больший эксцесс) или $\beta_2<3$ (меньший эксцесс).

3. В **критерии Шапиро-Уилка** для вариационного ряда $X_{(1)} \le X_{(2)} \le ... \le X_{(n)}$, полученного по наблюдаемой выборке $X_1, X_2, ..., X_n$, вычисляют величину

$$S = \sum_k a_k [X_{(n+1-k)} - X_{(k)}],$$

где индекс k изменяется от 1 до $n/2$ или от 1 до $(n-1)/2$ при четном и нечетном n соответственно. Коэффициенты a_k приведены в ГОСТ Р ИСО 5479-2002 и первоисточниках. Статистика критерия имеет вид

$$W = S^2 \Big/ \sum_{i-1}^{n} (X_i - \bar{X})^2 . \qquad (9.3)$$

Гипотеза о нормальности отвергается при малых значениях статистики W.

4. Статистика **критерия Эппса-Палли** для наблюдаемой выборки $X_1, X_2, ..., X_n$ имеет вид

$$T_{EP} = 1 + \frac{n}{\sqrt{3}} + \frac{2}{n} \sum_{k=2}^{n} \sum_{j=1}^{k-1} \exp\left\{ -\frac{(X_j - X_k)^2}{2\hat{\hat{\mu}}_2} \right\} - \sqrt{2} \sum_{j=1}^{n} \exp\left\{ -\frac{(X_j - \bar{X})^2}{4\hat{\hat{\mu}}_2} \right\}, \qquad (9.4)$$

где $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$, $\hat{\mu}_2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$. Выборка может быть неупорядочена, порядок наблюдений произволен, но он должен быть неизменным в течение всех проводимых вычислений. Гипотезу о нормальности отвергают при больших значениях статистики.

5. **Модификация D'Agostino** критерия **проверки на симметричность**. В данной модификации на основании следующих соотношений статистика (9.1) преобразуется в статистику z_1, приближенно подчиняющуюся стандартному нормальному закону:

$$b = \frac{3(n^2 + 27n - 70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)}, \qquad \omega^2 = -1 + \{2(b-1)\}^{1/2},$$

$$\delta = \frac{1}{\{\log(\sqrt{\omega^2})\}^{1/2}}, \qquad y = \sqrt{\hat{\beta}_1} \left\{ \frac{\omega^2 - 1}{2} \cdot \frac{(n+1)(n+3)}{6(n-2)} \right\}^{1/2},$$

$$z_1 = \delta \log\{y + (y^2 + 1)^{1/2}\}. \qquad (9.5).$$

6. **Модификация D'Agostino** критерия одновременной **проверки на симметричность и значение эксцесса**. Здесь предложено преобразование статистик (9.2) и (9.1) к статистике z_2, приближенно распределенной в соответствии со стандартным нормальным законом. Преобразование осуществляется с помощью следующих соотношений:

$$\delta = (n-3)(n+1)(n^2+15n-4), \quad a = \frac{(n-2)(n+5)(n+7)(n^2+27n-70)}{6\delta}$$

$$c = \frac{(n-7)(n+5)(n+7)(n^2+2n-5)}{6\delta}, \quad k = \frac{(n+5)(n+7)(n^3+37n^2+11n-313)}{12\delta},$$

$$\alpha = a + \beta_1 c, \qquad \chi = (\hat{\beta}_2 - 1 - \hat{\beta}_1)2k,$$

$$z_2 = \left\{ \left(\frac{\chi}{2\alpha} \right)^{1/3} - 1 + \frac{1}{9\alpha} \right\} (9\alpha)^{1/2} . \qquad (9.6)$$

Дополнительная информация по критериям доступна по адресам:
http://www.ami.nstu.ru/~headrd/seminar/publik_html/Izm_T_7.htm
http://www.ami.nstu.ru/~headrd/seminar/Kontrol_Q

Порядок выполнения

Используя программу **ISW**, исследовать описанные выше критерии проверки отклонения наблюдаемых данных от нормального закона. В частности:

1. Исследовать зависимость распределений статистик (9.1) и (9.2) от объема выборок в случае принадлежности наблюдений нормальному закону. При некотором объеме выборок смоделировать эмпирические распределения статистик (9.1) и (9.2) при законе, отличающемся от нормального. Сравнить распределения статистик со случаем нормальности наблюдаемого закона.
2. Исследовать распределения статистик критерия Шапиро-Уилка при различных объемах выборок. Оценить близость получаемых эмпирических распределений статистик к «теоретическим» по процентным точкам таблиц, соответствующим данному критерию. При некотором объеме выборок (n=10) смоделировать распределение статистики критерия при семействе экспоненциальных распределений (двустороннее экспоненциальное) при параметре формы, равном 4÷6. Сравнить с ситуацией, соответствующей справедливой проверяемой гипотезе о нормальном законе. Оценить мощность критерия.
3. Аналогично предыдущему пункту исследовать распределения статистики критерия Эппса-Палли.
4. Исследовать распределения статистики z_1. Проверить близость эмпирических распределений статистики стандартному нормальному закону.
5. Исследовать распределения статистики z_2. Проверить близость эмпирических распределений статистики стандартному нормальному закону.
6. Оценить мощность критериев со статистиками (9.3), (9.4), (9.5), (9.6) относительно заданной альтернативы.

Варианты заданий

№	$F_1(x,\theta_1)$	№	$F_1(x,\theta_1)$
1	Двустороннее экспоненциальное с параметром формы 3	8	Лапласа
2	Двустороннее экспоненциальное с параметром формы 5	9	Коши
3	Двустороннее экспоненциальное с параметром формы 10	10	Двустороннее экспоненциальное с параметром формы 0.5
4	Двустороннее экспоненциальное с параметром формы 7	11	Гамма-распределение с параметром формы 5
5	Распределение экстремальных значений (минимум)	12	Распределение экстремальных значений (максимум)
6	Экспоненциальное	13	Нормальное с масштабом 10
7	Логистическое	14	Нормальное со сдвигом 10

Контрольные вопросы

1. Какие альтернативы плохо различимы критериями проверки отклонения от нормальности Шапиро-Уилка и Эппса-Палли?
2. Можно ли по результатам проверки симметричности закона и проверки эксцесса утверждать о нормальности проверяемых данных?
3. Начиная с каких объемов выборок, критерии согласия не уступают по мощности критериям проверки на нормальность (Шапиро-Уилка, Эппса-Палли)?

[Назад][Вперед] [Содержание]

Лабораторная работа № 10. Исследование распределений статистик в регрессионном анализе при ошибках измерений, отличных от нормального

Цель работы. Статистическое моделирование распределений статистик, используемых при проверке статистических гипотез в регрессионном анализе, при различных законах распределения ошибок измерений. Исследование законов распределения статистик.

Методические указания

1. Линейная регрессия. Линейной регрессией назовем зависимость

$$E[y\,|\,x] = \eta(x,\boldsymbol{\theta}) = f^T(x)\boldsymbol{\theta} = \sum_{i=1}^{m} f_i(x)\boldsymbol{\theta}. \qquad (10.1)$$

Пусть проведено n измерений величины y при некоторых значениях вектора x. Подставив результаты измерений в уравнение (10.1), получаем:

$$y_1 = f_1(x_1)\theta_1 + \ldots + f_m(x_1)\theta_m + e_1,$$
$$\ldots$$
$$y_n = f_1(x_n)\theta_1 + \ldots + f_m(x_n)\theta_m + e_n,$$

где $e_1, e_2, \ldots e_n$ - ошибки измерений.
Обозначим

$$\mathbf{Y} = (y_1, \ldots y_n)^T, \quad \mathbf{X} = \begin{bmatrix} f_1(x_1) & \cdots & f_m(x_1) \\ & \cdots & \\ f_1(x_n) & \cdots & f_m(x_n) \end{bmatrix}, \quad \mathbf{e} = (e_1, \cdots e_n)^T.$$

Тогда уравнение (10.1) принимает вид:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{e}. \qquad (10.2)$$

Здесь $\mathbf{Y}$ – вектор наблюдений размерности $(n \times 1)$, который называют также откликом системы, $\mathbf{X}$ – матрица независимых переменных размерности $(n \times m)$, называемая регрессором системы, θ - вектор оцениваемых параметров размерности $(m \times 1)$, $\mathbf{e}$ – вектор случайных отклонений системы размерности $(n \times 1)$. Наличие в уравнении случайных ошибок говорит о том, что зависимость (10.2) является стохастической.

