

УДК 519.2:303.732.4

UDC 519.2:303.732.4

СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ ДЛЯ СГРУППИРОВАННЫХ ДАННЫХ**STATISTICAL ESTIMATION FOR THE GROUPED DATA**

Орлов Александр Иванович
д.э.н., д.т.н., к.ф.-м.н., профессор

Orlov Alexander Ivanovich
Dr.Sci.Econ., Dr.Sci.Tech., Cand.Phys-Math.Sci.,
professor
Bauman Moscow State Technical University,
Moscow, Russia

Московский государственный технический университет им. Н.Э. Баумана, Россия, 105005, Москва, 2-я Бауманская ул., 5, prof-orlov@mail.ru

Описана вероятностная модель группировки данных, в том числе многомерных. Обобщена формулы Эйлера-Маклорена. С ее помощью получены поправки Шеппарда и поправки на группировку для коэффициента корреляции. Найдены и изучены асимптотические поправки на группировку в общем случае. Оценена точность приближения

The probabilistic model of grouping data (including multidimensional data) is described. We have also generalized Euler-Maclaurin's formulas. With its help Sheppard's corrections and corrections on grouping for correlation coefficient are received. We have found and studied asymptotical corrections on grouping data generally. Accuracy of approach has been estimated

Ключевые слова: СТАТИСТИЧЕСКИЕ МЕТОДЫ, МАТЕМАТИЧЕСКАЯ СТАТИСТИКА, ПРИКЛАДНАЯ СТАТИСТИКА, МОДЕЛЬ ГРУППИРОВКИ ДАННЫХ, ОБОБЩЕНИЕ ФОРМУЛЫ ЭЙЛЕРА-МАКЛОРЕНА. ПОПРАВКИ НА ГРУППИРОВКУ, ТОЧНОСТЬ ПРИБЛИЖЕНИЯ

Keywords: STATISTICAL METHODS, MATHEMATICAL STATISTICS, APPLIED STATISTICS, MODEL OF GROUPING DATA, GENERALIZED EULER-MACLAURIN'S FORMULA, CORRECTIONS ON GROUPING, ACCURACY OF APPROACH

1. Введение

При вычислении различных статистических характеристик часто пользуются сгруппированными данными. Погрешность группировки - один из видов погрешностей наблюдений. До появления компьютеров группировку проводили для облегчения расчетов. К сожалению, устаревшие рекомендации по обязательному проведению группировки укоренились в курсах по «общей теории статистики». В отмененном из-за низкого научного уровня ГОСТ 11.006-74 (СТ СЭВ 1190-78) «Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим» группировку предписывалось проводить даже при проверке согласия по критериям Колмогорова и омега-квадрат, что приводило к ошибочным статистическим выводам.

В современных условиях нет необходимости сокращать число арифметических операций при анализе статистических данных. Вместо гистограмм для описания распределения в настоящее время рекомендуют

использовать непараметрические оценки плотности [1]. Однако с существующей традицией приходится считаться, следовательно, надо уметь оценивать влияние группировки данных на статистические характеристики.

Другая причина использования сгруппированных данных – неточность измерений, приводящих либо к автоматическому (с помощью средств измерения) округлению, либо к округлению, проводимому специалистом, осуществляющим измерение.

В настоящей статье рассмотрим статистические методы анализа сгруппированных данных. Начнем с одномерных данных.

2. Вероятностная модель группировки

Пусть выборка объема n взята из непрерывного распределения числовой случайной величины X с плотностью вероятности $f(x)$. Пусть элементы выборки сгруппированы по интервалам длины h , центры которых находятся в точках $a_i = a_0 + ih$, где $i = 0, \pm 1, \pm 2, \dots$ (Поскольку для непрерывной случайной величины вероятность попадания в точку соприкосновения интервалов равна 0, то нет необходимости указывать, к какому из интервалов относятся такие точки.) В таких случаях при вычислении моментов и других выборочных характеристик по сгруппированным данным обычно предполагается [2, с.393-394], что все выборочные значения, принадлежащие некоторому интервалу, совпадают с центром этого интервала. Тогда фактически рассматривается выборка из дискретного распределения, в соответствии с которым случайная величина Y принимает значения $a_i = a_0 + ih$ с вероятностями

$$p_i = P(Y = a_i) = \int_{a_i - \frac{h}{2}}^{a_i + \frac{h}{2}} f(x) dx .$$

Пользуясь сгруппированными данными, оценивают те или иные выборочные характеристики, например, моменты

$$M(Y^k) = \sum_{-\infty < i < +\infty} a_i^k p_i,$$

которые будем называть групповыми. Во многих случаях исследователя интересуют моменты исходной непрерывной случайной величины X :

$$M(X^k) = \int_{-\infty}^{+\infty} x^k f(x) dx.$$

Поэтому важно изучить соотношения между этими двумя множествами моментов. При некоторых условиях регулярности приближенные значения моментов $M(X^k)$ можно получить, введя поправки к групповым моментам $M(Y^k)$.

Групповые моменты запишем в виде

$$M(Y^k) = \sum_{-\infty < i < +\infty} a_i^k \int_{a_i - \frac{h}{2}}^{a_i + \frac{h}{2}} f(x) dx = \sum_{-\infty < i < +\infty} g(a_i), \quad g(a) = a^k \int_{a - \frac{h}{2}}^{a + \frac{h}{2}} f(x) dx.$$

Для решения поставленной задачи применим формулу Эйлера-Маклорена [2, с.394-398] и ее обобщение, полученное в [3].

3. Обобщение формулы Эйлера-Маклорена

В различных статистических методах анализа данных возникает необходимость вычисления сумм следующего вида:

$$S(g, a, h, n) = \sum_{i=1}^n g(a + (i-1)h),$$

где $g(x)$ – достаточно гладкая функция. Другими словами, требуется просуммировать значения функции $g(x)$ в n точках, отстоящих друг от друга на расстояние h , начиная с точки a . Типичным примером является суммирование биномиальных вероятностей с целью вычисления попадания биномиально распределенной случайной величины в заданный интервал.

