Министерство образования и науки Российской Федерации

НОВОСИБИРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ
УНИВЕРСИТЕТ

Компьютерные технологии анализа данных и исследования статистических закономерностей

Методические указания
к выполнению лабораторных работ
для студентов I-го курса магистратуры ФПМИ (**семестр 2**)
по направлению 01.04.02
дневного отделения

Новосибирск, 2017

Методические указания предназначены для студентов, выполняющих лабораторные работы по курсу "Компьютерные технологии анализа данных и исследования статистических закономерностей" (направление 010400.68). 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*,
канд. техн. наук, доц. *Е.В. Чимитова*

Работа подготовлена на кафедре
прикладной математики

Содержание

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://postovalov.net. Структурированные указания на материалы по различным разделам курса размещены по адресу http://ami.nstu.ru/~headrd/seminar/start.htm.

Лабораторная работа № 1. 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 $D_{m,n}$ рекомендуется вычислять 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

$$S_C = \sqrt{\frac{mn}{m+n}} D_{m,n} \tag{1}$$

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 работе Леманом (Lehmann E.L.) и подробно исследован в Розенблаттом (Rosenblatt M.). Статистика критерия имеет вид

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

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

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

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 $a1(t)$:

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

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

В отличие от статистики критерия Смирнова распределение статистики T быстро сходится к предельному закону $a1(T)$.

3. Критерий однородности Андерсона–Дарлинга. Статистика двухвыборочного критерия Андерсона–Дарлинга (критерия однородности Андерсона–Дарлинга-Петита) определяется выражением

$$A^2 = \frac{mn}{m+n}\int_{-\infty}^{\infty}\frac{[G_m(x) - F_n(x)]^2}{(1 - H_{m+n}(x))H_{m+n}(x)}dH_{m+n}(x). \tag{3}$$

Для выборок непрерывных случайных величин выражение для этой статистики принимает простой вид

$$A^2 = \frac{1}{mn}\sum_{i=1}^{m+n-1}\frac{\left(M_i(m+n) - mi\right)^2}{i(m+n-i)}, \tag{4}$$

где M_i – число элементов первой выборки, меньших или равных i-му элементу вариационного ряда объединенной выборки.

Предельным распределением статистики (4) при справедливости проверяемой гипотезы H_0 является то же самое распределение $a2(t)$, которое является предельным для статистики критерия согласия Андерсона–Дарлинга.

4. Многовыборочный критерий Андерсона–Дарлинга. Многовыборочный вариант критерия однородности Андерсона–Дарлинга для k выборок строится следующим образом. Пусть x_{ij} j-е наблюдение i-й выборки $j=\overline{1,n_i}$, $i=\overline{1,k}$. Предположим, что i-й выборке соответствует непрерывная функция распределения $F_i(x)$. Необходимо проверить гипотезу вида $H_0: F_1(x)=F_2(x)=\ldots=F_k(x)$ без указания общего для них закона распределения. Обозначим эмпирическую функцию распределения, соответствующую i-й выборке, как $F_{in_i}(x)$, а эмпирическую функцию распределения, соответствующую объединённой выборке объёмом $n=\sum_{i=1}^{k} n_i$, как $H_n(x)$. Статистика k-выборочного критерия Андерсона–Дарлинга определяется выражением

$$A_{kn}^2=\sum_{i=1}^{k} n_i \int_{B_n} \frac{[F_{in_i}(x)-H_n(x)]^2}{\left(1-H_n(x)\right)H_n(x)} dH_n(x), \tag{5}$$

где $B_n=\{x\in R: H_n(x)<1\}$. Для $k=2$ соотношение (5) сводится к (4). В предположении о непрерывности $F_i(x)$, упорядочив объединённую выборку $Z_1\le Z_2\le\ldots\le Z_n$, непосредственно из (5) можно получить простое выражение для вычисления статистики:

$$A_{kn}^2=\frac{1}{n}\sum_{i=1}^{k}\frac{1}{n_i}\sum_{j=1}^{n-1}\frac{\left(nM_{ij}-jn_i\right)^2}{j(n-j)}, \tag{6}$$

где M_{ij} – число элементов в i-й выборке, которые не больше чем Z_j. Проверяемая гипотеза H_0 отклоняется при больших значениях статистики (6).

Вместо статистики (6) используется статистика вида:

$$T_{kn}=\frac{A_{kn}^2-(k-1)}{\sqrt{D[A_{kn}^2]}}. \tag{7}$$

Дисперсия статистики A_{kn}^2 определяется выражением

$$D\left[A_{kn}^2\right]=\frac{an^3+bn^2+cn+d}{(n-1)(n-2)(n-3)} \tag{8}$$

при

$a=(4g-6)(k-1)+(10-6g)H$,

$b=(2g-4)k^2+8hk+(2g-14h-4)H-8h+4g-6$,

$c=(6h+2g-2)k^2+(4h-4g+6)k+(2h-6)H+4h$,

$d=(2h+6)k^2-4hk$,

где

$$H=\sum_{i=1}^{k}\frac{1}{n_i},\quad h=\sum_{i=1}^{n-1}\frac{1}{i},\quad g=\sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\frac{1}{(n-i)j}.$$

Зависимость предельных распределений статистики (2.11) от числа сравниваемых выборок k иллюстрирует рис. 2.20. С ростом числа сравниваемых выборок это распределение медленно сходится к стандартному нормальному закону. Сходимость распределений статистики к предельному с ростом объёмов выборок (при равных n_i) для $k=3$ демонстрирует рис. 2.21.

Более подробно об описании критериев однородности законов см. в [4].

Указание: При моделировании выборок, соответствующих различным законам задавайте ***различные*** начальные значения для ГСЧ!

Порядок выполнения

Используя программу ISW, исследовать распределения статистик критериев проверки однородности Смирнова, Лемана–Розенблатта и Андерсона–Дарлинга.

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.
3. Исследовать сходимость к предельным распределениям (см. [4]) распределений статистики (7) k-выборочного критерия Андерсона–Дарлинга при заданном k. При k=2 сравнить мощность критерия с двухвыборочными.

Варианты заданий.

№	$F(x,\theta)$	$F_1(x,\theta_1)$
1	Нормальное (0, 1)	Ообобщённый нормальный закон с параметром формы (3)
2	Нормальное (0, 2)	Коши (0,1)
3	Нормальное (0, 1)	Ообобщённый нормальный закон с параметром формы (1)
4	Экспоненциальное (0, 1)	Гамма-распределение с параметром формы (2)

5	Нормальное (0, 1)	Su-Джонсона (0, 1)
6	Экспоненциальное (0, 1)	Вейбулла (2,1)
7	Логистическое (0, 1)	Нормальное (0,2)
8	Нормальное (0, 1)	Ообобщённый нормальный закон с параметром формы (5)
9	Логарифмически нормальное (0,1)	Вейбулла (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?

Лабораторная работа № 2. 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 ним, например, 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

$$z = \frac{\bar{x}_1 - \bar{x}_2 - \mu_1 + \mu_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} , \qquad (1)$$

где $\bar{x}_i = \dfrac{1}{n_i} \sum_{j=1}^{n_i} x_j$, 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:

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

где n_i – объем i-й выборки,

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

В случае принадлежности выборок нормальному закону эта статистика при справедливости гипотезы H_0 подчиняется 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:

$$t = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{\sqrt{\left[\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right]}}, \qquad (3)$$

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

В случае нормального закона и справедливости гипотезы H_0 статистика (3) подчиняется распределению 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 $n_1 = n_2 = n$ –

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

а число степеней свободы распределения Стьюдента –

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

4. F-критерий однородности средних. В случае справедливости гипотезы о постоянстве (о равенстве) дисперсий с помощью F-критерия можно проверять гипотезу об однородности математических ожиданий по k выборкам.

Пусть у нас имеется k выборок объемом n. Сумма квадратов отклонений по всем выборкам определяется соотношением

$$Q_{kn} = \sum_{i=1}^{k}\sum_{j=1}^{n}\left(x_{ij} - \overline{x}_{kn}\right)^2,$$

где

$$\overline{x}_{kn} = \frac{1}{kn}\sum_{i=1}^{k}\sum_{j=1}^{n} x_{ij} = \overline{\overline{x}}_k$$

– среднее по всем выборкам. Общая сумма Q_{kn} раскладывается на два компонента

$$Q_{kn} = Q_1 + Q_2,$$

$$Q_1 = n\sum_{i=1}^{k}\left(\overline{x}_{in} - \overline{\overline{x}}_k\right)^2 = n\sum_{i=1}^{k}\left(\overline{x}_{in}^2 - k\overline{\overline{x}}_k\right)^2,$$

$$Q_2 = \sum_{i=1}^{k}\sum_{j=1}^{n}\left(x_{ij} - \overline{x}_{in}\right)^2 = (n-1)\sum_{i=1}^{k} s_{in}^2.$$

Компонент Q_1, например, в задачах контроля качества является мерой различия в уровнях настройки между k выборками, в то время как Q_2 определяет различие в уровнях настройки внутри этих k выборок.