Предположим, что:

1) На вектор неизвестных параметров регрессии (10.2) не наложено ограничений, т.е. $\Theta = R^m$, где Θ - множество априорных значений вектора параметров θ.

2) Вектор $\mathbf{e} = (e_1 \cdots e_n)^T$ – случайный, отсюда следует, что $\mathbf{Y} = (y_1, \ldots y_n)^T$ – тоже случайный вектор.

3) Математическое ожидание вектора случайных отклонений равно нулю, т.е. $E[e_i] = 0, \quad i = \overline{1, n}$.

4) Для любых $i_1 \neq i_2$ $E[e_{i_1} \times e_{i_2}] = 0$, $E[e_i^2] = \sigma^2$ для всех $i = \overline{1, n}$. Другими словами, $\mathrm{cov}(\mathbf{e}) = \sigma^2 I_n$, здесь σ^2 – дисперсия отклонений, $\mathrm{cov}(\mathbf{e})$ - матрица ковариаций отклонений размерности $(n \times n)$, I_n-единичная матрица размерности $(n \times n)$, т.е.

$$\mathrm{cov}(e) = \begin{bmatrix} \sigma^2 & & \mathrm{O} \\ & \ddots & \\ \mathrm{O} & & \sigma^2 \end{bmatrix}_{n \times n}.$$

5) Матрица $\mathbf{X}$ является детерминированной, т.е. x_{ij} не являются случайными переменными.

6) Ранг матрицы регрессоров $rg(\mathbf{X}) = m$.

Будем считать, что *модель линейной регрессии* задана, если нам известны $\mathbf{Y}, \mathbf{X}$ и закон распределения вектора ошибок.

2. Оценивание параметров линейной регрессии методом максимального правдоподобия. Предположим, что известно распределение вектора случайных отклонений $\mathbf{e}$, а, следовательно, и распределение отклика системы. Пусть отклик системы имеет плотность распределения $f(\mathbf{Y}, \theta)$, где $\theta = (\theta_1, \ldots \theta_m, \sigma^2)$ – вектор неизвестных параметров системы, значение которых необходимо оценить.

Определение. *Функцией правдоподобия* случайной величины $\mathbf{Y} = (y_1, \ldots y_n)^T$ назовем функцию вида:

$$f(\mathbf{Y}, \theta) = \prod_{i=1}^{n} f(y_i, \theta).$$

Определение. *Логарифмической функцией правдоподобия* случайной величины $\mathbf{Y} = (y_1, \ldots y_n)^T$ назовем функцию следующего вида:

$$L(\mathbf{Y}, \theta) = \ln \prod_{i=1}^{n} f(y_i, \theta) = \sum_{i=1}^{n} \ln f(y_i, \theta).$$

Определение. *Оценкой максимального правдоподобия* (ОМП) вектора параметров θ назовем такое значение вектора параметров $\hat{\theta} \in \Theta$ (здесь Θ – множество возможных значений параметра θ), при котором достигается максимум функции правдоподобия или логарифмической функции правдоподобия (что эквивалентно, поскольку функции $\varphi(x)>0$ и $ln\varphi(x)$ достигают экстремумов в одних и тех же точках). Вычислить значение ОМП можно, воспользовавшись одним из методов поиска максимума функции многих переменных.

3. Проверка гипотез в линейном регрессионном анализе. Критерий отношения правдоподобия. Существует общий способ построения критериев проверки статистических гипотез, предложенный Нейманом и Пирсоном. Он аналогичен методу максимального правдоподобия в статистическом оценивании и называется *критерием отношения правдоподобия.*

Пусть плотность распределения $\mathbf{Y}$ равна $f(\mathbf{Y}, \theta)$, т.е. зависит от некоторого неизвестного вектора параметров $\theta \in \Theta$. Для каждого $\mathbf{Y}$ найдем $\max\limits_{\theta \in \Theta_H} f(\mathbf{Y}, \theta)$ и $\max\limits_{\theta \in \Theta} f(\mathbf{Y}, \theta)$ (считаем, что максимум достигается при $\hat{\theta}$).

В качестве критического множества при проверке гипотезы $H: \theta \in \Theta_H$ выбираем

$$W_K = \left\{ \mathbf{Y} \in R^n : \max_{\theta \in \Theta_H} f(\mathbf{Y}, \theta) \Big/ \max_{\theta \in \Theta} f(\mathbf{Y}, \theta) < \varphi_\alpha \right\}, \qquad (10.3)$$

где φ_α – фиксированное число, которое задает верхнюю границу вероятности совершения ошибки первого рода $P_\theta(W_K) \le \alpha$ для всех $\theta \in \Theta_H$. Статистика

$$\max_{\theta \in \Theta_H} f(\mathbf{Y}, \theta) / \max_{\theta \in \Theta} f(\mathbf{Y}, \theta) \tag{10.4}$$

называется *статистикой критерия отношения правдоподобия*.

Рассмотрим линейную гипотезу в самом общем виде

$$H : \mathbf{R}\theta = \mathbf{r}, \tag{10.5}$$

где $\mathbf{R}$ – известная матрица $k \times m$, $rg(\mathbf{R}) = k \le m$; $\mathbf{r}$ – заданный вектор $k \times 1$. Таким образом, в гипотезе H на θ накладывается k независимых линейных ограничений.

Для проверки гипотез необходимо знать вид распределения $\mathbf{Y}$. Положим, что $\mathbf{e} \sim N(0, \sigma^2 \mathbf{I}_n)$, тогда $\mathbf{Y} \sim N(\mathbf{X}\theta, \sigma^2 \mathbf{I}_n)$, и в этом случае статистика критерия отношения правдоподобия имеет вид:

$$Q = \frac{(\mathbf{r} - \mathbf{R}\hat{\theta})^T (\mathbf{R}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{R}^T)^{-1}(\mathbf{r} - \mathbf{R}\hat{\theta})}{(\mathbf{Y} - \mathbf{X}\hat{\theta})^T (\mathbf{Y} - \mathbf{X}\hat{\theta})} \cdot \frac{n-m}{k} \tag{10.6}$$

Отметим частный случай гипотезы (10.5). Если матрица $\mathbf{R} = \mathbf{I}$, т.е. проверяется гипотеза о всех параметрах регрессии ($k = m$), то вид статистики (10.6) упрощается:

$$Q = \frac{(\mathbf{r} - \hat{\theta})^T (\mathbf{X}^T\mathbf{X})(\mathbf{r} - \hat{\theta})}{(\mathbf{Y} - \mathbf{X}\hat{\theta})^T (\mathbf{Y} - \mathbf{X}\hat{\theta})} \cdot \frac{n-m}{m}. \tag{10.7}$$

Доказано [8], что при справедливости проверяемой гипотезы о параметрах линейной регрессии с отклонениями, *распределенными по нормальному закону*, статистика (10.7) критерия отношения правдоподобия подчиняется $F_{k,n-m}$-распределению Фишера со степенями свободы k и $n-m$.

Таким образом, если вычисленное значение статистики не превышает критического $Q < F_{k,n-m,1-\alpha}$, где α – заданный уровень значимости критерия, то проверяемая гипотеза (10.5) не отклоняется.

Распределение Фишера является частным случаем бета-распределения II рода. При использовании программы ISW, задавая необходимое $F_{k,n}$-распределение через бета-распределения II рода, следует учитывать связь между параметрами этих законов:

$$F_{k,n} = BeII\left(\frac{k}{2}, \frac{n}{2}, \frac{n}{k}, 0\right).$$

В случае проверки гипотез относительно параметров регрессии с отклонениями (с ошибками наблюдений), подчиненными законам, отличным от нормального, вид распределения статистики (10.6) не известен.

4. Моделирование значений статистики критерия отношения правдоподобия. На основе известной линейной регрессионной модели вида $Y = X\theta$ проводим эксперимент по следующей схеме:

1) Моделируем вектор ошибок e по заданному закону распределения в соответствии с вариантом задания.
2) Значения матрицы X и вектора θ задаем самостоятельно. На основе известных X, θ, e вычисляем новое значение Y: $Y = X\theta + e$.
3) Находим оценки неизвестных параметров системы методом максимального правдоподобия.
4) Проверяем гипотезу (10.4). Для этого на основе полученных значений вычисляем значение Q статистики критерия по формуле (10.6).
5) Эксперимент повторяем N раз для того, чтобы выборки значений оценок $\hat{\theta}$ и критерия Q имели бы достаточный для исследования объем.

Порядок выполнения.