Классическая формула Эйлера-Маклорена такова [4, п.465]:

$$S(g, a, h, n) = \frac{1}{h} \int_a^{a+nh} g(x) dx + A_1 \{g(a+nh) - g(a)\} + A_2 h \{g'(a+nh) - g'(a)\} + \dots + A_m h^{m-1} \{g^{(m-1)}(a+nh) - g^{(m-1)}(a)\} + R_m.$$

Здесь R_m – остаточный член порядка h^m , выражение для которого приведено в [4, с.540],

$$A_1 = -\frac{1}{2}, \quad A_{2p-1} = 0, \quad p > 1, \quad A_{2p} = (-1)^{p-1} \frac{B_p}{(2p)!}, \quad p \geq 1,$$

где B_p есть p -е число Бернулли [4, п.449]. В частности,

$$A_2 = \frac{1}{12}, \quad A_4 = -\frac{1}{720}, \quad A_6 = \frac{1}{30240}.$$

Очевидным недостатком классической формулы Эйлера-Маклорена является ее несимметричность – левый конец отрезка $[a, a+nh]$, по которому проводится интегрирование, входит в число точек, значения функции в которых суммируется, а правый – нет. Этот недостаток связан с тем, что каждому слагаемому в рассматриваемой сумме ставится в соответствие отрезок длины h , значение функции $g(x)$ в *левом* конце которого и есть рассматриваемое слагаемое. В результате нет симметричности – формула меняется при изменении направления оси x -ов.

Чтобы избавиться от несимметричности, достаточно слагаемому $g(x_0)$ (где $x_0 = a + (i-1)h$ при некотором i) поставить в соответствие отрезок $\left[x_0 - \frac{h}{2}; x_0 + \frac{h}{2}\right]$. Опишем подход к получению асимптотических разложений, впервые предложенный в статье [3] в связи с изучением скорости сходимости распределения классической статистики омега-квадрат (Крамера-Мизеса-Смирнова) к предельному распределению. Будем исходить из формулы Тейлора с дополнительным членом в виде определенного интеграла [4, п.318]:

$$g(x) = g(x_0) + \frac{g'(x_0)}{1!}(x-x_0) + \frac{g''(x_0)}{2!}(x-x_0)^2 + \dots + \frac{g^{(m)}(x_0)}{m!}(x-x_0)^m + \frac{1}{m!} \int_{x_0}^x g^{(m+1)}(t)(x-t)^m dt$$

Проинтегрируем почленно обе части этого равенства по x по отрезку отрезок $\left[x_0 - \frac{h}{2}; x_0 + \frac{h}{2} \right]$. Получим, что

$$\int_{x_0 - \frac{h}{2}}^{x_0 + \frac{h}{2}} g(x) dx = g(x_0)h + \frac{g'(x_0)}{1!} \int_{x_0 - \frac{h}{2}}^{x_0 + \frac{h}{2}} (x-x_0) dx + \frac{g''(x_0)}{2!} \int_{x_0 - \frac{h}{2}}^{x_0 + \frac{h}{2}} (x-x_0)^2 dx + \dots + \frac{g^{(m)}(x_0)}{m!} \int_{x_0 - \frac{h}{2}}^{x_0 + \frac{h}{2}} (x-x_0)^m dx + \frac{1}{m!} \int_{x_0 - \frac{h}{2}}^{x_0 + \frac{h}{2}} \left\{ \int_{x_0}^x g^{(m+1)}(t)(x-t)^m \right\} dx = g(x_0)h + \frac{g''(x_0)}{2!} \frac{h^3}{12} + \frac{g^{(4)}(x_0)}{4!} \frac{h^5}{80} + \dots + \frac{g^{(m)}(x_0)}{m!} \frac{h^{m+1}}{2^m(m+1)} + \frac{1}{m!} \int_{x_0 - \frac{h}{2}}^{x_0 + \frac{h}{2}} \left\{ \int_{x_0}^x g^{(m+1)}(t)(x-t)^m \right\} dx$$

(при четном m). Весьма важно, что обращаются в 0 все интегралы, соответствующие нечетным степеням.

Из только что полученного соотношения следует, что

$$g(x_0)h = \int_{x_0 - \frac{h}{2}}^{x_0 + \frac{h}{2}} g(x) dx - \frac{g''(x_0)h^3}{24} - \frac{g^{(4)}(x_0)h^5}{1920} - \dots - \frac{g^{(m)}(x_0)}{m!} \frac{h^{m+1}}{2^m(m+1)} - Q_m,$$

где Q_m – остаточный член, указанный в предыдущей формуле.

Применим аналогичную процедуру ко второй производной - функции $g''(x)$. Исходим из формулы Тейлора

$$g''(x) = g''(x_0) + \frac{g^{(3)}(x_0)}{1!}(x-x_0) + \frac{g^{(4)}(x_0)}{2!}(x-x_0)^2 + \dots$$

Интегрируя, получаем, что

$$\int_{x_0 - \frac{h}{2}}^{x_0 + \frac{h}{2}} g''(x) dx = g''(x_0)h + \frac{g^{(4)}(x_0)}{2!} \frac{h^3}{12} + \dots,$$

откуда

$$g''(x_0)h = \int_{x_0-\frac{h}{2}}^{x_0+\frac{h}{2}} g''(x)dx - \frac{g^{(4)}(x_0) h^3}{2! \cdot 12} - \dots = \left\{ g' \left(x_0 + \frac{h}{2} \right) - g' \left(x_0 - \frac{h}{2} \right) \right\} - \frac{g^{(4)}(x_0) h^3}{2! \cdot 12} - \dots$$

Подставляя в формулу для $g(x_0)h$, получаем, что

$$\begin{aligned} g(x_0)h &= \int_{x_0-\frac{h}{2}}^{x_0+\frac{h}{2}} g(x)dx - \frac{h^2}{24} \left\{ \int_{x_0-\frac{h}{2}}^{x_0+\frac{h}{2}} g''(x)dx - \frac{g^{(4)}(x_0) h^3}{2! \cdot 12} - \dots \right\} - \frac{g^{(4)}(x_0)h^5}{1920} + O(h^7) = \\ &= \int_{x_0-\frac{h}{2}}^{x_0+\frac{h}{2}} g(x)dx - \frac{h^2}{24} \int_{x_0-\frac{h}{2}}^{x_0+\frac{h}{2}} g''(x)dx + \frac{7g^{(4)}(x_0)h^5}{5760} + O(h^7) = \int_{x_0-\frac{h}{2}}^{x_0+\frac{h}{2}} g(x)dx - \\ &- \frac{h^2}{24} \left\{ g' \left(x_0 + \frac{h}{2} \right) - g' \left(x_0 - \frac{h}{2} \right) \right\} + \frac{7g^{(4)}(x_0)h^5}{5760} + O(h^7). \end{aligned}$$

