

С.С. Вожов, Е.В. Чимитова

ПРОВЕРКА ГИПОТЕЗЫ О ВИДЕ РАСПРЕДЕЛЕНИЯ ПО ИНТЕРВАЛЬНЫМ ДАННЫМ

*Работа выполнена при поддержке Министерства образования и науки РФ
в рамках проектной части государственного задания (№ 2.541.2014К).*

Рассматривается ICM (Iterative Convex Minorant)-алгоритм построения непараметрической оценки функции распределения по интервальным данным. Предлагаются критерии согласия, основанные на расстоянии между предполагаемой функцией распределения и её непараметрической оценкой. Применение критериев согласия опирается на статистическое моделирование условных распределений статистик данных критериев в интерактивном режиме проверки гипотезы о виде распределения.

Ключевые слова: интервальные данные, алгоритм Тёрнбулла, ICM-алгоритм, критерии согласия.

Основы анализа интервальных данных были заложены в метрологии, где интервал неопределенности вводится естественным образом. Предполагается, что каждое наблюдение – это величина, измеренная прибором, который имеет абсолютную ошибку измерения Δ . Таким образом, если точное значение наблюдаемого параметра объекта равно \dot{x} , ошибка измерения равна $e \in [-\Delta, \Delta]$, то точечное наблюдение выглядит как $x = \dot{x} + e$. В этом случае мы имеем дело с точечной выборкой $X_n = \{X_1, \dots, X_n\}$. При проведении статистического анализа на основе точечных выборок мы, по сути, не учитываем величину погрешности измерения Δ . Однако существует и другой подход, который заключается в представлении наблюдения в виде интервала $(x - \Delta, x + \Delta) = (L, R)$, тогда для выборки из n наблюдений получаем интервальную выборку вида

$$I_n = \{(L_1, R_1), \dots, (L_n, R_n)\}.$$

Интервальные наблюдения рассматриваются во многих современных исследованиях, например в [1–8]. В каждой из них аргументируется необходимость развития методов математической статистики, когда результаты наблюдений – не числа, а интервалы.

1. Непараметрическая оценка функции распределения. ICM-алгоритм

Основная идея построения непараметрической оценки функции распределения по интервальным данным заключается в том, что находится максимум логарифма функции правдоподобия

$$\ln L(I_n) = \sum_{i=1}^n \ln(F(R_i) - F(L_i))$$

по значениям функции распределения в граничных точках интервалов наблюдений при соблюдении условия монотонности функции распределения. Однако решение данной оптимизационной задачи методом штрафных функций требует больших вычислительных ресурсов. Вместо этого целесообразно использование специальных алгоритмов. Одним из таких алгоритмов является алгоритм Тёрнбулла [9], подробное описание которого представлено в [5, 10, 11]. Однако данный алгоритм требует значительных вычислительных затрат, поэтому рекомендуется использовать ICM-алгоритм, предложенный в [12–14]. В [10, 11] показано, что время вычисления оценки по ICM-алгоритму существенно меньше, чем по алгоритму Тёрнбулла.

Идея ИСМ-алгоритма заключается в том, чтобы свести задачу максимизации функции правдоподобия к задаче построения изотонической регрессии [12]. Для оценивания функции распределения $F(x)$ вводится разбиение $0 < \tau_0 < \tau_1 < \dots < \tau_m$, состоящее из всех неповторяющихся упорядоченных границ интервалов L_i и R_i , $i = \overline{1, n}$.

В ИСМ-алгоритме для всех $j = \overline{1, m}$ и $i = \overline{1, n}$ определяются веса:

$$\alpha_{ji} = \begin{cases} 1, & \text{если } \tau_j = L_i, \\ -1, & \text{если } \tau_j = R_i, \\ 0, & \text{если } \tau_j \notin (L_i, R_i]. \end{cases}$$

Итерационный процесс ИСМ-алгоритма для произвольного шага алгоритма $k \geq 1$:

1. Для всех $j = \overline{1, m}$ найти точки $(G_j^{(k)}, V_j^{(k)})$ в соответствии со следующими выражениями:

$$G_j^{(k)} = G_{j-1}^{(k)} + \sum_{i=1}^n \frac{1}{|\alpha_{ji}| \cdot \left(\hat{F}^{(k)}(\tau_j) - \hat{F}^{(k)}(\tau_{j-1}) \right)^2},$$

$$W_j^{(k)} = W_{j-1}^{(k)} + \sum_{i=1}^n \frac{1}{\alpha_{ji} \cdot \left(\hat{F}^{(k)}(L_i) - \hat{F}^{(k)}(R_i) \right)},$$

$$D_j^{(k)} = D_{j-1}^{(k)} + \sum_{i=1}^n \frac{\hat{F}^{(k)}(\tau_j)}{|\alpha_{ji}| \cdot \left(\hat{F}^{(k)}(\tau_j) - \hat{F}^{(k)}(\tau_{j-1}) \right)^2},$$

$$V_j^{(k)} = W_j^{(k)} + D_j^{(k)}.$$

2. Установить $l = 0$.

3. Оценка функции распределения равна левой производной выпуклой миноранты:

$$\hat{F}^{(k)}(\tau_j) = \min_{l+1 \leq s \leq m} \left(\frac{V_s^{(k)} - V_l^{(k)}}{G_s^{(k)} - G_l^{(k)}} \right),$$

где $j = \overline{l+1, s}$, s – индекс угловой точки выпуклой миноранты (функция, значение которой не больше соответствующих значений данной функции).

4. Изменить $l = s$. Если $l < m$ переходим к пункту 3.

Алгоритм повторяем до тех пор, пока для всех $j = \overline{1, m}$ не выполнится условие:

$$\left| \hat{F}^{(k)}(\tau_j) - \hat{F}^{(k-1)}(\tau_j) \right| \leq 10^{-7}.$$

2. Критерии согласия для проверки сложной гипотезы о виде распределения по интервальным данным

Пусть имеется априорная информация о том, что выборка наблюдений принадлежит некоторому параметрическому семейству распределений $F_0(x; \theta)$, $\theta \in \Theta$. Для оценивания неизвестных параметров распределения $F_0(x; \theta)$ можно использовать метод максимального правдоподобия, который заключается в максимизации логарифма функции правдоподобия:

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \ln L(\mathcal{I}_n; \theta) = \arg \max_{\theta \in \Theta} \sum_{i=1}^n \ln (F(R_i; \theta) - F(L_i; \theta)).$$

Прежде чем делать выводы на основе полученного распределения, необходимо убедиться в адекватности полученной вероятностной модели. Для этого проверяется сложная гипотеза о виде распределения

$$H_0 : F(t) \in \{F_0(t; \theta), \theta \in \Theta\}.$$

Сложность гипотезы определяется тем, что неизвестные параметры предполагаемого распределения $F_0(x; \theta)$ оцениваются по той же выборке, по которой проверяется гипотеза.

Для проверки гипотезы H_0 по интервальной выборке рассмотрим модификации известных критериев согласия типа Колмогорова, Крамера–Мизеса–Смирнова и Андерсона–Дарлингга. Суть модификации заключается в том, что вместо эмпирической функции распределения в выражениях статистик данных критериев будем использовать непараметрическую оценку функции распределения по интервальной выборке \mathcal{I}_n . В этом случае статистика критерия типа Колмогорова вычисляется следующим образом:

$$D_n = \sup_{0 < t < \tau_m} |\hat{F}_n(t) - F_0(t; \hat{\theta})|,$$

статистика критерия типа Крамера–Мизеса–Смирнова вычисляется по формуле

$$S_{\omega^2} = \int_0^{\tau_m} (\hat{F}_n(t) - F_0(t; \hat{\theta}))^2 dF_0(t; \hat{\theta})$$

и статистика Андерсона–Дарлингга имеет вид

$$S_W = \int_0^{\tau_m} (\hat{F}_n(t) - F_0(t; \hat{\theta}))^2 \frac{dF_0(t; \hat{\theta})}{F_0(t; \hat{\theta})(1 - F_0(t; \hat{\theta}))},$$

где $\hat{F}_n(t)$ – непараметрическая оценка функции распределения по интервальным данным, которая вычисляется с помощью ИСМ-алгоритма.