1. Смоделировать выборки объемом не меньше 500 наблюдений, состоящие из значений статистики критерия отношения правдоподобия при ошибках наблюдений, подчиняющихся закону, указанному в варианте при $n = 10,\ 15,\ 20;\quad m = 3, 4,\ 5$.
2. Проанализировать выборки на предмет принадлежности к бета-распределению II рода, предварительно оценив неизвестные параметры этого распределения.
3. Для получения достоверных результатов пункты 1 - 2 повторить не менее 10 раз для каждой модели регрессии, а затем найти среднее значение каждого из параметров теоретического бета-распределению II рода.
4. Заполнить таблицу усредненных значений параметров бета-распределения II рода для различных значений количеств наблюдений n и количеств неизвестных параметров системы m.

Варианты заданий

№ варианта	Распределение отклонений (ошибок) регрессии	Оцениваемые параметры
1.	Коши	θ, σ^2
2.	Двустороннее экспоненциальное с параметром формы 0,1	θ, σ^2
3.	Двустороннее экспоненциальное с параметром формы 0,5	θ, σ^2
4.	Двустороннее экспоненциальное с параметром формы 1,0	θ, σ^2

5.	Двустороннее экспоненциальное с параметром формы 3,0	θ, σ^2
6.	Двустороннее экспоненциальное с параметром формы 5,0	θ, σ^2
7.	Двустороннее экспоненциальное с параметром формы 10,0	θ, σ^2
8.	Коши	θ, σ^2
9.	Двустороннее экспоненциальное с параметром формы λ=0,1	$\theta, \sigma^2, \lambda$
10.	Двустороннее экспоненциальное с параметром формы λ=0,5	$\theta, \sigma^2, \lambda$
11.	Двустороннее экспоненциальное с параметром формы λ=1,0	$\theta, \sigma^2, \lambda$
12.	Двустороннее экспоненциальное с параметром формы λ=3,0	$\theta, \sigma^2, \lambda$
13.	Двустороннее экспоненциальное с параметром формы λ=5,0	$\theta, \sigma^2, \lambda$
14.	Двустороннее экспоненциальное с параметром формы λ=10,0	$\theta, \sigma^2, \lambda$

Контрольные вопросы

1. Что такое линейная регрессия?
2. Методы оценивания параметров линейной регрессии?
3. Проверка гипотезы о равенстве вектора параметров заданному вектору при нормальном законе распределения ошибок наблюдений?
4. Проверка гипотезы о равенстве вектора параметров заданному вектору при отклонении закона распределения ошибок наблюдений от нормального?

[Назад][Вперед] [Содержание]

Лабораторная работа № 11. Исследование статистик многомерного анализа

Цель работы. Вычисление статистик многомерного анализа по эмпирическим данным, полученным при моделировании псевдослучайных векторов в соответствии с многомерным нормальным законом распределения с различными параметрами. Исследование распределений статистик многомерного анализа в зависимости от объемов выборок.

Методические указания

Введем для дальнейшего использования следующие обозначения:

– $X_1, X_2, \ldots, X_n$ – выборка m-мерного случайного вектора объема n;

– $M = [m_i]_{i=1}^m$ – вектор математического ожидания случайного вектора X;
– $\Sigma = [\sigma_{ij}]_{i,j=1}^m$ – ковариационная матрица случайного вектора X;
– $\hat{M}$ и $\hat{\Sigma}$ – оценки максимального правдоподобия (ОМП) для вектора математического ожидания и ковариационной матрицы, вычисляемые по негруппированным данным:

$$\hat{M} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \hat{\Sigma} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{M})(X_i - \hat{M})^T.$$

В данной работе исследуются распределения статистик критериев, связанных с проверкой следующих гипотез.

1. Проверка гипотезы о равенстве вектора математических ожиданий некоторому известному вектору. Важной статистической проблемой является проблема проверки гипотезы о том, что вектор математических ожиданий (МО) наблюдаемых многомерных случайных величин равен заданному. Такая задача возникает на практике, когда, например, на основании наблюдений некоторого технологического процесса желают убедиться, что эти показатели равны номинальному значению, т.е. процесс протекает нормально, а отклонения наблюдаемых значений от номинальных объясняются лишь ошибками наблюдений (измерений).

Проверяемая гипотеза имеет вид $H_0 : M = M_0$. Возможны две ситуации.

а) Когда нам известна ковариационная матрица Σ (например, из ранее проводимых экспериментов или предположений), то для вывода распределения статистики критерия, используется факт, что разность между векторами среднего значения выборки и среднего значения генеральной совокупности распределена нормально с нулевым вектором математического ожидания и известной ковариационной матрицей.

В этом случае статистика критерия

$$X_m^2 = n(\hat{M} - M_0)^T \Sigma^{-1}(\hat{M} - M_0) \qquad (11.1)$$

при справедливости проверяемой гипотезы H_0 подчиняется χ_m^2- распределению с числом степеней свободы m.

б) В случае неизвестной ковариационной матрицы используется критерий со статистикой

$$T^2 = \frac{n(n-m)}{m(n-1)}(\hat{M} - M_0)^T \hat{\Sigma}^{-1}(\hat{M} - M_0). \qquad (11.2)$$

При справедливости гипотезы H_0 статистика подчиняется $F_{m,n-m}$ - распределению Фишера с параметрами m и n-m.

2. Проверка гипотез о коэффициенте парной корреляции. При анализе совокупности случайных величин нас может интересовать взаимосвязь между несколькими случайными величинами, или зависимость одной или большего числа величин от остальных. Когда рассматривается взаимосвязь двух величин,

то речь идет о парной корреляции. Взаимозависимость же величин при устранении влияния совокупности других – это частная корреляция, а зависимость одной величины от группы величин характеризуется множественной корреляцией.

Под парной корреляцией понимается обычная корреляция между двумя величинами. Если оценка ковариационной матрицы $\hat{\Sigma}$ уже известна, то оценка соответствующего парного коэффициента корреляции определяется выражением

$$\hat{r}_{ij} = \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}}.$$

Можно находить оценки парного коэффициента корреляции, проверять гипотезы о его значимости ($H_0 : r_{ij} = 0$) и о равенстве его определенному значению ($H_0 : r_{ij} = r_0$).

При проверке гипотезы $H_0 : r_{ij} = 0$ используется статистика

$$t = \frac{\sqrt{n-2}\left|\hat{r}_{ij}\right|}{\sqrt{1-\hat{r}_{ij}^2}}, \qquad (11.3)$$

которая при справедливости H_0 подчиняется t_{n-2} - распределению Стьюдента с числом степеней свободы n-2.

В случае проверки гипотезы $H_0 : r_{ij} = r_0$ используется статистика

$$z_0 = \sqrt{n-3}\left(\frac{1}{2}\ln\left(\frac{1+\hat{r}_{ij}}{1-\hat{r}_{ij}}\right) - \frac{1}{2}\ln\left(\frac{1+r_0}{1-r_0}\right) - \left(\frac{r_0}{2(n-1)}\right)\right), \qquad (11.4)$$

имеющая в качестве предельного стандартное нормальное распределение.

3. Проверка гипотез о коэффициенте частной корреляции. В случае двух нормальных или почти нормальных величин коэффициент корреляции между ними может быть использован в качестве меры взаимозависимости. На практике при интерпретации «взаимозависимости» встречаются определенные трудности. Если одна величина коррелирована с другой, то это может быть всего лишь отражением того факта, что они обе коррелированы с некоторой третьей величиной или с совокупностью величин. Это приводит к необходимости анализа условных корреляций между двумя величинами при фиксированных значениях остальных величин. Это так называемые *частные корреляции*.

Если корреляция между двумя величинами уменьшается при фиксации некоторой другой случайной величины, то это означает, что их взаимозависимость возникает частично через воздействие этой величины. Если же частная корреляция равна нулю или очень мала, то делается вывод, что их взаимозависимость целиком обусловлена этим воздействием. Когда же частная корреляция оказывается больше первоначальной корреляции между двумя величинами, то следует, что другие компоненты ослабляли связь, или, можно

сказать, «маскировали» корреляцию. Необходимо помнить, что даже в последнем случае нельзя предполагать наличие причинной связи, так как некоторая, совершенно отличная от рассматриваемых при анализе, величина может быть источником этой корреляции. Как при обычной корреляции, так и при частных корреляциях предположение о причинности должно иметь внестатистические основания.