Дальше можно провести аналогичные выкладки для четвертой производной, и т.д. С помощью продемонстрированного подхода можно получить, например, следующий результат:

$$g(x_0) = \frac{1}{h} \int_{x_0-\frac{h}{2}}^{x_0+\frac{h}{2}} A(x)dx,$$

где

$$\begin{aligned} A(x) &= g(x) - \frac{h^2}{24} g''(x) + \frac{7h^4}{5760} g^{(4)}(x) - \frac{31h^6}{967680} g^{(6)}(x) + \\ &+ \int_{x_0}^x g^{(8)}(t) \left\{ -\frac{(x-t)^7}{5040} + \frac{(x-t)^5 h^2}{2880} - \frac{7(x-t)^3 h^4}{34560} + \frac{31(x-t)h^6}{967680} \right\} dt. \end{aligned}$$

Таким образом, справедливо следующее соотношение, напоминающее классическую формулу Эйлера-Маклорена:

$$\begin{aligned} S(g, a, h, n) &= \sum_{i=1}^n g(a + (i-1)h) = \frac{1}{h} \int_{a-\frac{h}{2}}^{a+\left(n-\frac{1}{2}\right)h} g(x)dx - \frac{h}{24} \left\{ g' \left(a + \left(n - \frac{1}{2} \right) h \right) - g' \left(a - \frac{h}{2} \right) \right\} + \\ &+ \frac{7h^3}{5760} \left\{ g^{(3)} \left(a + \left(n - \frac{1}{2} \right) h \right) - g^{(3)} \left(a - \frac{h}{2} \right) \right\} - \frac{31h^5}{967680} \left\{ g^{(5)} \left(a + \left(n - \frac{1}{2} \right) h \right) - g^{(5)} \left(a - \frac{h}{2} \right) \right\} + O(h^7). \end{aligned}$$

Сравнение с классической формулой Эйлера-Маклорена показывает, что в последней формуле отсутствует первый член асимптотического

разложения, соответствующий $A_1\{g(a+nh) - g(a)\}$, а коэффициенты при остальных членах меньше по абсолютной величине:

$$\frac{1}{24} < A_2 = \frac{1}{12}; \frac{7}{5760} < |A_4| = \frac{1}{720}; \frac{31}{967680} < A_6 < \frac{1}{30240}.$$

Полученная формула симметрична – не меняется при изменении направлении оси x -ов.

Из сказанного вытекает рекомендация: вместо классической формулы Эйлера-Маклорена употреблять полученную в статье [3] формулу, в которой каждому входящему в сумму значению аргумента ставится в соответствие отрезок, для которого это значение является центром. Полученная формула позволила, в частности, построить компактные таблицы для специальных функций, используемых при оценивании параметров гамма-распределения, и вывести формулы для использования вне таблиц, при разработке государственного стандарта ГОСТ 11.011-83 [5]. Многомерный аналог полученной формулы, в котором каждой точке целочисленной решетки ставился в соответствие прямоугольный параллелепипед, для которого эта точка являлась центром, позволил разработать оригинальный метод оценивания скорости сходимости распределений непараметрических статистик типа Колмогорова, Смирнова и омега-квадрат [3].

4. Поправки Шеппарда

С помощью формулы Эйлера-Маклорена групповые моменты $M(Y^k)$ могут быть выражены линейными функциями от «истинных» моментов $M(X^k)$ (в предположении, что остаточным членом можно пренебречь). Решая последовательно уравнения относительно «истинных» моментов $M(X^k)$, получаем [2, с.395]:

$$\begin{aligned}
 M(X) &= M(Y), \\
 M(X^2) &= M(Y^2) - \frac{1}{12}h^2, \\
 M(X^3) &= M(Y^3) - \frac{1}{4}M(Y)h^2, \\
 M(X^4) &= M(Y^4) - \frac{1}{2}M(Y^2)h^2 + \frac{7}{240}h^4, \dots
 \end{aligned}$$

Таким образом, получены поправки к результатам расчетов по сгруппированным данным, позволяющие более точно оценить моменты исходного распределения. Эти поправки были впервые выведены В.Ф. Шеппардом в 1898 г. [6] и поэтому называются в статистической литературе «поправками Шеппарда». Общая формула имеет вид [7]:

$$M(X^k) = \sum_{j=0}^k C_k^j (2^{1-j} - 1) B_j M(Y^{k-j}) h^j,$$

где B_j – числа Бернулли [4, п.449].

Хотя поправки Шеппарда получены еще в XIX веке, статистическая теория обработки сгруппированных данных продолжает развиваться. Ниже рассмотрена непараметрическая многомерная постановка, в которой границы группировки задаются статистиком. Параметрическую теорию развивал Г. Куллдорф [8]. Модель группировки со случайным сдвигом начала координат изучал Н.А. Бодин [9-11]. Вместо округления результаты измерений могут быть представлены в виде случайных интервалов, как это принято в статистике интервальных данных [1, гл. 12].

5. Многомерная группировка

Дадим описание группировки данных в случае m -мерного случайного вектора $Z = (Z_1, Z_2, \dots, Z_m)$ с плотностью распределения $p(x_1, x_2, \dots, x_m)$. Пусть выборочные значения координаты Z_i группируются по интервалам длины h_i со средними точками (центрами) $a_{i,k} = a_i + kh_i, k = 0, \pm 1, \pm 2, \dots$. Вектор $h = (h_1, h_2, \dots, h_m)$ назовем шагом группировки.