Гипотеза H_0 отвергается при больших значениях статистик. Аналитический вид распределений рассматриваемых статистик при справедливости нулевой гипотезы неизвестен. При этом достигнутые уровни значимости (p -value) можно оценить с использованием методов статистического моделирования. Для этого необходимо задать количество повторений N . Порядок проверки гипотезы о виде распределения в этом случае имеет следующий вид:

1. Установить уровень значимости α .
2. По исходной интервальной выборке вычислить оценку неизвестного параметра θ распределения $F_0(t; \theta)$.
3. В соответствии с выбранным критерием согласия вычислить значение статистики критерия $S(\mathcal{I}_n)$.
4. Положить $M = 0$ и повторить следующую последовательность действий N раз:

4.1) сгенерировать интервальную выборку $\mathcal{Z}_n = \{(L_1, R_1), \dots, (L_n, R_n)\}$:

$$(L_i, R_i) = \left(X_i - \frac{\Delta}{2}, X_i + \frac{\Delta}{2} \right), X_i = F_0^{-1}(\xi_i; \hat{\theta}), \xi_i \in \text{Uniform}(0, 1),$$

где $\hat{\theta}$ – оценка параметра θ , полученная в п. 2, Δ – абсолютная ошибка измерения в исходной выборке \mathcal{I}_n ;

- 4.2) по полученной выборке \mathcal{Z}_n оценить параметры распределения $F_0(t; \theta)$;
- 4.3) вычислить значение статистики критерия по смоделированной выборке;
- 4.4) если $S(\mathcal{Z}_n) > S(\mathcal{I}_n)$, то увеличить M на 1.

5. Вычислить оценку достигнутого уровня значимости $p = \frac{M}{N}$.

6. Если $p > \alpha$, то гипотеза H_0 не отвергается.

Важнейшим свойством критериев проверки статистических гипотез является мощность критерия. Чем выше мощность критерия, тем меньше вероятность допустить ошибку второго рода (вероятность принять гипотезу H_0 , когда она не верна). Целью исследования мощности критериев согласия может быть не только выявление наиболее мощного среди них, но и сравнение величины мощности критериев

по интервальным данным с аналогичной величиной мощности критериев по точечным данным. Точечная выборка содержит больше информации о распределении, чем интервальная, так как точечное наблюдение – это, по сути, измерение с нулевой погрешностью Δ (что на практике, как правило, невозможно). Это означает, что мощность критериев согласия для точечных выборок будет выше мощности соответствующих модифицированных критериев для интервальных данных. Однако то, насколько велики эти потери, неизвестно.

Для исследования мощности непараметрических критериев согласия по точечным и интервальным выборкам рассматривалась пара близких конкурирующих гипотез: H_0 , соответствующая нормальному распределению (сложная гипотеза), против H_1 , соответствующей логистическому распределению с параметрами $\theta_1 = 5$, $\theta_2 = \frac{\sqrt{3}}{\pi}$. В табл. 1 представлены оценки мощности критериев типа Колмогорова, Крамера–Мизеса–Смирнова и Андерсона–Дарлинга для точечных и интервальных выборок наблюдений объёмом $n = 50, 100, 200, 300, 500$. Оценки мощности вычислялись по выборкам статистик объёма $N = 20\,000$ при уровне значимости $\alpha = 0,1$. Функции плотности рассматриваемых распределений представлены в табл. 2.

Т а б л и ц а 1

Оценки мощности критериев согласия по точечным и интервальным данным

n	D_n		S_{ω^2}		S_W	
	X_n	I_n	X_n	I_n	X_n	I_n
50	0,18	0,17	0,21	0,21	0,24	0,23
100	0,23	0,21	0,29	0,29	0,33	0,32
200	0,36	0,31	0,46	0,45	0,49	0,49
300	0,45	0,36	0,58	0,58	0,67	0,62
500	0,65	0,50	0,79	0,78	0,89	0,82

Из табл. 1 видно, что среди рассмотренных критериев наиболее мощным является критерий Андерсона–Дарлинга как для точечных, так и для интервальных выборок. Как и ожидалось, мощность для всех видов критериев выше для точечных выборок. При этом эта разница увеличивается с ростом объёма выборки.