Для проверки гипотезы используется критерий со статистикой

$$F = \frac{Q_1/(k-1)}{Q_2/[k(n-1)]}. \qquad (4)$$

Если все выборки извлекаются из нормальной генеральной совокупности, то при справедливости гипотезы H_0 статистка (4) подчиняется F_{ν_1,ν_2}-распределению Фишера со степенями свободы $\nu_1 = k-1$ и $\nu_2 = k(n-1)$. Критерий правосторонний. Проверяемая гипотеза отклоняется при больших значениях статистики (4).

5. 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 $m+n$ значений объединенной выборки, определяют сумму рангов R_1, соответствующую элементам первой выборки, и сумму рангов второй R_2. Вычисляются

$$U_1 = mn + \frac{m(m-1)}{2} - R_1,$$

$$U_2 = mn + \frac{n(n-1)}{2} - R_2.$$

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

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

$$\tilde{z} = \frac{\left| U - \frac{mn}{2} \right|}{\sqrt{\frac{mn(m+n+1)}{12}}}, \qquad (5)$$

распределение которой приближенно описывается стандартным нормальным законом.

6. 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

$$H = \left[\frac{12}{n(n+1)}\right]\left[\sum_{i=1}^{k} \frac{R_i^2}{n_i}\right] - 3(n-1). \qquad (6).$$

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

7. Критерий Ван дер Вардена. Критерий предназначен для анализа 2-х выборок. Статистика непараметрического критерия Ван дер Вардена вычисляется в соответствии с выражением:

$$V = \sum_{i=1}^{n_2} u_{R_i/(n_1+n_2+1)}, \qquad (7)$$

где u_γ – γ-квантиль стандартного нормального закона, R_i, $i = \overline{1,n_2}$ – ранг i-го наблюдения, например, как в (3.10), второй выборки в общем вариационном ряду объединенной выборки из $n_1 + n_2$ наблюдений.

Считается, что при $n_1, n_2 \geq 20$ распределение статистики (7) удовлетворительно описывается нормальным законом с нулевым математическим ожиданием и дисперсией

$$D[V]=\frac{n_1 n_2}{(n_1+n_2)(n_1+n_2-1)}\sum_{i=1}^{n_1+n_2}u^2_{i/(n_1+n_2+1)}.$$

Нормализованная статистика

$$V^*=\frac{V}{\sqrt{D[V]}} \tag{8}$$

должна подчиняться стандартному нормальному закону.

Критерий двусторонний, проверяемая гипотеза отклоняется при больших по модулю значениях статистики (8).

8. Многовыборочный критерий Ван дер Вардена. Статистика многовыборочного критерия Ван дер Вардена для проверки гипотезы о равенстве средних по k выборкам имеет вид

$$T=(n-1)\sum_{i=1}^{k}\frac{1}{n_i}\left(\sum_{j=1}^{n_i}u_{R_{ij}/(n+1)}\right)^2 \Bigg/ \sum_{i=1}^{n}u^2_{i/(n+1)}, \tag{9}$$

где $n=\sum_{i=1}^{k}n_i$, u_γ – γ-квантиль стандартного нормального закона, R_{ij}, $j=\overline{1,n_i}$, – ранг j-го элемента i-й выборки в вариационном ряду объединённой выборки объёма n.

При справедливости проверяемой гипотезы H_0 статистика (9) хорошо описывается χ^2_{k-1}-распределением. Отклонением распределения статистики от χ^2_{k-1}-распределения можно пренебречь при $n_i>30$.

Более подробно об описании критериев однородности законов см. в [4].

Указание: При моделировании выборок, соответствующих различным законам задавайте ***различные*** начальные значения для ГСЧ!

Порядок выполнения.

В процессе выполнения работы использовать программу **isw.exe.**

1. По указанию преподавателя исследовать распределения статистик (1), (2) или (3) и(4) в условиях принадлежности наблюдений нормальным законам с заданными параметрами при некоторых объемах выборок. 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 (5) критерия Манна-Уитни к стандартному нормальному закону с ростом объемов выборок. Убедиться, что распределения статистики (5) не зависят от вида закона, из которого извлекаются выборки.
5. Исследовать сходимость распределения статистики (6) критерия к соответствующему χ^2_{k-1}-распределению в зависимости от объемов выборок. Проверить, что критерий Краскела–Уаллиса действительно является непараметрическим, и распределения статистики (6) 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.
7. Исследовать сходимость распределений статистик (8) и (9) критериев Ван дер Вардена к асмптотическим.
8. Сравнить мощность критериев Ван дер Вардена с другими параметрическими и непараметрическими.

Варианты для выбора законов.

№	$F(x,\theta)$	$F_1(x,\theta_1)$
1	Нормальное (0, 1)	Ообобщённый нормальный закон с параметром формы (1)
2	Нормальное (0.1, 1)	Ообобщённый нормальный закон с параметром формы (3)
3	Нормальное (0.1, 2)	Ообобщённый нормальный закон с параметром формы (5)
4	Нормальное (0.5, 1)	Коши (0,1)
5	Нормальное (0.5, 2)	Ообобщённый нормальный закон с параметром формы (0.5)
6	Нормальное (0, 1.5)	Ообобщённый нормальный закон с параметром формы (0.2)
7	Нормальное (0, 2)	Гамма-распределение с параметром формы (2)
8	Нормальное (0, 3)	Распределение минимального

		значения
9	Экспоненциальное (0, 1)	Распределение максимального значения
10	Нормальное (1, 3)	Показательное распределение

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?

Лабораторная работа № 3. Исследование распределений статистик критериев проверки однородности дисперсий. Часть 1

Цель работы. Исследовать распределения статистик и мощность ряда параметрических 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 применяются различные параметрические (и непараметрические) критерии.

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

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

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 $S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} \left(X_{ji} - \overline{X}_i \right)^2$, 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_{m-1}^2-распределению.

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

$$Q = \frac{S_{\max}^2}{S_1^2 + S_2^2 + \cdots + S_m^2} , \tag{2}$$

где $S_{\max}^2 = \max\left(S_1^2, S_2^2, \ldots, S_m^2\right)$, 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

$$F = \frac{s_1^2}{s_2^2}. \tag{3}$$

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

$$F = \frac{s_{\max}^2}{s_{\min}^2}. \tag{4}$$

Степенями свободы для распределения статистики являются число выборок $\nu_1 = m$ и $\nu_2 = n - 1$. В литературе приводятся таблицы процентных точек для статистики (3).

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:

$$W = \frac{N - m}{m - 1} \frac{\sum_{i=1}^{m} n_i \left(\bar{Z}_{i\bullet} - \bar{Z}_{\bullet\bullet}\right)^2}{\sum_{i=1}^{m} \sum_{j=1}^{n_i} \left(Z_{ij} - \bar{Z}_{i\bullet}\right)^2}, \tag{5}$$

где Z_{ij} определяется соотношением $Z_{ij} = \left| X_{ij} - \bar{X}_{i\bullet} \right|$, в котором $\bar{X}_{i\bullet}$ − есть среднее в i-й выборке. $\bar{Z}_{i\bullet}$ − есть среднее Z_{ij} по i-й выборке, $\bar{Z}_{\bullet\bullet}$ − есть 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 $\tilde{X}_{i\bullet}$ − есть медиана в i-й выборке; $Z_{ij} = \left| X_{ij} - \overline{X}'_{i\bullet} \right|$ где $\overline{X}'_{i\bullet}$ − есть усеченное среднее в i-й выборке). В этих случаях критерий становится устойчивей нарушению предположений о нормальности, однако, надо полагать, распределение статистики должно отличаться от F_{ν_1,ν_2}-распределения.

Указание: При моделировании выборок, соответствующих различным законам задавайте ***различные*** начальные значения для ГСЧ!

Порядок выполнения.