При вычислении выборочных характеристик предполагается, что все выборочные значения вектора Z , координаты которого удовлетворяют неравенствам

$$a_{i,k(i)} - \frac{h_i}{2} < Z_i \leq a_{i,k(i)} + \frac{h_i}{2}, \quad i = 1, 2, \dots, m,$$

совпадают с центром $(a_{1,k(1)}, a_{2,k(2)}, \dots, a_{m,k(m)})$ данного m -мерного прямоугольного параллелепипеда, т.е. фактически мы обрабатываем выборочные значения дискретного m -мерного вектора $W = (W_1, W_2, \dots, W_m)$ с распределением

$$P\{W = (a_{1,k(1)}, a_{2,k(2)}, \dots, a_{m,k(m)})\} = \int_A p(x_1, x_2, \dots, x_m) dx_1 dx_2 \dots dx_m,$$

где

$$A = \left\{ (x_1, x_2, \dots, x_m) : a_{i,k(i)} - \frac{h_i}{2} < x_i \leq a_{i,k(i)} + \frac{h_i}{2}, \quad i = 1, 2, \dots, m \right\}.$$

Представляется естественным разложить разность между характеристиками вектора W и характеристиками вектора Z по степеням координат шага группировки и оценить возникающий при этом остаточный член, т.е. получить обобщение поправок Шеппарда для моментов. Будем рассматривать характеристики вида $Mf(Z)$, где f – достаточно гладкая функция.

6. Поправки на группировку для коэффициента корреляции

В качестве примера рассмотрим линейный парный коэффициент корреляции Пирсона $\rho(Z_1, Z_2)$. Как известно, для удовлетворяющей некоторым условиям регулярности [2, п.27.9] плотности $p(x_1, x_2)$ справедливы соотношения

$$M(W_1) = M(Z_1), \quad M(W_2) = M(Z_2), \quad M(W_1 W_2) = M(Z_1 Z_2),$$

$$D(W_1) = D(Z_1) + \frac{h_1^2}{12}, \quad D(W_2) = D(Z_2) + \frac{h_2^2}{12}$$

при достаточно малых h_1 и h_2 с точностью до членов более высокого порядка по h_1 и h_2 . Из приведенных соотношений и определения коэффициента корреляции $\rho(Z_1, Z_2)$ с помощью элементарных преобразований получаем, что

$$\rho(W_1, W_2) - \rho(Z_1, Z_2) = \rho(Z_1, Z_2) \left(\frac{1}{\sqrt{1 + \frac{h_1^2}{12D(Z_1)}} \sqrt{1 + \frac{h_2^2}{12D(Z_2)}}} - 1 \right)$$

с той же точностью. Поскольку при малых y

$$\frac{1}{\sqrt{1+y}} = 1 - \frac{y}{2} + O(y^2),$$

то при достаточно малых h_1 и h_2 с точностью до членов более высокого порядка

$$\rho(W_1, W_2) - \rho(Z_1, Z_2) = \rho(Z_1, Z_2) \left(-\frac{h_1^2}{24D(Z_1)} - \frac{h_2^2}{24D(Z_2)} \right).$$

Из последней формулы вытекает, что

$$\rho(Z_1, Z_2) = \rho(W_1, W_2) \left(1 + \frac{h_1^2}{24D(Z_1)} + \frac{h_2^2}{24D(Z_2)} \right).$$

с той же точностью. Воспользовавшись приведенными выше соотношениями для вторых моментов координат двумерного вектора, получаем окончательную формулу

$$\rho(Z_1, Z_2) = \rho(W_1, W_2) \left(1 + \frac{h_1^2}{24D(W_1)} + \frac{h_2^2}{24D(W_2)} \right),$$

с точностью до членов более высокого порядка, в которой поправка на группировку определяется только по сгруппированным данным. Отметим, что группировка приводит к уменьшению коэффициента корреляции (по абсолютной величине):

$$|\rho(Z_1, Z_2)| > |\rho(W_1, W_2)| \quad .$$

Поправка на группировку для коэффициента корреляции была получена в статье [12], возникшей в полемике со статьей [13], в которой без какого-либо основания утверждалось, что при группировке коэффициент корреляции увеличивается. Несмотря на элементарность выкладок, поправка на группировку для коэффициента корреляции не была ранее известна, что и отмечено в реферате статьи [12] в *Mathematical Review*.

7. Асимптотические поправки на группировку

Перейдем к общему случаю. Часть результатов была получена в статье [14], остальные публикуются впервые.

Будем использовать следующий многомерный аналог классической формулы Эйлера-Маклорена:

$$\sum_{v \in Z^m} g(a_1 + v_1 h_1, \dots, a_m + v_m h_m) = \int_{R^m} g(a_1 + x_1 h_1, \dots, a_m + x_m h_m) dx_1 \dots dx_m + \sum_{t=1}^m (-1)^t \int_{R^m} \sum_{\mu \in C(t,m)} \left(\prod_{i \in \mu} P_1(x_i) \frac{\partial}{\partial x_i} \right) g(a_1 + x_1 h_1, \dots, a_m + x_m h_m) dx_1 \dots dx_m,$$

где Z^m – множество точек с целыми координатами m -мерного пространства R^m , функция $g : R^m \rightarrow R^1$ абсолютно интегрируема, $C(t,m)$ – совокупность всех подмножеств мощности t множества $\{1,2,\dots,m\}$ и

$$P_1(q) = \sum_{j=1}^{\infty} \frac{\sin(2\pi qj)}{j\pi}.$$

Справедлив следующий многомерный аналог полученного нами ранее обобщения одномерной классической формулы Эйлера-Маклорена:

$$Mf(W) = Mf(Z) + \sum_{i=1}^m \frac{h_i^2}{24} \int_{R^m} p(x) \frac{\partial^2 f(x)}{\partial x_i^2} dx + \sum_{i=1}^m \frac{h_i^4}{1920} \int_{R^m} p(x) \frac{\partial^4 f(x)}{\partial x_i^4} dx + \sum_{\substack{i,j=1 \\ i \neq j}}^m \frac{h_i^2 h_j^2}{3456} \int_{R^m} p(x) \frac{\partial^4 f(x)}{\partial x_i^2 \partial x_j^2} dx + \dots$$

(в этой формуле $x = (x_1, x_2, \dots, x_m)$, соответственно, $dx = dx_1 dx_2 \dots dx_m$, используются введенные ранее обозначения). Описан закон получения

членов разложения и указана оценка остаточного члена [14]. Однако в общем случае они весьма громоздки, поэтому ограничимся несколькими доведенными до числа примерами.