3. Пример проверки гипотезы о виде распределения по интервальной выборке

Рассмотрим выборку цен потребительского спроса на биоэнергетический напиток НПК «САВА», которые были получены путём анкетирования потенциальных потребителей продукции НПК «САВА» [5]. Участникам предлагалось указать диапазон для цены, которую они готовы заплатить за биоэнергетический напиток. В результате были получены интервальные данные по ценовым предпочтениям. В [5] по полученной выборке с помощью алгоритма Тёрнбулла была вычислена оценка функции распределения. В данной работе подберем параметрическое распределение, которое достаточно хорошо описывает распределение наблюдаемой случайной величины. В табл. 2 представлены результаты оценки параметров ряда распределений методом максимального правдоподобия и соответствующие значения информационных критериев Акаике (AIC) и Шварца (BIC), позволяющих сравнивать полученные модели:

$$AIC = 2k - 2l, \quad BIC = k \ln n - 2l,$$

где k – количество оцененных параметров, l – значение логарифмической функции правдоподобия для построенной модели. Чем меньше значение критерия, тем лучше модель.

На основе полученных результатов по критериям Акаике и Шварца в качестве распределения, описывающего цену потребительского спроса, выбрано нормальное распределение. На рис. 1 представлены графики непараметрической оценки функции распределения по ICM-алгоритму и функции распределения нормального закона с оцененными параметрами.

Выбор распределения

Распределение	Функция плотности	Оценки параметров	l	BIC	AIC
Нормальное	$\frac{1}{\theta_2 \sqrt{2\pi}} \exp\left(-\frac{(x-\theta_1)^2}{2\theta_2^2}\right)$	(38,88; 12,48)	-66,59	139,69	137,18
Логистическое	$\frac{\exp\left(-\frac{x-\theta_1}{\theta_2}\right)}{\theta_2 \left(\exp\left(-\frac{x-\theta_1}{\theta_2}\right) + 1\right)^2}$	(38,99; 7,69)	-67,91	142,34	139,82
Логнормальное	$\frac{1}{\sqrt{2\pi}(x-\theta_1) \cdot \theta_3} \exp\left(-\frac{1}{2} \left(\frac{1}{\theta_3} \ln\left(\frac{x-\theta_1}{\theta_2}\right)\right)^2\right)$	(16; 18,50; 0,74)	-69,72	149,22	145,44
Гамма	$\frac{1}{\theta_2 \cdot \Gamma(\theta_3)} \left(\frac{x-\theta_1}{\theta_2}\right)^{\theta_3-1} \exp\left(-\frac{x-\theta_1}{\theta_2}\right)$	(16; 9,78; 2,35)	-67,29	144,37	140,59
Экспоненциальное	$\frac{1}{\theta_2} \exp\left(-\frac{x-\theta_1}{\theta_2}\right)$	(16; 22,83)	-71,32	149,15	146,63
Вейбулла	$\frac{\theta_3}{x-\theta_1} \left(\frac{x-\theta_1}{\theta_2}\right)^{\theta_3} \exp\left(-\left(\frac{x-\theta_1}{\theta_2}\right)^{\theta_3}\right)$	(16; 25,65; 1,85)	-66,15	142,07	138,29
Логлогистическое	$\frac{\theta_3 \left(\frac{x-\theta_1}{\theta_2}\right)^{\theta_3-1}}{\theta_2 \left(1 + \left(\frac{x-\theta_1}{\theta_2}\right)^{\theta_3}\right)^2}$	(16; 20,08; 2,36)	-69,66	149,09	145,31