В процессе выполнения работы использовать программу **isw.exe.**

1. Исследовать сходимость распределений статистики (5.1) 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 варианта $F_1(x,\theta_1)$).
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 (2) критерия Кокрена в зависимости от объемов выборок при извлечении выборок из нормальной совокупности.
5. Аналогично п.2 исследовать распределения статистики (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 (3).
8. Исследовать распределения и устойчивость критерия Хартли со статистикой (4).

9. Исследовать распределения и устойчивость критерия Левене со статистикой (5).

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

Варианты заданий

№	$F(x,\theta)$	$F_1(x,\theta_1)$
1	Нормальное (0, 1)	Обобщённый нормальный закон с параметром формы (1)
2	Нормальное (0.1, 1)	Обобщённый нормальный закон с параметром формы (3)
3	Нормальное (0.1, 2)	Обобщённый нормальный закон с параметром формы (5)
4	Нормальное (0.5, 1)	Коши (0,1)
5	Нормальное (0.5, 2)	Обобщённый нормальный закон с параметром формы (0.5)
6	Нормальное (0, 1.5)	Обобщённый нормальный закон с параметром формы (0.2)
7	Нормальное (0, 2)	Гамма-распределение с параметром формы (2)
8	Нормальное (0, 3)	Распределение минимального значения
9	Экспоненциальное (0, 1)	Распределение максимального значения
10	Нормальное (1, 3)	Показательное распределение

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?

Лабораторная работа № 4. Исследование распределений статистик критериев проверки однородности дисперсий. Часть 2

Цель работы. Исследовать распределения статистик и мощность некоторых параметрических и непараметрических 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 применяются различные параметрические (и непараметрические) критерии.

1. Критерий Неймана-Пирсона. Статистика критерия Неймана–Пирсона (критерия отношения правдоподобия) определяется отношением арифметического среднего всех оценок дисперсий s_i^2 к их геометрическому среднему:

$$h = \frac{1}{k}\sum_{i=1}^{k} s_i^2 \Bigg/ \left(\prod_{i=1}^{k} s_i^2\right)^{\frac{1}{k}} , \qquad (1)$$

где k – количество выборок, $s_i^2 = \dfrac{1}{n_i - 1}\sum_{j=1}^{n_i}\left(x_{ij} - \bar{x}_i\right)^2$ – оценки выборочных дисперсий, $\bar{x}_i = \dfrac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$ – выборочное среднее значение, x_{ij} – j-й элемент i-й выборки. Обычно предполагается, что $n_1 = n_2 = \ldots = n_k = n$. Критерий правосторонний. Проверяемая гипотеза H_0 отклоняется при больших значениях статистики (1), когда $h > h_{1-\alpha}$.

2. Критерий О`Брайена. При формировании статистики критерия каждый j-й элемент i-й выборки x_{ij} преобразуется в соответствии с формулой

$$V_{ij} = \frac{(n_i - 1.5)n_i\left(x_{ij} - \bar{x}_i\right)^2 - 0.5 s_i^2 (n_i - 1)}{(n_i - 1)(n_i - 2)} ,$$

где n_i – объём, $\bar{x}_i$ – среднее значение, s_i^2 – оценка дисперсии i-й выборки.

Статистика критерия имеет вид

$$V=\frac{\frac{1}{k-1}\sum_{i=1}^{k}n_i\left(\overline{V_i}-\overline{\overline{V_i}}\right)^2}{\frac{1}{N-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(V_{ij}-\overline{V_i}\right)^2},\qquad(2)$$

где $\overline{V_i}=\frac{1}{n_i}\sum_{j=1}^{n_i}V_{ij}$, $\overline{\overline{V_i}}=\frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i}V_{ij}$, $N=\sum_{i=1}^{k}n_i$.

Критерий правосторонний, и проверяемая гипотеза H_0 отклоняется при больших значениях статистики (2). Предельным распределением статистики критерия О'Брайена при справедливости H_0 является $F_{k-1,N-k}$-распределение Фишера с $k-1$ и $N-k$ степенями свободы.

3. Критерий Линка. Критерий Линка (критерий отношения размахов) является аналогом критерия Фишера и используется только при анализе 2-х выборок ($m=2$). Статистика критерия имеет вид:

$$F^*=\frac{\omega_{n_1}}{\omega_{n_2}},\qquad(3)$$

где $\omega_{n_1}=x_{1,\max}-x_{1,\min}$, $\omega_{n_1}=x_{2,\max}-x_{2,\min}$ – размахи, а $x_{1,\max}$, $x_{2,\max}$, $x_{1,\min}$, $x_{2,\min}$ – максимальные и минимальные элементы сравниваемых выборок.

Критерий двусторонний. Проверяемая гипотеза отклоняется с уровнем значимости α, если $F^*>F^*_{1-\alpha/2}$ или $F^*<F^*_{\alpha/2}$, где $F^*_{1-\alpha/2}$ и $F^*_{\alpha/2}$ – верхнее и нижнее критические значения статистики.

4. Критерий Ньюмана. Статистика критерия Ньюмана (стьюдентизированного размаха) имеет вид:

$$q=\frac{\omega_{n_1}}{s_{n_2}},\qquad(4)$$

где $\omega_{n_1}=x_{1,\max}-x_{1,\min}$, $s_{n_2}=\sqrt{\frac{1}{n_2-1}\sum_{i=1}^{n_2}\left(x_{2i}-\bar{x}_2\right)^2}$.

Как и предыдущий, этот критерий также является двусторонним. Проверяемая гипотеза H_0 о равенстве дисперсий отклоняется, если $q<q_{\alpha/2}$ или $q>q_{1-\alpha/2}$, где $q_{\alpha/2}$ и $q_{1-\alpha/2}$ – нижнее и верхнее критические значения статистики при заданном уровне значимости α.

5. Критерий Блиса–Кокрена–Тьюки. Критерий предложен в качестве аналога критерия Кокрена. Статистика имеет вид

$$c=\frac{\max\limits_{1\le i\le k}\omega_i}{\sum_{i=1}^{k}\omega_i},\qquad(5)$$

где k – количество сравниваемых выборок, $\omega_i=\max\limits_{1\le j\le n_i}x_{ij}-\min\limits_{1\le j\le n_i}x_{ij}$ – размах i-й выборки.

Критерий правосторонний. Если статистика $c > c_{1-\alpha}$, где $c_{1-\alpha}$ – верхнее критическое значение при заданном уровне значимости α, то проверяемая гипотеза H_0 о равенстве дисперсий отклоняется.

6. Критерий Кадуэлла–Лесли–Брауна. Этот критерий был предложен в качестве аналога критерия Хартли с заменой в статистике отношений оценок дисперсий на отношения размахов

$$K = \frac{\max\limits_{1\le i\le k} \omega_i}{\min\limits_{1\le i\le k} \omega_i}, \tag{6}$$

где k – количество выборок, ω_i – размах i-й выборки.

Критерий правосторонний. При $K > K_{1-\alpha}$, где $K_{1-\alpha}$ – верхнее критическое значение статистики при заданном уровне значимости α, проверяемая гипотеза H_0 отклоняется.

7. Z–критерий Оверолла–Вудворда. Статистика критерия имеет вид:

$$Z = \frac{1}{k-1}\sum_{i=1}^{k} Z_i^2, \tag{7}$$

где $Z_i = \sqrt{\frac{c_i\left(n_i-1\right)s_i^2}{MSE}} - \sqrt{c_i\left(n_i-1\right)-\frac{c_i}{2}}$, $MSE = \frac{1}{N-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(x_{ij}-\bar{x}_i\right)^2$, m – количество выборок, $c_i = 2 + 1/n_i$, n_i – размер i-й выборки, s_i^2 – несмещенная оценка дисперсии i-й выборки, $N = \sum_{i=1}^{k} n_i$, x_{ij} – j-й элемент i-й выборки, $\bar{x}_i$ – среднее значение i-й выборки.

При справедливости проверяемой гипотезы H_0 об однородности дисперсий и принадлежности анализируемых выборок нормальному закону предельное распределение статистики (4.15) не зависит от размера выборки и подчиняется $F_{k-1,\infty}$-распределению Фишера.

8. Модифицированный Z–критерий. В целях построения критерия, более устойчивого к нарушению стандартного предположения о нормальности авторами предложена модификация Z–критерия, статистика которого отличается вычислением величин c_i:

$$c_i = 2.0\left[\frac{1}{K_i}\left(2.9+\frac{0.2}{n_i}\right)\right]^{\frac{1.6\left(n_i-1.8K_i+14.7\right)}{n_i}}, \tag{8}$$

где $K_i = \frac{1}{n_i-2}\sum_{j=1}^{n_i} G_{ij}^4$ – оценка коэффициента эксцесса i-й выборки, $G_{ij} = \left(x_{ij}-\bar{x}_i\right)\Big/\sqrt{\frac{n_i-1}{n_i}s_i^2}$.