Пусть $m = 2$, вектор (Z_1, Z_2) имеет нормальное распределение. Тогда

$$M(Z_j) = M(W_j) - R(M(Z_j)), \quad D(W_j) = D(Z_j) + \frac{h_j^2}{12} + R(D(Z_j)), \quad j = 1, 2,$$

$$\text{cov}(W_1, W_2) = \text{cov}(Z_1, Z_2) + R.$$

На основе описанных выше общих результатов можно получить точные формулы и оценки остаточных членов в этих формулах. Оказывается, остаточные члены $R(M(Z_j))$ и $R(D(Z_j))$, $j = 1, 2$, весьма малы даже при достаточно большом шаге группировки. Так,

$$R(M(Z_j)) = \sum_{j=1}^{\infty} \frac{(-1)^j}{j} \exp\left\{-2\left(\frac{j\pi\sigma_j}{h_j}\right)^2\right\} \sin\left(2j\pi \frac{M(Z_j) - a_j}{h_j}\right)$$

и $|R(M(Z_j))|$ не превосходит $10^{-9}\sigma_j$ при $h_j \leq \sigma_j$ и не превосходит $5 \times 10^{-3}\sigma_j$ при $h_j \leq 2\sigma_j$, где $\sigma_j^2 = D(Z_j)$, а a_j – начало отсчета по j -й координате, $j = 1, 2$.

Полученные оценки зависят от $M(Z_j) - a_j$. Так, при $a_j - M(Z_j) = h_j/4$ относительная погрешность определения истинной дисперсии по сгруппированной с использованием поправок Шеппарда, т.е.

$$\left| \frac{D(Z_j) - D(W_j) + \frac{h_j^2}{12}}{D(Z_j)} \right| = \left| \frac{R(D(Z_j))}{D(Z_j)} \right|,$$

не превосходит $6,8 \times 10^{-2}$ при $h_j \leq 3\sigma_j$, меньше $0,25 \times 10^{-2}$ при $h_j \leq 2\sigma_j$ и $6,8 \times 10^{-11}$ при $h_j \leq \sigma_j$. Таким образом (это подтверждается равномерной оценкой остаточного члена), при шаге группировки, равном среднеквадратическому отклонению, формула

$$D(W_j) = D(Z_j) + \frac{h_j^2}{12}$$

становится практически точной.

Иным оказывается положение при расчете по сгруппированным данным коэффициента ковариации. При малых и средних значениях коэффициент корреляции ρ (при $|\rho| \leq 0,8$) относительная погрешность невелика даже при довольно большом шаге группировки (так, $\left| \frac{R}{\rho\sigma_1\sigma_2} \right| \leq 10^{-3}$ при $h_1 = \sigma_1$ и $h_2 = \sigma_2$). При больших коэффициентах корреляции, если ставится цель добиться достаточно хорошего приближения, шаг группировки необходимо уменьшать, поскольку при $|\rho| \rightarrow 1$ основной член в оценке относительной погрешности таков:

$$\left| \frac{R}{\rho\sigma_1\sigma_2} \right| \approx 4 \exp \left\{ -2(1-|\rho|)\pi^2 \left[\left(\frac{\sigma_1}{h_1} \right)^2 + \left(\frac{\sigma_2}{h_2} \right)^2 \right] \right\}.$$

Часто оказывается необходимым использовать коэффициент корреляции ρ в случае произвольного распределения случайного вектора при естественном и обычно выполняющемся условии существования достаточного числа моментов и гладкости плотности вероятности. На основе описанного выше общего подхода получена следующая приближенная формула для выражения истинного коэффициента корреляции $\rho(Z) = \rho(Z_1, Z_2)$ через «сгруппированный» коэффициент корреляции $\rho(W) = \rho(W_1, W_2)$ и «сгруппированные» дисперсии $D(W_1)$ и $D(W_2)$:

$$\rho(Z_1, Z_2) = \rho(W_1, W_2) \left(1 + \frac{h_1^2}{24D(W_1)} + \frac{h_2^2}{24D(W_2)} - \frac{1}{288} \frac{h_1^2 h_2^2}{D(W_1)D(W_2)} \right) = \rho_1(W_1, W_2).$$

Эта формула уточняет ранее приведенную формулу, полученную элементарными средствами. Оценена относительная погрешность определения коэффициента корреляции:

$$\left| \frac{\rho(Z_1, Z_2) - \rho_1(W_1, W_2)}{\rho(Z_1, Z_2)} \right| \leq R_D + \frac{1}{120} \left(\frac{h_1^2}{D(W_1)} + \frac{h_2^2}{D(W_2)} \right) + \left| \frac{R}{\rho(Z_1, Z_2)\sigma_1\sigma_2} \right|,$$

где R_D – максимум относительных погрешностей в определении истинных дисперсий $D(Z_1)$ и $D(Z_2)$ через «сгруппированные» дисперсии $D(W_1)$ и $D(W_2)$ с использованием поправок Шеппарда, R – абсолютная погрешность при определении истинного коэффициента ковариации через «сгруппированный» (см. выше).

Подводя итоги статьи, следует отметить, что хотя теория статистического анализа сгруппированных данных в настоящее время всё ещё недостаточно развита и систематизирована, принципиальных трудностей на пути расчета поправок на группировку нет. Во многих случаях шаг группировки порядка среднего квадратического отклонения исходных случайных величин дает с учетом поправок вполне приемлемую точность [15]. По сравнению с другими видами погрешностей, в частности, рассматриваемыми в статистике интервальных данных ([1, гл.12]), группировка сравнительно слабо влияет на свойства статистических процедур.

8. Оценки скорости сходимости распределения статистики Крамера – Мизеса - Смирнова

Как уже отмечалось, обобщение формулы Эйлера – Маклорена, в том числе на многомерный случай, было разработано с целью получения асимптотических разложений при изучении скорости сходимости распределения классической статистики омега-квадрат (Крамера-Мизеса-Смирнова) к предельному распределению. Расскажем об этом цикле работ подробнее.