Представим случайный вектор X в следующем виде:

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix},$$

где $X_1 = (x_1, x_2, \ldots, x_l)^T$, $X_2 = (x_{l+1}, x_{l+2}, \ldots, x_m)^T$. Соответственно вектор математических ожиданий и ковариационную матрицу –

$$M = \begin{bmatrix} M_1 \\ M_2 \end{bmatrix}, \ \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

Если случайный вектор X подчиняется нормальному закону с вектором средних M и ковариационной матрицей Σ, то условное распределение подвектора X_1 при известном X_2 является нормальным с математическим ожиданием $M_1 + B(X_2 - M_2)$ и ковариационной матрицей $\Sigma_{11\bullet 2}$, где $B = \Sigma_{12}\Sigma_{22}^{-1}$, $\Sigma_{11\bullet 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

ОМП для частного коэффициента корреляции определяется следующим соотношением

$$\hat{r}_{ij;l+1,\ldots,m} = \frac{\hat{\sigma}_{ij;l+1,\ldots,m}}{\sqrt{\hat{\sigma}_{ii;l+1,\ldots,m}\hat{\sigma}_{jj;l+1,\ldots,m}}},$$

где $\hat{\sigma}_{ij\bullet l+1,\ldots,m}$ - элемент i-й строки и j-го столбца матрицы $\Sigma_{11\bullet 2}$, l – число компонент в условном распределении $f(X_1 | X_2)$, $2 \le l \le m$.

В случае оценки взаимозависимости между компонентами x_i и x_j $(i, j \le l)$ многомерной случайной величины X исключается влияние компонент $x_{l+1}, x_{l+2}, \ldots, x_m$.

В критериях проверки гипотез вида $H_0 : r_{ij\bullet l+1,\ldots,m} = 0$ и $H_0 : r_{ij\bullet l+1,\ldots,m} = r_0$ используются статистики, аналогичные применяемым в критериях для парного коэффициента корреляции. В этом случае в соответствующих соотношениях n заменяется на $n - m + l$.

Статистика критерия проверки гипотезы $H_0 : r_{ij;l+1,\ldots,m} = 0$

$$t = \frac{\sqrt{n - m + l - 2}\left|\hat{r}_{ij\bullet l+1,\ldots,m}\right|}{\sqrt{1 - \hat{r}_{ij\bullet l+1,\ldots,m}}}. \qquad (11.5)$$

подчиняется $t_{n-m+l-2}$ – распределению Стьюдента с числом степеней свободы n-m+l-2.

Используемая при проверке гипотезы $H_0 : r_{ij;l+1,\ldots,m} = r_0$ статистика

$$z_0 = \sqrt{n-3}\left(\frac{1}{2}\ln\left(\frac{1+\hat{r}_{ij\bullet l+1,\ldots,m}}{1-\hat{r}_{ij\bullet l+1,\ldots,m}}\right) - \frac{1}{2}\ln\left(\frac{1+r_0}{1-r_0}\right) - \left(\frac{r_0}{2(n-1)}\right)\right) \qquad (11.6)$$

в пределе распределена по стандартному нормальному закону.

4. Проверка гипотезы о коэффициенте множественной корреляции. Множественный коэффициент корреляции является мерой взаимозависимости компоненты многомерной случайной величины и некоторого множества компонент. Можно рассматривать корреляцию между одной компонентой случайного вектора и множеством всех остальных или каким-то подмножеством.

Подчеркнем, что множественный коэффициент корреляции случайной величины x_i относительно некоторого множества других случайных величин всегда не меньше, чем абсолютная величина любого парного коэффициента корреляции r_{ij} с таким же первичным индексом. Более того, множественный коэффициент корреляции никогда нельзя уменьшить путем расширения множества величин, относительно которых измеряется зависимость x_i.

Если коэффициент корреляции между x_i и множеством всех остальных компонент многомерной случайной величины равен нулю ($r_i = 0$), то все коэффициенты корреляции этой величины относительного любого подмножества также равны 0, т.е. величина x_i полностью некоррелирована со всеми остальными величинами.

С другой стороны, если r_i относительно множества всех остальных компонент равен единице $r_i = 1$, то, по крайней мере, один из коэффициентов корреляции относительно некоторого подмножества компонент должен быть равен 1.

Надо отметить, что коэффициент корреляции, например, между x_1 и множеством всех остальных компонент является обычным коэффициентом корреляции между x_1 и условным математическим ожиданием $E[x_1 \mid x_2, x_3, \ldots, x_n]$.

Если представить случайный вектор X в том виде, как это было показано в разделе частной корреляции, то ОМП множественного коэффициента корреляции между x_i, $i \le l$ и множеством компонент $x_{l+1}, x_{l+2}, \ldots, x_m$ определяется соотношением

$$\hat{r}_{i;l+1,\ldots,m} = \sqrt{\frac{\hat{\sigma}_{(i)}\Sigma_{22}^{-1}\hat{\sigma}_{(i)}^T}{\hat{\sigma}_{ii}}},$$

где $\sigma_{(i)}$ – i-я строка матрицы Σ_{12}, σ_{ii} – элемент матрицы Σ_{11}.

При проверке гипотезы $H_0 : r_{i\bullet l+1,\ldots,m} = 0$ статистика критерия

$$F = \frac{n-m+l-1}{m-l} \frac{\hat{r}^2_{i\bullet l+1,\ldots,m}}{1-\hat{r}^2_{i\bullet l+1,\ldots,m}} \qquad (11.7)$$

подчиняется $F_{m-l,n-m+l-1}$ – распределению Фишера со степенями свободы *m-l* и *n-m+l*-1.

Порядок выполнения

1. Вычислить, если это необходимо, оценки коэффициентов корреляции.

2. Используя программу **STAR System.exe**, в соответствии с вариантом многомерного нормального закона и вариантом критерия проверки гипотез смоделировать выборки статистик с сохранением в файл. Исследовать влияние различных значений объемов выборок многомерной случайной величины на близость эмпирических распределений статистик рассматриваемых критериев соответствующим теоретическим законам. Близость оценивать по достигаемым уровням значимости критериев согласия при проверке гипотез о соответствии смоделированных эмпирических распределений статистик теоретическим законам. Определить какие статистики многомерного анализа зависят от объемов выборок многомерной случайной величины.

3. Построить графики эмпирических функций распределения смоделированных выборок статистик корреляционного анализа и их теоретических распределений.

4. На основании полученных результатов сделать выводы об исследуемых распределениях статистик многомерного (корреляционного) анализа.

Варианты заданий

Наборы параметров для моделирования многомерных нормальных величин:

$$M = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad (P1)$$

$$M = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \Sigma = \begin{bmatrix} 5 & 1 & 2{,}5 \\ 1 & 6 & 1 \\ 2{,}5 & 1 & 5 \end{bmatrix} \qquad (P2)$$

$$M = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0{,}5 \\ 0 & 0{,}5 & 1 \end{bmatrix} \qquad (P3)$$

$$M = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{bmatrix}, \ \Sigma = \begin{bmatrix} 10 & 2 & 1 & 0{,}5 & 0{,}3 \\ 2 & 10 & 2 & 1 & 0{,}5 \\ 1 & 2 & 10 & 2 & 1 \\ 0{,}5 & 1 & 2 & 10 & 2 \\ 0{,}3 & 0{,}5 & 1 & 2 & 10 \end{bmatrix} \qquad \text{(P4)}$$

Варианты проверяемых гипотез:

1. О равенстве МО некоторому известному вектору при известной ковариационной матрице.
2. О равенстве МО некоторому известному вектору при неизвестной ковариационной матрице.
3. О нулевом парном коэффициенте корреляции.
4. О парном коэффициенте корреляции, которой равен заданному значению.
5. О нулевом частном коэффициенте корреляции.
6. О частном коэффициенте корреляции, которой равен заданному значению.
7. О нулевом множественном коэффициенте корреляции.

№ варианта	Параметры моделирования	Гипотезы	Индексы коэффициентов корреляции
1.	(P4)	1, 2	
2	(P2), (P3)	3 с параметрами (3), 4 с параметрами (2)	$R_{1,2}$
3.	(P2), (P3)	5 с параметрами (3), 6 с параметрами (2)	$R_{1,2}$ для 5, $R_{1,2}$ для 6
4.	(P3)	5, 7	$R_{1,2}$
5.	(P1), (P4)	1, 2	

Контрольные вопросы

1. Какие законы используются в качестве распределений статистик при проверке гипотез о векторе математического ожидания, и как осуществляется проверка гипотез?
2. Какие законы используются в качестве распределений статистик при проверке гипотез о парном коэффициенте корреляции, и как осуществляется проверка гипотез?
3. Какие законы используются в качестве распределений статистик при проверке гипотез о частном коэффициенте корреляции, и как осуществляется проверка гипотез?
4. Какой закон используется в качестве распределения статистики при проверке гипотезы о множественном коэффициенте корреляции, и как осуществляется проверка гипотезы?