9. Критерий Ансари–Бредли. Непараметрические аналоги критериев проверки однородности дисперсий предназначены для проверки гипотез о принадлежности двух выборок с объемами n_1 и n_2 общей генеральной совокупности с одинаковыми характеристиками рассеяния. При этом, как правило, предполагается равенство средних.

Статистика критерия Ансари–Бредли может быть вычислена следующим образом:

$$S = \sum_{i=1}^{n_1} \left\{ \frac{n_1 + n_2 + 1}{2} - \left| R_i - \frac{n_1 + n_2 + 1}{2} \right| \right\}, \tag{9}$$

где R_i – ранги элементов первой выборки в общем вариационном ряду.

Критерий двусторонний, проверяемая гипотеза не отклоняется при $S_{\alpha/2} < S < S_{1-\alpha/2}$. В случае принадлежности выборок случайных величин одному и тому же закону распределение $G\left(S \middle| H_0\right)$ статистики (9) при справедливости проверяемой гипотезы H_0 не зависит от вида этого закона.

Математическое ожидание и дисперсия статистики (9) имеют вид:

$$E[S] = \begin{cases} \dfrac{n_1(n_1 + n_2 + 2)}{4} & \text{при четном } (n_1 + n_2), \\[2ex] \dfrac{n_1(n_1 + n_2 + 1)^2}{4(n_1 + n_2)} & \text{при нечетном } (n_1 + n_2); \end{cases}$$

$$D[S] = \begin{cases} \dfrac{n_1 n_2 (n_1 + n_2 - 2)(n_1 + n_2 + 2)}{48(n_1 + n_2 - 1)} & \text{при четном } (n_1 + n_2), \\[2ex] \dfrac{n_1 n_2 (n_1 + n_2 + 1)\left[(n_1 + n_2)^2 + 3\right]}{48(n_1 + n_2)^2} & \text{при нечетном } (n_1 + n_2). \end{cases}$$

При объемах выборок $n_1,\ n_2 > 10$ дискретное распределение нормированной статистики

$$S^* = (S - E[S]) / \sqrt{D[S]} \tag{10}$$

достаточно хорошо приближается стандартным нормальным законом.

10. Критерий Муда. Статистика критерия имеет вид

$$M = \sum_{i=1}^{n_1} \left(R_i - \frac{n_1 + n_2 + 1}{2} \right)^2, \tag{11}$$

где R_i – ранги элементов первой выборки в общем вариационном ряду двух выборок. Критерий двусторонний, проверяемая гипотеза не отклоняется при $M_{\alpha/2} < M < M_{1-\alpha/2}$.

При $n_1, n_2 > 10$ распределение нормированной статистики

$$M^* = \frac{\left(M - E[M] + \frac{1}{2}\right)}{\sqrt{D[M]}}, \tag{12}$$

где

$$E[M] = \frac{n_1(n_1 + n_2 + 1)(n_1 + n_2 - 1)}{12},$$

$$D[M] = \frac{n_1 n_2 \left(n_1 + n_2 + 1\right)(n_1 + n_2 + 2)(n_1 + n_2 - 2)}{180},$$

хорошо приближается стандартным нормальным.

11. Критерий Сижела–Тьюки. Статистика критерия строится следующим образом. Вариационный ряд, построенный по объединенной выборке, $x_1 \le x_2 \le ... \le x_n$, где $n = n_1 + n_2$, преобразуется в последовательность вида

$$x_1, x_n, x_{n-1}, x_2, x_3, x_{n-2}, x_{n-3}, x_4, x_5, \dots,$$

т. е. оставшийся ряд «переворачивается» каждый раз после приписывания рангов паре крайних значений. В качестве статистики критерия используется сумма рангов элементов первой выборки.

Статистика критерия имеет вид

$$R = \sum_{i=1}^{n_1} R_i . \tag{13}$$

Проверяемая гипотеза не отклоняется при $R_{\alpha/2} < R < R_{1-\alpha/2}$.

При $n_1, n_2 > 10$ распределение нормированной статистики

$$R^* = (R - E[R]) / \sqrt{D[R]}, \tag{14}$$

где

$$E[R] = \frac{n_1(n_1 + n_2 + 1)}{2}, \quad D[R] = \frac{n_1 n_2 \left(n_1 + n_2 + 1\right)}{12},$$

достаточно хорошо приближается стандартным нормальным законом.

12. Критерий Клотца. Статистика критерия имеет вид

$$L = \sum_{i=1}^{n_1} u^2_{R_i/(n_1+n_2+1)}, \tag{15}$$

где n_1 и n_2 – объемы сравниваемых выборок; R_i – ранг i-го элемента первой выборки в общем упорядоченном по возрастанию ряду $(n_1 + n_2)$ значений объединенной выборки; u_γ – γ-квантиль стандартного нормального распределения. Значения $u^2_{i/(N+1)}$, называемые метками критерия, несложно вычислить.

Критерий двусторонний, гипотеза о равенстве параметров масштаба не отклоняется с достоверностью α, если $L_{\alpha/2} < L < L_{1-\alpha/2}$.

При $n_1, n_2 > 10$ нормализованная статистика

$$L^* = \frac{L - E[L]}{\sqrt{D[L]}}, \tag{16}$$

где

$$E[L] = \frac{n_1}{n_1 + n_2} \sum_{i=1}^{n_1+n_2} u^2_{\frac{i}{n_1+n_2+1}},$$

$$D[L] = \frac{n_1 n_2}{(n_1+n_2)(n_1+n_2-1)} \sum_{i=1}^{n_1+n_2} u^4_{\frac{i}{n_1+n_2+1}} - \frac{n_2}{n_1(n_1+n_2-1)} \left[\frac{n_1}{n_1+n_2} \sum_{i=1}^{n_1+n_2} u^2_{\frac{i}{n_1+n_2+1}} \right]^2,$$

хорошо аппроксимируется стандартным нормальным законом.

13. k-выборочный критерий Флайне–Киллина. Модифицированный непараметрический критерий Флайне–Киллина предназначен для проверки однородности дисперсий $k \ge 2$ выборок с объёмами n_i, $i = \overline{1,k}$. Статистика критерия формируется следующим образом.

По исходным выборкам вычисляются абсолютные значения $z_{ji} = \left| x_{ji} - \tilde{x}_i \right|$, где $\tilde{x}_i$ – выборочная медиана i-й выборки, $i = \overline{1,k}$. Далее строится вариационный ряд объединённой выборки z_{ji}, $j = \overline{1,n_i}$, $i = \overline{1,k}$. Для элементов i-й выборки на основании рангов R_{ji} её элементов z_{ji} в объединённой выборке строятся метки

$$a_{n,R_{ji}} = \Phi^{-1}\left(\frac{1 + R_{ji}/(n+1)}{2} \right), \ j = \overline{1,n_i},$$

где $n = \sum_{i=1}^{k} n_i$, находятся $\bar{A}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} a_{n,R_{ji}}$. Статистика критерия имеет вид:

$$\chi_0^2 = \frac{\sum_{i=1}^{k} n_i \left(\bar{A}_i - \bar{a} \right)^2}{V^2}, \tag{17}$$

где $\bar{a} = \frac{1}{n} \sum_{j=1}^{n} a_{n,j}$, $V^2 = \frac{1}{n-1} \sum_{j=1}^{n} \left(a_{n,j} - \bar{a} \right)^2$.

Асимптотическим распределением статистики (17) при справедливости H_0 и больших объёмах выборок является χ^2_{k-1}-распределение.

Указание: При моделировании выборок, соответствующих различным законам задавайте ***различные*** начальные значения для ГСЧ!

Порядок выполнения.

В процессе выполнения работы использовать программу **isw.exe.** В соответствии с заданным (четным или нечетным) вариантом набора критериев:

1. Исследовать зависимость распределений статистик от объема анализируемых выборок и, при наличии, сходимость распределений статистик к асимптотическим.
2. Для параметрических критериев исследовать влияние нарушений предположения о нормальности наблюдаемых законов на распределения статистик. Для чего смоделировать распределение статистики при тех же объемах выборок: при различных симметричных законах наблюдаемых случайных величин с различной “тяжестью хвостов распределений”; при асимметричных законах (3 варианта $F_1(x,\theta_1)$).
3. Для непараметрических критериев исследовать, как изменится распределение статистики критерия при справедливости проверяемой гипотезы о равенстве дисперсий, в случае принадлежности выборок различным законам (но с одинаковыми дисперсиями).