Пусть $F_n(x)$ – эмпирическая функция распределения, построенная по выборке объема n из непрерывной функции распределения $F(x)$. Для проверки гипотезы

$$H_0 : F(x) \equiv F_0(x),$$

где $F_0(x)$ – заданная (однозначно определенная) функция распределения, в работах Г. Крамера [16], Р. фон Мизеса [17] и Н.В. Смирнова [18, 19] была предложена статистика

$$\omega_n^2 = n \int_{-\infty}^{+\infty} (F_n(x) - F_0(x))^2 dF_0(x),$$

которая теперь называется статистикой Крамера – Мизеса – Смирнова (ее свойства рассмотрены в статье [20]). Вслед за Н.В. Смирновым отметим вклад В.И. Гливенко, который предложил заменить интегрирование по dx , как было в первоначальном варианте, на интегрирование по $dF_0(x)$. Такое изменение вида статистики привело к инвариантности ее распределения относительно изменения функции распределения результатов измерений (наблюдений, испытаний, анализов, опытов), т.е. к тому, что при справедливости нулевой гипотезы H_0 распределение статистики ω_n^2 не зависит от вида непрерывной функции $F_0(x)$.

В работах [18, 19] Н.В. Смирнов показал, что функция распределения $S_n(z)$ статистики ω_n^2 сходится при $n \rightarrow \infty$ к некоторой предельной функции распределения $S_0(z)$ (ее вид указан в статье [20]). Следующий этап – оценка скорости сходимости, т.е. скорости убывания максимального расхождения допредельной и предельной функции распределения

$$\Delta_n = \sup_{-\infty < z < \infty} |S_n(z) - S_0(z)|.$$

Первую оценку дан Н.П. Канделаки в 1965 г.:

$$\Delta_n < C(\ln n)^{-1/4}$$

при всех n , где C – абсолютная константа [21].

Степенную оценку скорости сходимости Δ_n к 0 получил В.В. Сазонов. Он в 1968 г. показал [22], что для любого $\varepsilon > 0$ существует константа $c(\varepsilon)$ такая, что при всех n

$$\Delta_n < c(\varepsilon)n^{-1/10+\varepsilon}.$$

В 1969 г. ему удалось [23] заменить $1/10$ на $1/6$.

В том же 1969 г. В.А. Розенкранц доказал [24], что

$$\Delta_n < C(\ln n)^{3/2}n^{-1/5},$$

где C – некоторая абсолютная константа.

В начале 1971 г. автор настоящей статьи предложил метод оценки скорости сходимости на основе обобщения формулы Эйлера – Маклорена на многомерный случай [25], с помощью которого удалось доказать, что для любого $\varepsilon > 0$ существует константа $c(\varepsilon)$ такая, что при всех n

$$\Delta_n < c(\varepsilon)n^{-1/3+\varepsilon}.$$

В середине 1971 г. автором настоящей статьи был разработан принципиально новый метод – «процесс итерации формул» (опубликован в [3, п.6]), с помощью которого удалось заменить $1/3$ на $1/2$, т.е. доказать, что для любого $\varepsilon > 0$ существует константа $c(\varepsilon)$ такая, что при всех n

$$\Delta_n < c(\varepsilon)n^{-1/2+\varepsilon}.$$

Эти результаты были опубликованы в 1972 – 1974 г. – краткие формулировки в работах [26 - 29] и подробное изложение в статье [3].

Надо отметить, что в 1972 г. вышли две работы с более слабыми результатами. Я.Ю. Никитин доказал [30], что

$$\Delta_n < C(\ln n)^{5/2}n^{-1/4},$$

а Дж. Кифер [31] – что

$$\Delta_n < C(\ln n)^{3/2}n^{-1/4}.$$

В этих двух работах 1972 г. показатель степени был $1/4$, т.е. меньше $1/3$ в работе [25] 1971 г.

Более подробно исследования по оценке скорости сходимости распределений классической статистики омега-квадрат (Крамера-Мизеса-Смирнова) и других статистик супремумного и интегрального типов к

предельным распределениям рассмотрены в разделе 2.3 «Оценки скорости сходимости в принципе инвариантности» монографии [32, с.45 - 53].

Таким образом, полученное автором настоящей статьи обобщение формулы Эйлера – Маклорена, в том числе на многомерный случай, оказалось полезным не только при анализе сгруппированных данных, но и при исследованиях, казалось бы, в совсем иной области – оценке скорости сходимости распределений непараметрических статистик.

9. Заключительные замечания

Рассмотренные выше методы анализа сгруппированных данных естественно сопоставить со статистикой интервальных данных [33]. В обоих случаях элементы выборки - интервалы. При анализе сгруппированных данных центры интервалов принадлежат арифметической прогрессии $a_i = a_0 + ih$ (см. начало статьи), а длина равна h . В статистике интервальных данных длина интервалов также фиксирована, однако центрами интервалов являются значения независимых одинаково распределенных случайных величин, т.е. любые действительные числа.

Области практического применения этих двух теорий несколько различаются. Анализ сгруппированных данных нацелен на учет округления исходных выборочных значений, а статистика интервальных данных – на учет погрешностей измерений.

Однако методологические подходы анализа сгруппированных данных и статистики интервальных данных весьма близки, что позволяет отнести их к новому направлению теоретической и вычислительной математики – системной нечеткой интервальной математике [34, 35]

Литература

1. Орлов А.И. Прикладная статистика. – М.: Экзамен, 2006. – 672 с.
2. Крамер Г. Математические методы статистики. – М.: Мир, 1975. – 648 с.