Лабораторная работа № 12. Исследование распределений статистик критериев проверки однородности выборок

Цель работы. Исследовать распределения статистик критериев однородности Смирнова и Лемана-Розенблатта. Исследовать и проанализировать мощность критериев.

Методические указания

Задача проверки однородности двух выборок формулируется следующим образом. Пусть имеется две упорядоченные по возрастанию выборки размера m и n:

$$x_1 < x_2 < \ldots < x_m \quad \text{и} \quad y_1 < y_2 < \ldots < y_n .$$

Для определенности обычно полагают, что $m \le n$. Проверяется гипотеза о том, что две выборки извлечены из одной и той же генеральной совокупности, т.е. H_0: $F(x) = G(x)$ при любом x.

1. Критерий однородности Смирнова. Предполагается, что функции распределения $F(x)$ и $G(x)$ являются непрерывными. Статистика критерия Смирнова измеряет различие между эмпирическими функциями распределения, построенными по выборкам

$$D_{m,n} = \sup_x \left| G_m(x) - F_n(x) \right| .$$

При практическом использовании критерия значение статистики $D_{m,n}$ вычисляют в соответствии с соотношениями:

$$D^+_{m,n} = \max_{1 \le r \le m} \left[\frac{r}{m} - F_n(x_r) \right] = \max_{1 \le s \le n} \left[G_m(y_s) - \frac{s-1}{n} \right],$$

$$D^-_{m,n} = \max_{1 \le r \le m} \left[F_n(x_r) - \frac{r-1}{m} \right] = \max_{1 \le s \le n} \left[\frac{s}{n} - G_m(y_s) \right],$$

$$D_{m,n} = \max(D^+_{m,n}, D^-_{m,n}) .$$

Если гипотеза H_0 справедлива, то при неограниченном увеличении объемов выборок $\lim\limits_{m \to \infty} P\left\{ \sqrt{\frac{mn}{m+n}} D_{m,n} < s \right\} = K(s)$, т.е. статистика

$$S_C = \sqrt{\frac{mn}{m+n}} D_{m,n} \tag{12.1}$$

в пределе подчиняется распределению Колмогорова $K(s)$.

2. Критерий однородности Лемана-Розенблатта. Критерий однородности Лемана-Розенблатта представляет собой критерий типа ω^2. Критерий был предложен в Леманом и подробно исследован в Розенблаттом. Статистика критерия имеет вид

$$T = \frac{mn}{m+n}\int_{-\infty}^{\infty}[G_m(x) - F_n(x)]^2 dH_{m+n}(x),$$

где $H_{m+n}(x) = \frac{m}{m+n}G_m(x) + \frac{n}{m+n}F_n(x)$ – эмпирическая функция распределения, построенная по вариационному ряду объединения двух выборок. Как правило, статистика T используется в форме

$$T = \frac{1}{mn(m+n)}\left[n\sum_{i=1}^{n}(r_i - i)^2 + m\sum_{i=1}^{m}(s_i - j)^2\right] - \frac{4mn-1}{6(m+n)}, \qquad (12.2)$$

где r_i – порядковый номер (ранг) y_i, s_j – порядковый номер (ранг) x_j в объединенном вариационном ряде.

Розенблаттом было показано, что статистика (2) в пределе распределена как $a1(t)$ [4]:

$$\lim_{\substack{m\to\infty \\ n\to\infty}} P\{T < t\} = a1(t).$$

Это то же самое распределение, которому подчинена статистика критерия согласия ω^2 Крамера-Мизеса-Смирнова при проверке простых гипотез.

Более подробная информация о критериях однородности доступна по адресу: http://www.ami.nstu.ru/~headrd/seminar/publik_html/Izm_T_8.htm

Порядок выполнения

Используя программу **ISW**, исследовать распределения статистик критериев проверки однородности Смирнова и Лемана-Розенблатта.

1. Исследовать сходимость к предельным распределениям, для чего смоделировать распределения статистик критериев при различных объемах выборок (n_i=10, 20, 50, 100) [H_0: $F(x) = G(x) = F(x,\theta)$] при $m_i = n_i$. Объем выборок статистик 5000-10000 наблюдений. Оценить близость к предельным распределениям визуально (по графикам) и по критериям согласия. Провести данные исследования отдельно для критерия Смирнова и критерия Лемана-Розенблатта. Для критерия Смирнова исследовать влияние на распределение статистки выбора различных объемов выборок $m_i \neq n_i$.
2. Оценить мощность критериев относительно заданной альтернативы [H_1: $F(x) = F(x,\theta)$; $G(x) = F_1(x,\theta)$] при различных объемах выборок.

Варианты заданий.

№	$F(x,\theta)$	$F_1(x,\theta_1)$
1	Нормальное (0, 1)	Двустороннее экспоненциальное с

		параметром формы 3
2	Нормальное (0, 2)	Коши
3	Нормальное (0, 1)	Двустороннее экспоненциальное с параметром формы 1
4	Экспоненциальное	Гамма-распределение с параметром формы 2
5	Нормальное (0, 1)	Su-Джонсона
6	Экспоненциальное	Вейбулла-Гнеденко (2,1)
7	Логистическое (0, 1)	Нормальное (0,2)
8	Нормальное (0, 1)	Двустороннее экспоненциальное с параметром формы 5
9	Логарифмически нормальное (0,1)	Вейбулла-Гнеденко (2)

Контрольные вопросы

1. В чем отличие критериев однородности от критериев однородности средних и однородности дисперсий?
2. Какие недостатки можно отметить в критерии Смирнова и как устранить влияние негативных факторов?
3. Можно ли однозначно отдать предпочтение некоторому критерию однородности?

[Назад][Вперед] [Содержание]

Лабораторная работа № 13. Исследование распределений статистик критериев проверки гипотез о математических ожиданиях и дисперсиях

Цель работы. Исследовать, что происходит с распределениями классических статистик, используемых в критериях проверки гипотез о математических ожиданиях и дисперсиях, если наблюдаемый закон в той или иной мере отличается от нормального. Проверить, насколько будут корректны статистические выводы, базирующиеся на классических результатах, если нарушено предположение о нормальности.

Методические указания

Пусть мы имеем выборку n случайных величин, распределенных по нормальному закону $x_1,...,x_n \in N(\mu_{ист},\sigma^2_{ист})$. В этом случае задачи проверки гипотез о математических ожиданиях и дисперсиях формулируются следующим образом.

1. В критерии проверки гипотез вида $H_0 : \mu = \mu_0$ при известной дисперсии $\sigma^2_{ист}$ используется статистика $T_1 = \frac{1}{n}\sum_{i=1}^{n} x_i - \mu_0$, которая при справедливости гипотезы H_0 подчиняется нормальному распределению: $G(T_1 \mid H_0) = N\left(0, \frac{\sigma^2_{ист}}{n}\right)$. Проверяемая гипотеза H_0 отклоняется при больших отклонениях T_1 от 0.

2. Для проверки гипотезы $H_0 : \mu = \mu_0$ при неизвестной дисперсии $\sigma^2_{ист}$ используется статистика $T_2 = \frac{\bar{x} - \mu_0}{\bar{\sigma}}\sqrt{n}$, где $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\bar{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$. При справедливости H_0 статистика T_2 распределена как $G(T_2 \mid H_0) = t_{n-1}$ – распределение Стьюдента.

3. Для проверки гипотезы вида $H_0 : \sigma^2 = \sigma_0^2$ при известном математическом ожидании $\mu_{ист}$ вычисляется статистика $T_3 = \frac{1}{\sigma_0^2}\sum_{i=1}^{n}(x_i - \mu_{ист})^2$, условным распределением которой является $G(T_3 \mid H_0) = \chi^2_n$ – распределение.

4. В критерии проверки гипотезы вида $H_0 : \sigma^2 = \sigma_0^2$ при неизвестном математическом ожидании $\mu_{ист}$ используется статистика $T_4 = \frac{1}{\sigma_0^2}\sum_{i=1}^{n}(x_i - \bar{x})^2$, подчиняющаяся $G(T_3 \mid H_0) = \chi^2_{n-1}$ – распределению.