В случае принадлежности моделируемой выборке обобщённому нормальному закону с плотностью

$$De(\theta_2) = f\left(x;\theta_0,\theta_1,\theta_2\right) = \frac{\theta_2}{2\theta_1\Gamma(1/\theta_2)}\exp\left(-\left(\frac{|x-\theta_0|}{\theta_1}\right)^{\theta_2}\right),$$

для которого $E[x]=\theta_0$, $D[x]=\theta_1^2\dfrac{\Gamma(3/\theta_2)}{\Gamma(1/\theta_2)}$, при заданных дисперсии и параметре формы θ_2, требуемый параметр масштаба определится соотношением

$$\theta_1 = \sqrt{\frac{D[x]\Gamma(1/\theta_2)}{\Gamma(3/\theta_2)}}.$$

4. Исследовать (оценить) мощность рассматриваемых критериев относительно заданных конкурирующих гипотез и заданных объемов выборок в случае нормального закона. Рассматривать те же конкурирующие гипотезы, что и в предшествующей работе.
5. Сравнить мощность исследованных критериев с оценками мощности критериев, рассмотренных в предыдущей работе.
6. Исследовать и сравнить мощность k-выборочных критериев (из вашего набора критериев) при k =3.

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

Нечетный вариант: Неймана-Пирсона, Линка, Блиса–Кокрена–Тьюки, Z–критерий Оверолла–Вудворда, Ансари–Бредли, Муда, k-выборочный критерий Флайне–Киллина.
Четный вариант: О`Брайена, Ньюмана, Кадуэлла–Лесли–Брауна, модифицированный Z–критерий Оверолла–Вудворда, Сижела–Тьюки, Клотца, k-выборочный критерий Флайне–Киллина.

Варианты задания законов распределения $F(x,\theta)$ и $F_1(x,\theta_1)$ те же, что и в предыдущей работе по критериям однородности дисперсий.

№	$F(x,\theta)$	$F_1(x,\theta_1)$
1	Нормальное (0, 1)	Ообобщённый нормальный закон с параметром формы (1)
2	Нормальное (0.1, 1)	Ообобщённый нормальный закон с параметром формы (3)
3	Нормальное (0.1, 2)	Ообобщённый нормальный закон с параметром формы (5)
4	Нормальное (0.5, 1)	Коши (0,1)
5	Нормальное (0.5, 2)	Ообобщённый нормальный закон с параметром формы (0.5)
6	Нормальное (0, 1.5)	Ообобщённый нормальный закон с параметром формы (0.2)
7	Нормальное (0, 2)	Гамма-распределение с параметром формы (2)
8	Нормальное (0, 3)	Распределение минимального значения
9	Экспоненциальное (0, 1)	Распределение максимального значения
10	Нормальное (1, 3)	Показательное распределение

Контрольные вопросы

1. Виды статистик, используемых в критериях проверки гипотез об однородности дисперсий?
2. Распределения статистик каких параметрических критериев проверки гипотез об однородности дисперсий слабо зависят от объемов выборок?
3. Параметрические критерии, наиболее устойчивые к нарушению стандартного предположения о нормальности?
4. Наиболее предпочтительные параметрические критерии однородности дисперсий? В связи с чем?
5. Наиболее предпочтительные непараметрические критерии однородности дисперсий? Почему?

Лабораторная работа № 5. Исследование критериев проверки отклонения от равномерного закона. Часть 1.

Цель работы. Исследование распределений статистик критериев, используемых при проверке отклонения эмпирических распределений наблюдаемых величин от равномерного закона Исследование распределений статистик критериев Шермана, критериев Неймана-Бартона на нескольких полиномах Лежандра. Исследование и сравнение мощности критериев относительно близких конкурирующих гипотез.

Методические указания

1. **Критерий Шермана** предназначен для проверки гипотез о равномерности наблюдаемого закона. Статистика критерия имеет вид

$$\omega_n = \frac{1}{2}\sum_{i=1}^{n+1}\left|U_i - U_{i-1} - \frac{1}{n+1}\right|, \tag{1}$$

где U_i – элементы вариационного ряда $U_1 < U_2 < \ldots U_n$, построенного по проверяемой выборке $X_1, X_2, \ldots, X_n$ объёмом n, $U_0 = 0$, $U_{n+1} = 1$. Критерий правосторонний: гипотеза о равномерности отклоняется при больших значениях статистики ω_n.

При справедливости гипотезы о равномерности и больших объёмах выборок n статистика ω_n приближенно подчиняется нормальному распределению.

Нормализованная статистика имеет вид

$$\omega_n^* = \frac{\omega_n - E[\omega_n]}{\sqrt{D[\omega_n]}}, \tag{2}$$

$$E[\omega_n] = \left(\frac{n}{n+1}\right)^{n+1}, \; D[\omega_n] = \frac{2n^2 + n(n-1)^{n+2}}{(n+2)(n+1)^{n+2}} - \left(\frac{n}{n+1}\right)^{2n+2},$$

и её распределение при $n \to \infty$ сходится к стандартному нормальному закону

2. **Критерий Кимбелла** похож на критерий Шермана. Статистика критерия имеет вид

$$A = \sum_{i=1}^{n+1}\left(U_i - U_{i-1} - \frac{1}{n+1}\right)^2, \tag{3}$$

где, как и в критерии Шермана, U_i – элементы вариационного ряда, построенного по выборке $X_1, X_2, \ldots, X_n$ объёмом n, $U_0 = 0$, $U_{n+1} = 1$.

Проверяемая гипотеза о равномерности выборки отклоняется при больших значениях статистики.

3. Статистика **критерия Морана** имеет вид

$$M_n = -\sum_{i=1}^{n+1} \ln\left[(n+1)(U_i - U_{i-1})\right], \qquad (4)$$

где $U_0 = 0,\ U_{n+1} = 1$.

Критерий является правосторонним, т.е. гипотеза о равномерности отклоняется при больших значениях статистики. В программе ISW критерий носит название «Критерий равномерности Морана 2».

Существует модификация статистики

$$\tilde{M}_n = \frac{12(n+1)}{7n+8} M_n,$$

которая подчиняется χ^2–распределению с n степенями свободы.

4. Статистика **критерия Ченга-Спиринга** по наблюдаемо выбор**ке** имеет вид

$$W_p = \frac{\left[\left(U_n - U_1\right)\frac{n+1}{n-1}\right]^2}{\sum_{i=1}^{n}\left(U_i - \bar{U}\right)^2}, \qquad (5)$$

где разность $U_n - U_1$ представляет собой выборочный размах, а $\bar{U}$ – среднее выборки. Критерий двусторонний: проверяемая гипотеза отклоняется как при больших, так и при малых значениях статистики.

5. **Критерии Хегази-Грина** для проверки равномерности были разработаны в виде аналогов соответствующих критериев нормальности. Критерии используются в форме 2-х модификаций: статистики, использующие математическое ожидание порядковых статистик:

$$T_1 = \frac{1}{n}\sum_{i=1}^{n}\left|U_i - \frac{i}{n+1}\right|,\ T_2 = \frac{1}{n}\sum_{i=1}^{n}\left(U_i - \frac{i}{n+1}\right)^2 \qquad (6)$$

и статистики, в которых вместо математических ожиданий порядковых статистик, используются их модальные значения:

$$T_1^* = \frac{1}{n}\sum_{i=1}^{n}\left|U_i - \frac{i-1}{n-1}\right|,\ T_2^* = \frac{1}{n}\sum_{i=1}^{n}\left(U_i - \frac{i-1}{n-1}\right)^2. \qquad (7)$$

Критерии правосторонние. Проверяемая гипотеза о равномерности распределения случайных величин отклоняется при больших значениях статистик.

6. Статистика **критерия Фросини** имеет вид

$$B_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left| U_i - \frac{i-0.5}{n} \right|. \qquad (8)$$

Критерий правосторонний: гипотеза равномерности отклоняется при больших значениях статистики.