3. Орлов А.И. Скорость сходимости распределения статистики Мизеса-Смирнова // Теория вероятностей и ее применения. 1974. Т.19. № 4. С.766-786.
4. Фихтенгольц Г.М. Курс дифференциального и интегрального исчисления. Т.2. – М.: Наука, 1966. – 800 с.
5. ГОСТ 11.011-83. Прикладная статистика. Правила определения оценок и доверительных границ для параметров гамма-распределения / Орлов А.И., Миронова Н.Г., Невельсон М.Б. и др. – М.: Изд-во стандартов, 1984. – 53 с.
6. Sheppard W.F. On the calculation of the most probable values of frequency constants for data arranged according to equidistant divisions of a scale // Proc. London Math. Soc., 1898, v.29, p.353.
7. Wold H. Sulla correzione di Sheppard // Giorn. Ist. Italiano d. Attari, 1934, v.5, p.304.
8. Кулльдорф Г. Введение в теорию оценивания по группированным и частично группированным выборкам. – М.: Наука, 1966. – 196 с.
9. Бодин Н.А. Об ошибках округления при многомерных измерениях // Труды Математического института им. В.А. Стеклова: Том 79. – М.-Л. Наука, 1965. - С.76-105.
10. Бодин Н.А. Оценка параметров распределения по группированным выборкам // Труды Математического института им. В.А. Стеклова: Том 111. – Л. Наука, 1970. – С.110-154.
11. Бодин Н.А. Об ошибках округления при измерениях с помощью случайно сдвигающейся шкалы // Труды Математического института им. В.А. Стеклова: Том 111. – Л. Наука, 1970. – С.155-162.
12. Орлов А.И. Поправка на группировку для коэффициента корреляции // Экономика и математические методы. – 1980. - № 4. – С.800-801.
13. Свободин В.А., Сирожитдинов И.К. К вопросу о корреляционных моделях по усредненным данным // Экономика и математические методы. – 1980. - № 4. – С.796-800.
14. Орлов А.И., Орловский И.В. О поправках на группировку // Прикладной многомерный статистический анализ. – М.: Наука, 1978. – С.339-342.
15. Орлов А.И. Оценивание для сгруппированных данных // Статистические методы оценивания и проверки гипотез: межвуз. сб. науч. тр. / Перм. гос. нац. иссл. ун-т. – Пермь, 2012. – Вып. 24. – С. 83-95.
16. Cramer H. On the composition of elementary errors // Skand. Aktuarietidskr. 1928. V.11. P. 17 – 34, 141 – 180.
17. Mises R. von. Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistic und theoretischen Physik. Leipzig. Wien. Deutsche. 1931. S.316-335.
18. Smirnov N.V. Sur la distribution de ω^2 (criterium de M.R. v. Mises) // Compt. Rend. Ser. Math. Paris. 1936. V.202. №6. P.449-452.
19. Смирнов Н.В. О распределении ω^2 – критерия Мизеса // Математический сборник. 1937. Т.2 (44). №5. С.973-993.
20. Орлов А.И. Непараметрические критерии согласия Колмогорова, Смирнова, Омега-квадрат и ошибки при их применении / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №03(097). С. 647 – 675. – IDA [article ID]: 0971403047. – Режим доступа: <http://ej.kubagro.ru/2014/03/pdf/47.pdf>, 1,812 у.п.л.
21. Канделаки Н.П. Об одной предельной теореме в пространстве Гильберта // Труды ВЦ АН СССР. 1965. Т.5. №1. С.46-55.

22. Sazonov V.V. On ω_n^2 – criterion // Sankhya. Ser.A. 1968. V.30. N 2. P. 205 – 209.
23. Сазонов В.В. Улучшение одной оценки скорости сходимости // Теория вероятностей и ее применения. 1969. Т.14. №4. С.667-678.
24. Rosenkrantz W.A. A rate of convergence for the von Mises statistic // Trans. Amer. Math. Soc. 1969. V.139. P.329 – 337.
25. Орлов А.И. Оценки скорости сходимости к пределу распределений некоторых статистик // Теория вероятностей и ее применения. 1971. Т.16. №3. С.583-584.
26. Орлов А.И. О проверке симметрии распределения // Теория вероятностей и ее применения. 1972. Т.17. №2. С.372-377.
27. Орлов А.И. Предельные теоремы для статистик интегрального типа // Тезисы докладов Международной конференции по теории вероятностей и математической статистике (Вильнюс, 25-30 июня 1973 г.). Т.2. - Вильнюс: Изд-во Вильнюсского госуниверситета, 1973. С.137-140.
28. Орлов А.И. Переход от сумм к интегралам и его применения в изучении асимптотических распределений статистик // Теория вероятностей и ее применения. 1973. Т.18. №4. С. 881-883.
29. Орлов А.И. Применение критериев типа омега-квадрат для проверки принадлежности функции распределения выборки некоторому семейству // Многомерный статистический анализ в социально-экономических исследованиях. - М.: Наука, 1974. С.401-403.
30. Никитин Я.Ю. Оценки скорости сходимости в некоторых предельных теоремах и статистических критериях // Доклады АН СССР, 1972. Т.202. №4. С.758-760.
31. Kiefer J. Skorohod embedding of multivariate RV's and the sample DF // Z. Wahrscheinlichkeitstheor. und verw. Geb. 1972. V.24. N 1. P.1-35.
32. Орлов А.И. Устойчивость в социально-экономических моделях. – М.: Наука, 1979. – 296 с.
33. Орлов А.И. Основные идеи статистики интервальных данных / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №10(094). С. 867 – 892. – IDA [article ID]: 0941310060. – Режим доступа: <http://ej.kubagro.ru/2013/10/pdf/60.pdf>, 1,625 у.п.л.
34. Орлов А.И. Системная нечеткая интервальная математика (СНИМ) – перспективное направление теоретической и вычислительной математики / А.И. Орлов, Е.В. Луценко // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №07(091). С. 255 – 308. – IDA [article ID]: 0911307015. – Режим доступа: <http://ej.kubagro.ru/2013/07/pdf/15.pdf>, 3,375 у.п.л.
35. Орлов А.И., Луценко Е.В. Системная нечеткая интервальная математика. Монография (научное издание). – Краснодар, КубГАУ. 2014. – 600 с.

References

1. Orlov A.I. Prikladnaja statistika. – М.: Jekzamen, 2006. – 672 s.
2. Kramer G. Matematicheskie metody statistiki. – М.: Mir, 1975. – 648 s.
3. Orlov A.I. Skorost' shodimosti raspredelenija statistiki Mizesa-Smirnova // Teoriya verojatnostej i ee primenenija. 1974. Т.19. № 4. S.766-786.