Более подробная информация доступна по адресу:
http://www.ami.nstu.ru/~headrd/seminar/publik_html/Izm_T_3.htm

Порядок выполнения

1. Исследовать распределения статистик T_1, T_2, T_3, T_4 в условиях принадлежности наблюдений нормальному закону с заданными параметрами при некотором объеме выборок n. Объем выборок статистик N задавать порядка 5000-10000. Убедиться в близости полученных распределений статистик теоретическим законам визуально. Проверить согласие полученных эмпирических распределений статистик с соответствующими теоретическими, для чего использовать все доступные критерии согласия.
2. Исследовать (оценить) мощность критериев относительно заданных конкурирующих гипотез ($H_1 : \mu = \mu_0 + \delta_\mu$, $H_1 : \sigma^2 = (\sigma_0 + \delta_\sigma)^2$) и заданного n в случае нормального закона.

3. Исследовать влияние отклонения наблюдаемого закона от нормального на распределения статистик T_1, T_2. Для этого смоделировать распределения статистик: при различных симметричных законах наблюдаемых случайных величин с различной “тяжестью хвостов распределений”; при асимметричных законах. Определить, какие отклонения от нормального закона не приводят к заметным изменениям распределений статистик, в каких случаях заметно изменение в распределениях, какие отклонения приводят к существенным изменениям распределений статистик T_1, T_2?
4. Провести аналогичные исследования распределений статистик T_3, T_4, убедиться в существенной зависимости распределений статистик от наблюдаемого закона.

В процессе выполнения работы использовать программу **ISW.**

Варианты заданий.

№	$F(x,\theta)$	$F_1(x,\theta_1)$
1	Нормальное (0, 1)	Двустороннее экспоненциальное с параметром формы 1
2	Нормальное (0.1, 1)	Двустороннее экспоненциальное с параметром формы 3
3	Нормальное (0.1, 2)	Двустороннее экспоненциальное с параметром формы 5
4	Нормальное (0.5, 1)	Коши
5	Нормальное (0.5, 2)	Двустороннее экспоненциальное с параметром формы 0.5
6	Нормальное (0, 1.5)	Двустороннее экспоненциальное с параметром формы 0.2
7	Нормальное (0, 2)	Гамма-распределение с параметром формы 2
8	Нормальное (0, 3)	Распределение экстремальных значений (минимум)
9	Нормальное (0, 3)	Распределение экстремальных значений (максимум)
10	Нормальное (1, 3)	Экспоненциальное распределение

Контрольные вопросы

1. Виды статистик, используемых в классических критериях проверки гипотез о математических ожиданиях и дисперсиях?
2. При каких видах нарушений исходных предположений применение классических критериев проверки гипотез о математических ожиданий остается корректным?

3. Варианты действий в ситуации, когда применение классических критериев проверки гипотез о математических ожиданий невозможно?
4. Варианты действий в случае проверки гипотез о дисперсиях и нарушении предположений о нормальности наблюдаемого закона?

[Назад][Вперед] [Содержание]

Лабораторная работа № 14. Исследование распределений статистик критериев проверки однородности средних

Цель работы. Исследовать распределения статистик параметрических и непараметрических критериев, используемых для проверки гипотез об однородности средних.

Методические указания

Проверяемая гипотеза о равенстве математических ожиданий (об однородности математических ожиданий) случайных величин, соответствующих двум выборкам, задается в виде

$$H_0 : \mu_1 = \mu_2 ,$$

а конкурирующая –

$$H_1 : \mu_1 \neq \mu_2$$

В общем случае, гипотеза о равенстве математических ожиданий имеет вид

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_m$$

при конкурирующей

$$H_1 : \mu_{i_1} \neq \mu_{i_2} .$$

Для проверки гипотезы H_0 может использоваться целый ряд критериев. Условием применения параметрических критериев является принадлежность наблюдений нормальному закону. К таким критериям относятся: критерий сравнения двух выборочных средних при известных дисперсиях; сравнения двух выборочных средних при неизвестных, но равных дисперсиях (критерий Стьюдента); сравнения двух выборочных средних при неизвестных и неравных дисперсиях (проблема Беренса-Фишера). Для этих же целей предназначена целая совокупность непараметрических критериев: U –критерий Уилкоксона, критерий Манна–Уитни, H –критерий Краскела–Уаллиса.

1. Сравнение двух выборочных средних при известных дисперсиях. Применение критерия сравнения двух выборочных средних (по двум выборкам) при известных и равных дисперсиях предусматривает вычисление статистики

$$z = \frac{x_1 - x_2 - \mu_1 + \mu_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}, \qquad (14.1)$$

где $\bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_j$, n_i - объем i-й выборки, $i = 1, 2$. В случае принадлежности наблюдений нормальным законам статистика z подчиняется стандартному нормальному закону.

2. Сравнение двух выборочных средних при неизвестных, но равных дисперсиях. Применение критерия сравнения двух выборочных средних при неизвестных, но равных дисперсиях предусматривает вычисление статистики t:

$$t = \frac{\bar{x}_1 - \bar{x}_2 - \mu_1 + \mu_2}{\sqrt{\left[\frac{n_1 + n_2}{n_1 n_2}\right]\left[\frac{Q_1 + Q_2}{n_1 + n_2 - 2}\right]}}, \qquad (14.2)$$

где n_i – объем i-й выборки,

$$Q_i = \sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2, \quad \bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}.$$

В случае принадлежности выборок нормальному закону при справедливости гипотезы H_0 эта статистика подчиняется t_ν-распределению Стьюдента с числом степеней свободы $\nu = n_1 + n_2 - 2$.

3. Сравнение двух выборочных средних при неизвестных и неравных дисперсиях. При неравных объемах выборок $n_1 \neq n_2$ статистика критерия имеет вид:

$$t = \frac{\bar{x}_1 - \bar{x}_2 - \mu_1 + \mu_2}{\sqrt{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]}}, \qquad (14.3)$$

где n_i – объем i-й выборки, $s_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2$, $\bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$.

В случае нормального закона и справедливости гипотезы H_0 статистика (14.3) подчиняется распределению t_ν-Стьюдента с числом степеней свободы

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}.$$

При равных объемах выборок $n_1 = n_2 = n$ –

$$t = \frac{\bar{x}_1 - \bar{x}_2 - \mu_1 + \mu_2}{\sqrt{\left[\frac{s_1^2 + s_2^2}{n}\right]}},$$

а число степеней свободы распределения Стьюдента –

$$v = n - 1 + \frac{2n-2}{\frac{s_1^2}{s_2^2} + \frac{s_2^2}{s_1^2}}.$$

4. U**–критерий Уилкоксона, Манна и Уитни.** Ранговый критерий Манна и Уитни основан на критерии Уилкоксона для независимых выборок. Он является непараметрическим аналогом t-критерия для сравнения двух средних значений непрерывных распределений. Для вычисления статистики упорядочивают $n_1 + n_2$ значений объединенной выборки, определяют сумму рангов R_1, соответствующую элементам первой выборки, и сумму рангов второй R_2. Вычисляются

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1,$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2.$$

Статистика критерия имеет вид: $U = \min\{U_1, U_2\}$.

Для достаточно больших выборок ($n_1 + n_2 > 60$), когда объемы выборок не слишком малы ($n_1 \geq 8, n_2 \geq 8$) используется статистика

$$\hat{z} = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 \quad n_1 + n_2 + 1}{12}}}, \tag{14.4}$$

которая приближенно распределена в соответствии со стандартным нормальным законом.

5. H**–критерий Краскела–Уаллиса**. H–критерий Краскела–Уаллиса является развитием U–критерия для проверки гипотезы (о равенстве средних) по k выборкам. Объединенную выборку $n = \sum_{i=1}^{k} n_i$ упорядочивают и вычисляют суммы рангов R_i для i-й выборки, $i = \overline{1,k}$. Статистика для проверки гипотезы H_0 имеет вид

$$H = \left[\frac{12}{n(n+1)}\right]\left[\sum_{i=1}^{k} \frac{R_i^2}{n_i}\right] - 3(n+1). \tag{14.5}$$

H представляет собой дисперсию ранговых сумм. При больших n_i и k (практически $n_i \geq 5,\ k \geq 4$) при справедливости гипотезы H_0 статистика подчиняется χ^2_{k-1}-распределению.

Порядок выполнения.