7. **Критерий Гринвуда–Кэсенберри–Миллера** представляет собой модификацию критерия равномерности, предложенного Гринвудом. Статистика критерия Гринвуда–Кэсенберри–Миллера имеет вид

$$Q = \sum_{i=1}^{n+1} \left(U_i - U_{i-1}\right)^2 + \sum_{i=1}^{n} \left(U_{i+1} - U_i\right)\left(U_i - U_{i-1}\right). \qquad (9)$$

Данная модификация по своим свойствам является предпочтительней оригинала. Критерий правосторонний: гипотеза равномерности отклоняется при больших значениях статистики.

8. Статистика **критерия Шварца** имеет вид

$$A_n^* = \frac{n}{2} \sum_{i=1}^{n} \left(\frac{U_{i+1} - U_{i-1}}{2} - \frac{1}{n} \right)^2, \qquad (10)$$

где крайние значения порядковых статистик отличаются от используемых в других критериях равномерности $U_0 = -U_1$, $U_{n+1} = 2 - U_n$. Критерий правосторонний: гипотеза о равномерности отвергается при больших значениях статистики.

9. **Критерии Неймана-Бартона**. Данное семейство критериев основано на отношении правдоподобия. По элементам U_i вариационного ряда, построенного по выборке $X_1, X_2, ..., X_n$, вычисляют величины

$$V_j = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \pi_j \left(U_i - 0.5\right),$$

где $\pi_j(y)$ – полиномы Лежандра, ортогональные на отрезке [0,1]. Как правило, ограничиваются использованием первых 4-х полиномов:

$$\pi_1(y) = 2\sqrt{3}y; \quad \pi_2(y) = \sqrt{5}\left(6y^2 - 0{,}5\right);$$

$$\pi_3(y) = \sqrt{7}\left(20y^3 - 3y\right); \quad \pi_4(y) = 3\left(70y^4 - 15y^2 + 0{,}375\right).$$

Статистики критерия имеет вид

$$N_K = \sum_{j=1}^{K} V_j^2. \qquad (11)$$

Проверяемая гипотеза о равномерности наблюдаемой выборки отклоняется при больших значениях статистик.

При $n > 20$ распределения статистик хорошо аппроксимируется χ^2 –распределениями с K степенями свободы.

Наиболее полная информация представлена в [1].

Порядок выполнения

Используя программу **ISW**, исследовать описанные выше критерии проверки отклонения наблюдаемых данных от равномерного закона.

1. Исследовать зависимость распределений статистики Шермана (1) от объема выборок в случае принадлежности наблюдений равномерном закону на интервале [0,1]. Оценить сходимость статистики к нормальному закону распределения. Оценить мощность критерия относительно конкурирующих законов в соответствии с вариантом задания при различных объемах выборок.
2. Исследовать распределения статистики критерия равномерности, указанного в варианте задания, при различных объемах выборок. Оценить близость получаемых эмпирических распределений статистик к «теоретическим» по процентным точкам таблиц, соответствующим данному критерию. Сделать вывод о возможной сходимости распределения статистики к предельному распределению.
3. Исследовать распределения статистик критерия Неймана-Бартона (11) при проверке равномерности для различных объемах выборок. Оценить сходимость статистики к χ^2 – распределению. Выбор количества полиномов, на которых основана статистика, осуществляется в соответствии с вариантом. Оценить мощность критерия.
4. Сравнить мощность всех рассмотренных критериев.

В качестве конкурирующих гипотез рассматриваются варианты бета-распределений 1-го рода.

Варианты заданий

№	Критерий равномерности	Конкурирующие гипотезы	Количество полиномов
1	Критерий Кимбелла (3)	$B_I(1.3,1.3)$ и $B_I(0.8,1)$	2
2	Критерий Морана (4)	$B_I(1,1.3)$ и $B_I(1.2,0.9)$	3
3	Критерий Ченга-Спиринга (5)	$B_I(1.2,1.3)$ и $B_I(1,0.6)$	2
4	Критерий Хегази-Грина T_1 (6)	$B_I(0.85,0.9)$ и $B_I(1.7,1)$	4
5	Критерий Хегази-Грина T_2 (6)	$B_I(0.7,1.3)$ и $B_I(1,1.3)$	3
6	Критерий Хегази-Грина T_1^* (7)	$B_I(1.4,1.2)$ и $B_I(1,0.7)$	2
7	Критерий Хегази-Грина T_2^* (7)	$B_I(1.4,1.4)$ и $B_I(0.8,0.6)$	4
8	Критерий Фросини (8)	$B_I(1.5,0.8)$ и $B_I(1,0.8)$	3
9	Критерий Гринвуда-Кэсенберри-Миллера (9)	$B_I(1,1.3)$ и $B_I(1.2,0.9)$	4
10	Критерий Шварца (10)	$B_I(1.2,1.3)$ и $B_I(1,0.6)$	2

Контрольные вопросы

1. В каких случаях критерий Шермана оказываются смещёнными?
2. Какие преимущества у критерия Неймана-Бартона относительно других критериев?
3. Перечислите недостатки рассмотренных критериев равномерности?

Лабораторная работа № 6. Исследование критериев проверки отклонения от равномерности закона. Часть 2.

Цель работы. Исследование распределений статистик критериев, используемых при проверке отклонения эмпирических распределений от равномерного закона. Исследование распределений статистик критериев, основанных на энтропийной оценке. Исследование и сравнение мощности критериев относительно близких конкурирующих гипотез.

Методические указания

Критерии, базирующиеся на оценках энтропии, представляют собой достаточно эффективные критерии проверки гипотез о принадлежности наблюдений равномерному закону. Для случайной величины X с функцией плотности $f(x)$ энтропия $H(f)$ этой величины, предложенная Шенноном, определяется соотношением

$$H(f) = -\int_{-\infty}^{\infty} \ln\left(f(x)\right) f(x)\,dx.$$

Используя то, что выражение для энтропии можно записать как

$$H(f) = \int_{0}^{1} \ln\left\{\frac{d}{dp} F^{-1}(p)\right\} dp,$$

была получена первая оценка величины энтропии.

1. **Критерий Дудевича-ван дер Мюлена** является аналогом первого энтропийного критерия Васичека для проверки нормальности. Статистика критерия Дудевича-ван дер Мюлена имеет вид

$$H(m,n) = -\frac{1}{n}\sum_{i=1}^{n} \ln\left\{\frac{n}{2m}\left(U_{i+m} - U_{i-m}\right)\right\}. \quad (1)$$

При вычислении статистики критерия в качестве размера окна m выбирается целое число $m \le \frac{n}{2}$, при этом, если $i+m \ge n$, то $U_{i+m} = U_n$, а если $i-m \le 1$, то $U_{i-m} = U_1$.

Критерий правосторонний, проверяемая гипотеза отклоняется при больших значениях статистики. Распределения статистик зависят от размера окна m, то есть для различных значений параметра m существуют свои критические значения.

2. В последние годы были предложены две **модификации энтропийного критерия**, применимые для проверки равномерности. Статистики этих модификаций имеют следующий вид:

$$HY_1(m,n) = -\frac{1}{n}\sum_{i=1}^{n}\ln\left(\frac{U_{i+m}-U_{i-m}}{\hat{F}(U_{i+m})-\hat{F}(U_{i-m})}\right), \tag{2}$$

$$HY_2(m,n) = -\sum_{i=1}^{n}\ln\left(\frac{U_{i+m}-U_{i-m}}{\hat{F}(U_{i+m})-\hat{F}(U_{i-m})}\right)\left(\frac{\hat{F}(U_{i+m})-\hat{F}(U_{i-m})}{\sum_{j=1}^{n}\left(\hat{F}(U_{j+m})-\hat{F}(U_{j-m})\right)}\right), \tag{3}$$

где

$$\hat{F}(U_i) = \frac{n-1}{n(n+1)}\left(i+\frac{1}{n-1}+\frac{U_i-U_{i-1}}{U_{i+1}-U_{i-1}}\right),\ i=2,\dots,n-1,$$

$$\hat{F}(U_1) = 1-\hat{F}(U_n) = \frac{1}{(n+1)}.$$

Для обеих модификаций гипотеза о равномерности отклоняется при больших значениях статистики. Характер зависимости свойств критериев от параметра окна m, такой же как и для критерия Дудевича-ван дер Мюлена.

3. Статистика **критерия Пардо** имеет вид

$$E_{m,n} = \frac{1}{n}\sum_{i=1}^{n}\frac{2m}{n\left(U_{i+m}-U_{i-m}\right)}, \tag{4}$$

где $U_0 = 0,\ U_{n+1} = 1$.

Критерий правосторонний, проверяемая гипотеза отклоняется при больших значениях статистики.