4. Fihtengol'c G.M. Kurs differencial'nogo i integral'nogo ischislenija. T.2. – M.: Nauka, 1966. – 800 s.
5. GOST 11.011-83. Prikladnaja statistika. Pravila opredelenija ocenok i doveritel'nyh granic dlja parametrov gamma-raspredelenija / Orlov A.I., Mironova N.G., Nevel'son M.B. i dr. – M.: Izd-vo standartov, 1984. – 53 s.
6. Sheppard W.F. On the calculation of the most probable values of frequency constants for data arranged according to equidistant divisions of a scale // Proc. London Math. Soc., 1898, v.29, p.353.
7. Wold H. Sulla correzione di Sheppard // Giorn. Ist. Italiano d. Attari, 1934, v.5, p.304.
8. Kulldorf G. Vvedenie v teoriju ocenivanija po gruppirovannym i chastichno gruppirovannym vyborkam. – M.: Nauka, 1966. – 196 s.
9. Bodin N.A. Ob oshibkah okruglenija pri mnogomernyh izmerenijah // Trudy Matematicheskogo instituta im. V.A. Steklova: Tom 79. – M.-L. Nauka, 1965. - S.76-105.
10. Bodin N.A. Ocenka parametrov raspredelenija po gruppirovannym vyborkam // Trudy Matematicheskogo instituta im. V.A. Steklova: Tom 111. – L. Nauka, 1970. – S.110-154.
11. Bodin N.A. Ob oshibkah okruglenija pri izmerenijah s pomoshh'ju sluchajno sdvigajushhejsja shkaly // Trudy Matematicheskogo instituta im. V.A. Steklova: Tom 111. – L. Nauka, 1970. – S.155-162.
12. Orlov A.I. Popravka na gruppirovku dlja kojefficienta korreljacionnoj zavisimosti // Jekonomika i matematicheskie metody. – 1980. - № 4. – S.800-801.
13. Svobodin V.A., Sirozhitdinov I.K. K voprosu o korreljacionnyh modeljah po usrednennym dannym // Jekonomika i matematicheskie metody. – 1980. - № 4. – S.796-800.
14. Orlov A.I., Orlovskij I.V. O popravkah na gruppirovku // Prikladnoj mnogomernyj statisticheskij analiz. – M.: Nauka, 1978. – S.339-342.
15. Orlov A.I. Ocenivanie dlja sgruppirovannyh dannyh // Statisticheskie metody ocenivanija i proverki gipotez: mezhvuz. sb. nauch. tr. / Perm. gos. nac. issl. un-t. – Perm', 2012. – Vyp. 24. – S. 83-95.
16. Cramer H. On the composition of elementary errors // Skand. Aktuarientidskr. 1928. V.11. P. 17 – 34, 141 – 180.
17. Mises R. von. Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistic und theoretischen Physik. Leipzig. Wien. Deutsche. 1931. S.316-335.
18. Smirnov N.V. Sur la distribution de (criterium de M.R. v. Mises) // Compt. Rend. Ser. Math. Paris. 1936. V.202. №6. P.449-452.
19. Smirnov N.V. O raspredelenii – kriterija Mizesa // Matematicheskij sbornik. 1937. T.2 (44). №5. S.973-993.
20. Orlov A.I. Neparаметрические критерии согласия Колмогорова, Смирнова, Омга-квадрат и ошибки при их применении / A.I. Orlov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2014. – №03(097). S. 647 – 675. – IDA [article ID]: 0971403047. – Rezhim dostupa: <http://ej.kubagro.ru/2014/03/pdf/47.pdf>, 1,812 u.p.l.
21. Kandelaki N.P. Ob odnoj predel'noj teoreme v prostranstve Gil'berta // Trudy VC AN GSSR. 1965. T.5. №1. S.46-55.
22. Sazonov V.V. On – criterion // Sankhya. Ser.A. 1968. V.30. N 2. P. 205 – 209.

23. Sazonov V.V. Uluchshenie odnoj ocenki skorosti shodimosti // Teorija verojatnostej i ee primenenija. 1969. T.14. №4. S.667-678.
24. Rosenkrantz W.A. A rate of convergence for the von Mises statistic // Trans. Amer. Math. Soc. 1969. V.139. P.329 – 337.
25. Orlov A.I. Ocenki skorosti shodimosti k predelu raspredelenij nekotoryh statistik // Teorija verojatnostej i ee primenenija. 1971. T.16. №3. S.583-584.
26. Orlov A.I. O proverke simmetrii raspredelenija // Teorija verojatnostej i ee primenenija. 1972. T.17. No.2. S.372-377.
27. Orlov A.I. Predel'nye teoremy dlja statistik integral'nogo tipa // Tezisy dokladov Mezhdunarodnoj konferencii po teorii verojatnostej i matematicheskoj statistike (Vil'njus, 25-30 ijunja 1973 g.). T.2. - Vil'njus: Izd-vo Vil'njusskogo gosuniversiteta, 1973. S.137-140.
28. Orlov A.I. Pehod ot summ k integralam i ego primenenija v izuchenii asimptoticheskikh raspredelenij statistik // Teorija verojatnostej i ee primenenija. 1973. T.18. №4. S. 881-883.
29. Orlov A.I. Primenenie kriteriev tipa omega-kvadrat dlja proverki prinadlezhnosti funkcii raspredelenija vyborki nekomu semejstvu // Mnogomernyj statisticheskij analiz v social'no-jekonomicheskikh issledovanijah. - M.: Nauka, 1974. S.401-403.
30. Nikitin Ja.Ju. Ocenki skorosti shodimosti v nekotoryh predel'nyh teoremah i statisticheskikh kriterijah // Doklady AN SSSR, 1972. T.202. №4. S.758-760.
31. Kiefer J. Skorohod embedding of multivariate RV's and the sample DF // Z. Wahrscheinlichkeitstheor. und verw. Geb. 1972. V.24. N 1. P.1-35.
32. Orlov A.I. Ustojchivost' v social'no-jekonomicheskikh modeljah. – M.: Nauka, 1979. – 296 s.
33. Orlov A.I. Osnovnye idei statistiki interval'nyh dannyh / A.I. Orlov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2013. – №10(094). S. 867 – 892. – IDA [article ID]: 0941310060. – Rezhim dostupa: <http://ej.kubagro.ru/2013/10/pdf/60.pdf>, 1,625 u.p.l.
34. Orlov A.I. Sistemnaja nechetkaja interval'naja matematika (SNIM) – perspektivnoe napravlenie teoreticheskoj i vychislitel'noj matematiki / A.I. Orlov, E.V. Lucenko // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2013. – №07(091). S. 255 – 308. – IDA [article ID]: 0911307015. – Rezhim dostupa: <http://ej.kubagro.ru/2013/07/pdf/15.pdf>, 3,375 u.p.l.
35. Orlov A.I., Lucenko E.V. Sistemnaja nechetkaja interval'naja matematika. Monografija (nauchnoe izdanie). – Krasnodar, KubGAU. 2014. – 600 s.