В процессе выполнения работы использовать программу **ISW.**

1. По указанию преподавателя исследовать распределения статистик (14.1), (14.2) или (14.3) в условиях принадлежности наблюдений нормальным законам с заданными параметрами при некоторых объемах выборок. Объем выборок статистик N задавать порядка 5000-10000. Убедиться в близости полученных распределений статистик соответствующим теоретическим законам визуально. Проверить согласие полученных эмпирических распределений статистик с соответствующими теоретическими, для чего использовать все доступные критерии согласия.
2. Исследовать влияние нарушений предположения о нормальности наблюдаемых законов на распределение исследуемой статистики. Для этого смоделировать распределение статистики при тех же объемах выборок: при различных симметричных законах наблюдаемых случайных величин с различной “тяжестью хвостов распределений”; при асимметричных законах. Определить, какие отклонения от нормального закона не приводят к заметным изменениям распределений статистики, в каких случаях заметно изменение в распределении, какие отклонения приводят к существенным изменениям распределения статистики?
3. Исследовать (оценить) мощность соответствующего критерия относительно заданных конкурирующих гипотез и заданных объемов выборок в случае нормального закона. Для расчета значений мощности критериев рассмотреть одну из конкурирующих гипотез, для которой величина математического ожидания одной из выборок отличается на величину $w_i \cdot \sigma$, где $w_1 = 0.1$, $w_2 = 0.2$, $w_3 = 0.3$, $w_4 = 0.4$, $w_5 = 0.5$ (выбор w_i по заданию преподавателя).
4. Исследовать сходимость распределения статистики (14.4) критерия Манна-Уитни к стандартному нормальному закону с ростом объемов выборок. Убедиться, что распределения статистики (14.4) не зависят от вида закона, из которого извлекаются выборки.
5. Исследовать сходимость распределения статистики (14.5) критерия Краскела–Уаллиса к соответствующему χ^2_{k-1}-распределению в зависимости от объемов выборок. Проверить, что критерий Краскела–Уаллиса действительно является непараметрическим, и распределения статистики (14.5) не зависят от вида закона, из которого извлекаются выборки.
6. Для тех же конкурирующих гипотез и объемов выборок (см. п.3) оценить мощность критериев Манна-Уитни и Краскела–Уаллиса.

Варианты для выбора законов совпадают с вариантами из лабораторной работы № 13.

Контрольные вопросы

1. Виды статистик, используемых в классических (параметрических) критериях проверки гипотез об однородности математических ожиданий?
2. При каких видах нарушений исходных предположений применение классических (параметрических) критериев проверки гипотез об однородности математических ожиданий остается корректным?
3. Что можно предпринять в ситуации, когда применение классических критериев проверки гипотез об однородности математических ожиданий невозможно?

[Назад][Вперед] [Содержание]

Лабораторная работа № 15. Исследование распределений статистик критериев проверки однородности дисперсий

Цель работы. Исследовать распределения статистик критериев, используемых для проверки гипотез об однородности дисперсий.

Методические указания

Проверяемая гипотеза о постоянстве дисперсии m выборок имеет вид:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_m^2.$$

а конкурирующая с ней гипотеза –

$$H_1 : \sigma_{i_1}^2 \neq \sigma_{i_2}^2,$$

где неравенство выполняется, по крайней мере, для одной пары индексов i_1, i_2. Для проверки такого вида гипотез применяется ряд критериев. Основополагающим предположением для всех критериев является принадлежность наблюдений (ошибок измерений) нормальному закону.

1. Критерий Бартлетта. Статистика критерия Бартлетта вычисляется в соответствии с соотношением:

$$\chi^2 = M\left[1 + \frac{1}{3(m-1)}\left(\sum_{i=1}^{m}\frac{1}{\nu_i} - \frac{1}{N}\right)\right]^{-1} \qquad (15.1)$$

где n_i – объемы выборок, $\nu_i = n_i$, если математическое ожидание известно, и $\nu_i = n_i - 1$, если неизвестно, $N = \sum_{i=1}^{m}\nu_i$,

$$M = N\ln\left(\frac{1}{N}\sum_{i=1}^{m}\nu_i S_i^2\right) - \sum_{i=1}^{m}\nu_i \ln S_i^2 ,$$

S_i^2 – оценки выборочных дисперсий. При неизвестном математическом ожидании оценки $S_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(X_{ji} - \overline{X}_i)^2$, где $\overline{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{ji}$ и $\nu_i = n_i - 1$. Если гипотеза H_0 верна, все $\nu_i > 3$ и выборки извлекаются из нормальной генеральной совокупности, то статистика (1) приближенно подчиняется χ^2_{m-1}-распределению.

2. Критерий Кокрена. Когда все n_i одинаковы, $n_1 = n_2 = ... = n_m = n$, возможно использование более простого критерия Кокрена. Статистика Q критерия Кокрена выражается формулой

$$Q = \frac{S^2_{\max}}{S_1^2 + S_2^2 + \cdots + S_m^2}, \qquad (15.2)$$

где $S^2_{\max} = \max\{S_1^2, S_2^2, ..., S_m^2\}$, где m – число независимых оценок дисперсий (число выборок).

Распределения статистики Кокрена сильно зависят от объема наблюдаемых выборок. Поэтому в справочной литературе приводятся только таблицы процентных точек, которые и используются при проверке гипотез.

3. Критерий Фишера для сравнения двух выборочных дисперсий из нормальных совокупностей. Для определения того, относятся ли две выборки к одной и той же генеральной совокупности, проверяется гипотеза вида $H_0 : \sigma_1^2 = \sigma_2^2$. Статистика для проверки гипотезы имеет вид

$$F = \frac{s_1^2}{s_2^2}. \qquad (15.3)$$

В случае принадлежности выборок нормальному закону и справедливости H_0 эта статистика подчинятся F_{ν_1,ν_2}-распределению Фишера с числом степеней свободы $\nu_1 = n_1 - 1$ и $\nu_2 = n_2 - 1$.

4. Проверка равенства нескольких дисперсий для выборок равного объема по Хартли. Статистика для проверки гипотезы имеет вид

$$F = \frac{s^2_{\max}}{s^2_{\min}}. \qquad (15.4)$$

Степенями свободы для распределения статистики являются число выборок $\nu_1 = m$ и $\nu_2 = n - 1$. В литературе для статистики (15.4) приводятся лишь таблицы процентных точек.

5. Критерий Левене об однородности дисперсий. Считается, что критерий Левене менее чувствителен к отклонениям от нормальности, чем критерий Бартлетта.

Пусть n_i − объем i-й выборки, $N = \sum_{i=1}^{m} n_i$, X_{ij} − j-е наблюдение в i-й выборке. Статистика критерия Левене имеет вид:

$$W = \frac{N-m}{m-1} \frac{\sum_{i=1}^{m} n_i (\overline{Z}_{i\bullet} - \overline{Z}_{\bullet\bullet})^2}{\sum_{i=1}^{m} \sum_{j=1}^{n_i} (Z_{ij} - \overline{Z}_{i\bullet})^2}, \qquad (15.5)$$

где Z_{ij} определяется выражением $Z_{ij} = \left| X_{ij} - \overline{X}_{i\bullet} \right|$, в котором $\overline{X}_{i\bullet}$ − среднее в i-й выборке. $\overline{Z}_{i\bullet}$ − среднее Z_{ij} по i-й выборке, $\overline{Z}_{\bullet\bullet}$ − среднее Z_{ij} по всем выборкам. В случае принадлежности выборок нормальному закону и справедливости H_0 эта статистика подчинятся F_{ν_1,ν_2}-распределению Фишера с числом степеней свободы $\nu_1 = m-1$ и $\nu_2 = N-m$.

В оригинальном критерии Левене предусмотрено использование только выборочных средних. В **ISW** возможно исследование данного случая. Brown и Forsythe расширили критерий Левене на случай использования выборочных медиан и усеченного среднего ($Z_{ij} = \left| X_{ij} - \tilde{X}_{i\bullet} \right|$, где $\tilde{X}_{i\bullet}$ − медиана в i-й выборке; $Z_{ij} = \left| X_{ij} - \overline{X}'_{i\bullet} \right|$, где $\overline{X}'_{i\bullet}$ − усеченное среднее в i-й выборке). В этих случаях критерий становится устойчивей к нарушению предположений о нормальности, однако, надо полагать, что распределение статистики должно отличаться от F_{ν_1,ν_2}-распределения.