4. **Выбор оптимального размера параметра** m для критериев, в качестве статистик которых используются различные оценки энтропии, представляет собой актуальную проблему. Большинство авторов предлагают использовать такие m, при которых значения соответствующих непараметрических оценок энтропии ближе к её теоретическому значению. Но размер окна m влияет также на мощность критериев. На выбор оптимального размера m, при котором мощность наиболее высока, влияет также объем проверяемой выборки n.

Наиболее распространенными способами выбора m является выбор в соответствии с соотношением $m=\sqrt{n}$ или выбор m^*, в соответствии с рекомендациями, приведенными в таблице, которые сформулированы на основании результатов анализа мощности:

Оптимальные значения m^*

n	m^*	n	m^*
$n \le 5$	1	$40 \le n \le 100$	6
$6 \le n \le 8$	2	$101 \le n \le 150$	7
$9 \le n \le 18$	3	$151 \le n \le 200$	8
$19 \le n \le 29$	4	$n > 200$	9
$30 \le n \le 39$	5		

Порядок выполнения

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

1. Исследовать зависимость распределений статистики Дудевича-ван дер Мюлена (1) от объема выборок в случае принадлежности наблюдений равномерном закону на интервале [0,1] для одного из методов выбора параметра m ($m = \sqrt{n}$ или m^*).
2. Оценить мощность критерия Дудевича-ван дер Мюлена относительно заданных конкурирующих гипотез (см. вариант предшествующей лабораторной работы) при выбора параметра m различным образом ($m = \sqrt{n}$ и m^*).
3. Оценить мощность критериев со статистиками (1) – (4) относительно тех же заданных альтернатив при m^* (см. таблицу).
4. Сравнить полученные оценки мощности с мощностями критериев, рассмотренных в предшествующей работе.

Контрольные вопросы

1. В каких ситуаций модификации (2), (3) более эффективны чем критерий Дудевича-ван дер Мюлена?
2. Что влияет на выбор оптимального значения параметра m?
3. Достоинства и недостатки энтропийных критериев по сравнению с другими критериями равномерности?

Лабораторная работа № 7. 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

Классические критерии проверки гипотез о математических ожиданиях и дисперсиях. 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 $T_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$, которая при справедливости гипотезы H_0 подчиняется нормальному распределению: $G(T_1 \mid H_0) = N\left(\mu_0, \frac{\sigma^2_{ист}}{n}\right)$. Проверяемая гипотеза H_0 отклоняется при больших отклонениях T_1 от μ_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 $T_3 = \frac{1}{\sigma_0^2}\sum_{i=1}^{n} x(\xi_i - \mu_{ист})^2$, 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 $T_4 = \frac{1}{\sigma_0^2}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$, 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.

В процессе выполнения работы использовать программу **isw.exe.**

4. 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.
5. 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.
6. 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, а в каких случаях заметно изменение в распределениях. Какие отклонения приводят к существенным изменениям распределений статистик T_1, T_2?
7. Провести аналогичные исследования распределений статистик T_3, T_4, убедиться в существенной зависимости распределений статистик от наблюдаемого закона.

Варианты заданий.

№	$F(x,\theta)$	$F_1(x,\theta_1)$
1	Нормальное (0, 1)	Экспоненциальное семейство с параметром формы (1)
2	Нормальное (0.1, 1)	Экспоненциальное семейство с параметром формы (3)
3	Нормальное (0.1, 2)	Экспоненциальное семейство с параметром формы (5)
4	Нормальное (0.5, 1)	Коши (0,1)

5	Нормальное (0.5, 2)	Экспоненциальное семейство с параметром формы (0.5)
6	Нормальное (0, 1.5)	Экспоненциальное семейство с параметром формы (0.2)
7	Нормальное (0, 2)	Гамма-распределение с параметром формы (2)
8	Нормальное (0, 3)	Распределение минимального значения
9	Экспоненциальное (0, 1)	Распределение максимального значения
10	Нормальное (1, 3)	Показательное 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?

Лабораторная работа № 8. Исследование статистик корреляционного анализа

Цель работы. Вычисление статистик корреляционного анализа по эмпирическим данным, полученным при моделировании многомерного нормального закона распределения с различными параметрами. Исследование распределений статистик корреляционного анализа с ростом 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_i;

– $\Sigma = [\sigma_{ij}]_{i,j=1}^{m}$ – ковариационная матрица случайного вектора X_i;

– $\hat{M}$ и $\hat{\Sigma}$ – оценки максимального правдоподобия (ОМП) для вектора математического ожидания и ковариационной матрицы, вычисляемые по негруппированным данным:

$$\hat{M} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ и } \hat{\Sigma} = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \hat{M}\right)\left(X_i - \hat{M}\right)^T .$$

В данной работе проведены исследования распределений статистик, связанных с проверкой следующих гипотез.

1. Проверка гипотезы о равенстве математического ожидания некоторому неизвестному вектору

Важной статистической проблемой является проблема проверки гипотезы о том, что вектор среднего значения нормального распределения является данным вектором. Такая задача очень часто 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). Рассмотрим эту проблему сначала в предположении, что ковариационная матрица известна, а затем – когда неизвестна.

Гипотеза имеет вид $H_0 : M = M_0$. И, как ранее отмечалось, здесь возможны две ситуации.

а) В случае, когда нам известна ковариационная матрица Σ (например, из ранее проводимых экспериментов или предположений), то для вывода предельных распределений статистик, используемых при проверке данной гипотезы, в многомерном случае используется факт, что разность между векторами среднего значения выборки и среднего значения генеральной совокупности распределена нормально с вектором математического ожидания, равным нулю, и известной ковариационной матрицей.

В этом случае статистика критерия имеет вид:

$$X_m^2 = n\left(\hat{M} - M_0\right)^T \Sigma^{-1}\left(\hat{M} - M_0\right). \quad (1)$$

При справедливости гипотезы H_0 предельным распределением статистики является $G(X_m^2 \mid H_0) = \chi_m^2$ - распределение, с числом степеней свободы m. Таким образом, гипотеза H_0 принимается, если

$$X_m^2 < \chi_{m,\alpha}^2,$$

где α - уровень значимости и

$$1-\alpha = P\left\{X_m^2 > \chi_{m,\alpha}^2\right\} = \frac{1}{2^{m/2}\Gamma(m/2)} \int_0^{\chi_{m,\alpha}^2} s^{m/2-1} e^{-s/2} ds.$$

б) Другая более важная группа статистических задач связана с вопросами, касающимися оценки математического ожидания многомерного нормального распределения, дисперсия которого неизвестна. Так в одномерном случае используется статистика, являющаяся частным от деления разности между выборочным средним значением и гипотетическим математическим ожиданием генеральной совокупности на среднее квадратичное отклонение. В предложен многомерный аналог данной величины, в которую также вовлечены математическое ожидание и обратная матрица к оценке ковариационной матрицы.

В этом случае согласно используется статистика:

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

Тогда при справедливости гипотезы H_0 предельное распределение статистики $G(T^2 \mid H_0) = F_{m,n-m}$ - распределение Фишера с параметрами m и n-m. В данном случае гипотеза H_0 принимается, если выполняется условие

$$T^2 < F_{m,n-m,\alpha}.$$

Величина $F_{m,n\text{-}m,\alpha}$ определятся из равенства

$$1-a = P\left\{T^2 > F_{m,n-m,\alpha}\right\} = \left(\frac{m}{n-m}\right) \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n-m}{2}\right)} \int_0^{F_{m,n-m,\alpha}} s^{m/2-1}\left(1+\frac{m}{n-m}s\right)^{-n/2} ds.$$

2. Проверка гипотез о коэффициенте парной корреляции

При анализе совокупности случайных величин нас может интересовать, например, взаимосвязь между несколькими случайными величинами, либо 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

– есть частная корреляция, а вот 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 парного коэффициента корреляции может быть найдена в соответствии с выражением

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

Здесь можно решать следующие задачи: определение оценки парного коэффициента корреляции, проверка гипотезы о его значимости (гипотеза вида: $H_0 : r_{ij} = 0$) и проверка гипотезы о равенстве его определенному значению (гипотеза вида: $H_0 : r_{ij} = r_0$).

Для проверки гипотезы $H_0 : r_{ij} = 0$ согласно вычисляется статистика:

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

При этом предельное распределение статистики $G(t \mid H_0) = t_{n-2}$ - распределение Стьюдента с числом степеней свободы n-2. При конкурирующей гипотезе $H_1 : r_{ij} \neq 0$ гипотеза H_0 принимается, если

$$t < t_{n-2,\alpha/2},$$

где α - уровень значимости. Величина $t_{n-2,\alpha/2}$ при $k = n-2$ определяется равенством

$$1-\alpha = P\{t < t_{k,\alpha/2}\} = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\,\Gamma\left(\frac{k}{2}\right)} \int_{-t_{k,\alpha/2}}^{t_{k,\alpha/2}} \left(1+\frac{s^2}{k}\right)^{-\frac{k+1}{2}} ds.$$

Если же гипотеза имеет вид $H_0 : r_{ij} = r_0$, то используется статистика:

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

При этом асимптотическим распределением статистики $G(z_0 \mid H_0)$ является стандартное нормальное распределение $N(0,1)$. Гипотеза H_0 принимается, если

$$z < t_{\alpha/2},$$

где $t_{\alpha/2}$ – квантиль стандартного нормального распределения и

$$1-a = P\{t < t_{\alpha/2}\} = \frac{1}{\sqrt{2\pi}} \int_{-t_{\alpha/2}}^{t_{\alpha/2}} e^{-s^2/2} ds.$$

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. Однако на практике при интерпретации «взаимозависимости» часто встречаются определенные трудности, а именно: 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. Указанная возможность приводит к рассмотрению условных корреляций между двумя величинами при фиксированных значениях остальных величин. Это так называемые *частные корреляции*.

Если корреляция между двумя величинами уменьшается при фиксировании 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. Наоборот, когда частная корреляция больше первоначальной корреляции между двумя величинами, то следует, что другие величины ослабляли связь, или, можно сказать, «маскировали» корреляцию. Но нужно 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 всегда 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, $2 \le l \le m$.
В данном случае при оценке взаимозависимости между компонентами x_i и x_j случайной величины X исключается влияние компонент $x_{l+1}, x_{l+2}, \ldots, x_m$.

При проверке 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 те же самые статистики, что и для парного коэффициента корреляции. В этом случае в соответствующих соотношениях *n* заменяется на $n - m + l$.

Для проверки гипотезы $H_0 : r_{ij;l+1,\ldots,m} = 0$ вычисляется статистика:

$$t = \frac{\sqrt{n-m+l-2}\left|\hat{r}_{ij\bullet l+1,\ldots,m}\right|}{\sqrt{1-\hat{r}_{ij\bullet l+1,\ldots,m}}}. \tag{5}$$

При этом предельным распределением статистики $G(t\,|\,H_0)$ является $t_{n-m+l-2}$ - распределение Стьюдента с числом степеней свободы *n*-*m*+*l*-2.
Если же гипотеза имеет вид $H_0 : r_{ij;l+1,\ldots,m} = r_0$, тогда используется статистика:

$$z_0 = \sqrt{n-3}\left(\frac{1}{2}\ln\left(\frac{1+\hat{r}_{ij\bullet l+1,\ldots,m}}{1-\hat{r}_{ij\bullet l+1,\ldots,m}}\right) - \frac{1}{2}\ln\left(\frac{1+r_0}{1-r_0}\right) - \left(\frac{r_0}{2(n-1)}\right)\right). \tag{6}$$

При этом предельное распределение статистики $G(z_0\,|\,H_0) = N(0,1)$ – стандартное нормальное распределение.

4. Проверка гипотезы о коэффициенте множественной корреляции

Множественный коэффициент корреляции является мерой зависимости компоненты многомерной случайной величины от некоторого множества компонент.

Можно рассматривать корреляцию между одной компонентой случайного вектора и множеством всех остальных или каким-то подмножеством.

Следует отметить, что множественный коэффициент корреляции r_i 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 $E\left[x_1 \middle| x_2, x_3, \ldots, x_n\right]$.

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:

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

Для гипотезы $H_0 : r_{i\bullet l+1,\ldots,m} = 0$ вычисляется статистика:

$$F = \frac{n-m+l-1}{m-l}\frac{\hat{r}_{i\bullet l+1,\ldots,m}^{2}}{1-\hat{r}_{i\bullet l+1,\ldots,m}^{2}}. \qquad (7)$$

При этом предельным распределением статистики $G(F\,|\,H_0)$ является $F_{m-l,n-m+l-1}$–распределение Фишера с числом степеней свободы m-l и n-m+l-1. Гипотеза H_0 не отклоняется, если

$$F < F_{m-l,n-m+l-1,\alpha},$$

где α - уровень значимости, а $F_{m-l,n-m+l-1,\alpha}$ – критическая точка критерия с уровнем значимости α.

Порядок выполнения.

1. Вычислить, если это необходимо, значения для коэффициентов корреляции.

Используя программу **STAR System.exe:**

2. Смоделировать выборки статистик корреляционного анализа по предложенным параметрам и гипотезам, соответствующим варианту задания, с сохранением в файл. Исследовать влияние различных значений объемов выборок многомерной случайной величины на значения критериев согласия. Определить какие статистики корреляционного 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 о распределениях статистик корреляционного анализа.

Варианты заданий

Наборы параметров для моделирования

Перечислим используемые параметры для моделирования законов распределения:

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

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

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

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

Гипотезы корреляционного анализа

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 предельные распределения используются при проверке гипотез о векторе математического ожидания, и как осуществляется проверка гипотез?
2. Какие предельные распределения используются при проверке гипотез о парном коэффициенте корреляции, и как осуществляется проверка гипотез?
3. Какие предельные распределения используются при проверке гипотез о частном коэффициенте корреляции, и как осуществляется проверка гипотез?
4. Какое предельное распределение используется при проверке гипотезы о множественном коэффициенте корреляции и как осуществляется проверка гипотезы?

Список литературы

1. Лемешко Б.Ю., Постовалов С.Н. Компьютерные технологии анализа данных и исследования статистических закономерностей: Учебное пособие. – Новосибирск: Изд-во НГТУ, 2004. – 119 с. https://ami.nstu.ru/~headrd/seminar/publik_html/Pos_KTAD.pdf.
2. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход : монография / Б.Ю. Лемешко, С.Б. Лемешко, С.Н. Постовалов, Е.В. Чимитова. – Новосибирск : Изд-во НГТУ, 2011. – 888 с. (серия «Монографии НГТУ»). https://ami.nstu.ru/~headrd/seminar/publik_html/Statistical_Data_Analysis.pdf.
3. Материалы к семинарам с указанием на различные разделы курса https://ami.nstu.ru/~headrd/seminar/start.htm
4. Лемешко Б.Ю. Критерии проверки гипотез об однородности. Руководство по применению. – М. : ИНФРА-М, 2017. – 208 с. https://ami.nstu.ru/~headrd/seminar/publik_html/Guide_homogeneity.pdf
5. Лемешко Б.Ю., Блинов П.Ю. Критерии проверки отклонения распределения от равномерного закона. Руководство по применению. – М.: ИНФРА-М, 2015. – 183 с. https://ami.nstu.ru/~headrd/seminar/publik_html/test_random_lection.pdf
6. Лемешко Б.Ю., Веретельникова В.И. Критерии проверки гипотез о случайности и отсутствии тренда. (*Препринт*) https://ami.nstu.ru/~headrd/seminar/publik_html/guid_trend_test.pdf
7. Помадин С.С. Исследование распределений статистик многомерного анализа данных при нарушении предположений о нормальности // Диссертация на соискание уч. ст. к.т.н. по специальности 05.13.17 — теоретические основы информатики / Научн. рук. д.т.н., проф. Лемешко Б.Ю./ Новосибирский государственный технический университет, 2004 г. https://ami.nstu.ru/~headrd/Korr_analiz.htm
8.

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, $\theta_0 \in R$, $\theta_1 > 0$, $x \in [-\infty,+\infty]$	$\frac{\pi}{\theta_0\sqrt{3}}\exp\left\{-\frac{\pi(x-\theta_1)}{\theta_0\sqrt{3}}\right\} \Bigg/ \left[1+\exp\left\{-\frac{\pi(x-\theta_1)}{\theta_0\sqrt{3}}\right\}\right]^2$
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 (обобщенный нормальный закон), $\theta_1 \in R$, $\theta_0 > 0$, $\theta_2 > 0$, $x \in [-\infty, +\infty]$	$f(x) = \frac{\theta_2}{2\theta_1 \Gamma(1/\theta_2)} \exp\left\{ -\left| \frac{x - \theta_0}{\theta_1} \right|^{\theta_2} \right\}$