Порядок выполнения

В процессе выполнения работы использовать программу **ISW.**

1. Исследовать сходимость распределений статистики (15.1) критерия Бартлетта к предельному χ^2_{m-1}-распределению в зависимости от объемов выборок.
2. Исследовать влияние нарушений предположения о нормальности наблюдаемых законов на распределение статистики критерия Бартлетта. Для этого смоделировать распределение статистики при тех же объемах выборок: при различных симметричных законах наблюдаемых случайных величин с различной “тяжестью хвостов распределений”; при асимметричных законах.
3. Исследовать (оценить) мощность критерия Бартлетта относительно заданных конкурирующих гипотез и заданных объемов выборок в случае нормального закона. Пусть конкурирующая гипотеза предполагает, что одна из выборок, например, выборка с номером m имеет некоторую другую дисперсию. Остальные $m-1$ выборки принадлежат нормальному закону с $\sigma = \sigma_0$ ($H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_m^2 = \sigma_0^2$). Для расчета значений мощности критерия рассмотреть одну из следующих конкурирующих гипотез: H_1^1:

$\sigma_m = 1.05\sigma_0$; H_1^2: $\sigma_m = 1.1\sigma_0$ и H_1^3: $\sigma_m = 1.2\sigma_0$ (выбор по заданию преподавателя).

4. Исследовать распределения статистики (15.2) критерия Кокрена в зависимости от объемов выборок при извлечении выборок из нормальной совокупности.
5. Аналогично п.2 исследовать распределения статистики (15.2) критерия Кокрена.
6. Для тех же конкурирующих гипотез и объемов выборок (см. п.3) оценить мощность критерия Кокрена.
7. Исследовать распределения и устойчивость критерия Фишера со статистикой (15.3).
8. Исследовать распределения и устойчивость критерия Хартли со статистикой (15.4).
9. Исследовать распределения и устойчивость критерия Левене со статистикой (15.5).

Содержание отчета. Цель работы, графики (наиболее наглядно характеризующие поведение распределений статистик), результаты статистического анализа распределений, оценки мощности (таблицы), выводы.

Варианты для выбора законов совпадают с вариантами из лабораторной работы № 13.

Контрольные вопросы

1. Виды статистик, используемых в критериях проверки гипотез об однородности дисперсий?
2. Распределения статистик каких критериев проверки гипотез об однородности дисперсий слабо зависят от объемов выборок?

[Назад][Вперед] [Содержание]

Литература

1. Лемешко Б.Ю., Постовалов С.Н. Компьютерные технологии анализа данных и исследования статистических закономерностей: Учебное пособие. – Новосибирск: Изд-во НГТУ, 2004. – 119 с.
2. Ермаков С.М., Михайлов Г.А. Статистическое моделирование. – Москва: Наука, 1982. – 296 с.
3. Хьюбер П. Робастность в статистике. М.: Мир, 1984. – 303 с.
4. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. - М.: Наука, 1983. - 416 с.
5. Денисов В.И., Лемешко Б.Ю., Постовалов С.Н. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим.

Методические рекомендации. Часть I. Критерии типа χ^2. - Новосибирск: Изд-во НГТУ, 1998. - 126 с.
6. Мирвалиев М., Никулин М.С. Критерии согласия типа хи-квадрат / Заводская лаборатория. 1992. Т. 58. № 3. С.52-58.
7. Лемешко Б.Ю., Постовалов С.Н. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Методические рекомендации. Часть II. Непараметрические критерии. - Новосибирск: Изд-во НГТУ, 1999. - 85 с.
8. Демиденко Е.З. Линейная и нелинейная регрессия. – М.: Финансы и статистика, 1981. – 302 с.

[Назад][В начало] [Содержание]

Приложение. Законы распределения наблюдаемых случайных величин

Закон распределения, параметры, область определения	Функция плотности $f(x,\theta)$
Экспоненциальный, $\theta_0 > 0$, $x \in 0,+\infty$	$\frac{1}{\theta_0} e^{-x/\theta_0}$
Полунормальный, $\theta_0 > 0$, $x \in 0,+\infty$	$\frac{2}{\theta_0\sqrt{2\pi}} e^{-x^2/2\theta_0^2}$
Рэлея, $\theta_0 > 0$, $x \in 0,+\infty$	$\frac{x}{\theta_0^2} e^{-x^2/2\theta_0^2}$
Максвелла, $\theta_0 > 0$, $x \in 0,+\infty$	$\frac{2x^2}{\theta_0^3\sqrt{2\pi}} e^{-x^2/2\theta_0^2}$
Лапласа, $\theta_1 \in R$, $\theta_0 > 0$, $x \in -\infty,+\infty$	$\frac{1}{2\theta_0} e^{-\lvert x-\theta_1\rvert/\theta_0}$
Нормальный, $\theta_1 \in R$, $\theta_0 > 0$, $x \in -\infty,+\infty$	$\frac{1}{\theta_0\sqrt{2\pi}} e^{-\frac{(x-\theta_1)^2}{2\theta_0^2}}$
Логнормальный, $\theta_1 \in R$, $\theta_0 > 0$, $x \in 0,+\infty$	$\frac{1}{x\theta_0\sqrt{2\pi}} e^{-(\ln x-\theta_1)^2/2\theta_0^2}$
Коши, $\theta_1 \in R$, $\theta_0 > 0$, $x \in -\infty,+\infty$	$\frac{\theta_0}{\pi[\theta_0^2+(x-\theta_1)^2]}$

Логистический, $\theta_0 \in R$, $\theta_1 > 0$, $x \in -\infty, +\infty$	$\frac{\pi}{\theta_0\sqrt{3}}\exp\left\{-\frac{\pi(x-\theta_1)}{\theta_0\sqrt{3}}\right\} \Big/ \left[1+\exp\left\{-\frac{\pi(x-\theta_1)}{\theta_0\sqrt{3}}\right\}\right]^2$
Экстремальных значений (максимум), $\theta_0 > 0$, $\theta_1 \in R$, $x \in -\infty, +\infty$	$\frac{1}{\theta_0}\exp\left\{-\frac{x-\theta_1}{\theta_0} - \exp\left(-\frac{x-\theta_1}{\theta_0}\right)\right\}$
Экстремальных значений (минимум), $\theta_0 > 0$, $\theta_1 \in R$, $x \in -\infty, +\infty$	$\frac{1}{\theta_0}\exp\left\{\frac{x-\theta_1}{\theta_0} - \exp\left(\frac{x-\theta_1}{\theta_0}\right)\right\}$
Вейбулла-Гнеденко, $\theta_0 > 0$, $\theta_1 > 0$, $x \in 0, +\infty$	$\frac{\theta_0 x^{\theta_0 - 1}}{\theta_1^{\theta_0}}\exp\left\{-\left(\frac{x}{\theta_1}\right)^{\theta_0}\right\}$
Гамма, $\theta_0 > 0$, $\theta_1 > 0$, $\theta_2 \in R$, $x \in \theta_2, +\infty$	$\frac{1}{\theta_1^{\theta_0}\Gamma(\theta_0)}(x-\theta_2)^{\theta_0-1}e^{-(x-\theta_2)/\theta_1}$
Бета-распределение I-го рода, $\theta_0 > 0$, $\theta_1 > 0$, $\theta_2 > 0$, $x \in 0, \theta_2$	$f(x) = \frac{\Gamma(\theta_0+\theta_1)}{\lambda^{\theta_0+\theta_1-1}\Gamma(\theta_0)\Gamma(\theta_1)}x^{\theta_0-1}(\theta_2-x)^{\theta_1-1}$
Бета-распределение II -го рода, $\theta_0 > 0$, $\theta_2 > 0$, $\theta_3 > 0$, $x \in \theta_1, +\infty$	$\frac{1}{\theta_0 \cdot \mathrm{B}(\theta_2,\theta_3)}\left(\frac{x-\theta_1}{\theta_0}\right)^{\theta_2-1}\left(1+\frac{x-\theta_1}{\theta_0}\right)^{-\theta_2-\theta_3}$
Sb-Джонсона, $\theta_1 \in R$, $\theta_3 \in R$, $\theta_1 > 0$, $\theta_2 > 0$, $x \in \theta_3, \theta_3+\theta_2$	$\frac{\theta_1\theta_2}{(x-\theta_3)(\theta_2+\theta_3-x)}\exp\left\{-\frac{1}{2}\left[\theta_0 - \theta_1 \ln\frac{x-\theta_3}{\theta_2+\theta_3-x}\right]^2\right\}$
Двустороннее экспоненциальное, $\theta_1 \in R$, $\theta_0 > 0$, $\theta_2 > 0$, $x \in -\infty, +\infty$	$f(x) = \frac{\theta_2}{2\theta_1\Gamma(1/\theta_2)}e^{-\left\lvert\frac{x-\theta_1}{\theta_0}\right\rvert^{\theta_2}}$