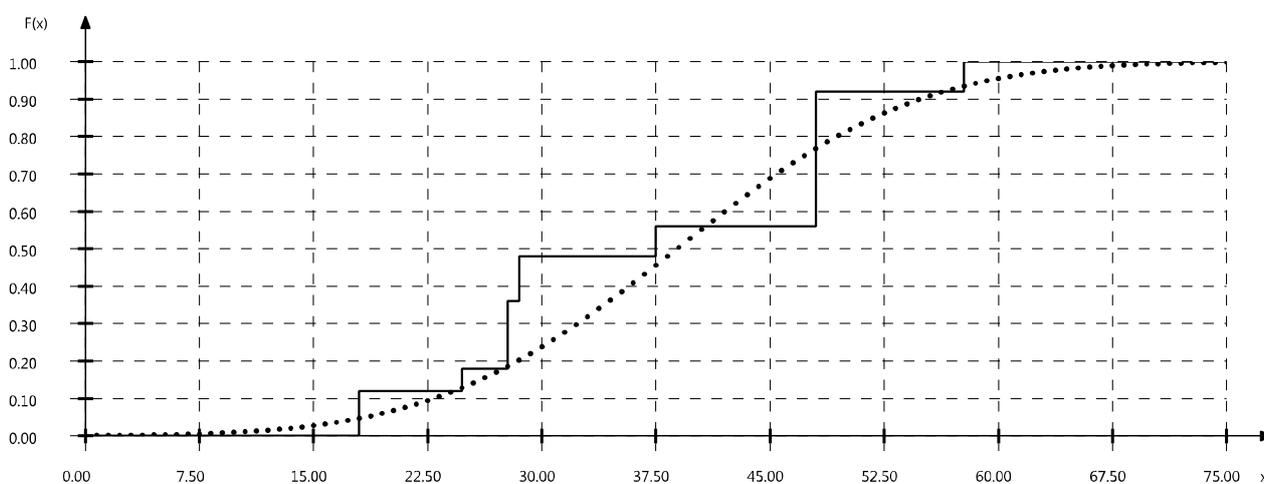


Рис. 1. Непараметрическая оценка функции распределения и нормальное распределение (пунктирная линия – нормальное распределение, сплошная – оценка функции распределения)

Информационные критерии позволяют сравнивать различные вероятностные модели, однако они не гарантируют адекватность выбранной модели. Для обоснования выбора модели необходима проверка статистической гипотезы о виде распределения.

Для проверки гипотезы о принадлежности рассматриваемой выборки нормальному распределению используем предложенные в данной работе модификации критериев согласия. Зададимся уровнем значимости $\alpha = 0,05$. В соответствии с представленным выше порядком проверки гипотезы о виде распределения при $N = 20\,000$ получены следующие оценки достигнутого уровня значимости (p -value): для критерия типа Колмогорова $p = 0,1853$, для критерия типа Крамера–Мизеса–Смирнова $p = 0,2786$ и для критерия типа Андерсона–Дарлингга $p = 0,3423$. Поскольку по всем критериям p -value больше

заданного уровня значимости, то нет оснований для отвержения гипотезы о принадлежности выборки нормальному распределению.

Заключение

В данной работе предложены модификации критериев согласия Колмогорова, Крамера–Мизеса–Смирнова и Андерсона–Дарлинга для проверки гипотез о виде распределения по интервальным данным. Применение предложенных критериев предполагает использование методов статистического моделирования для оценки достигнутого уровня значимости (p -value) для конкретной гипотезы при заданных значениях объема выборки и длины интервалов наблюдений. Сформулирован порядок проверки гипотезы о виде распределения по интервальным данным.

В результате исследования мощности предложенных критериев в качестве наиболее предпочтительного можно рекомендовать критерий согласия типа Андерсона–Дарлинга.

На основе данных о ценах потребительского спроса на биоэнергетический напиток НПК «САВА» проверена гипотеза о принадлежности данной выборки нормальному распределению.

ЛИТЕРАТУРА

1. Kreinovich V. Interval computations and interval-related statistical techniques: estimating uncertainty of the results of data processing and indirect measurements // *Advanced Mathematical and Computational Tools in Metrology and Testing X*. Singapore : World Scientific, 2015. P. 38–49. (Book series: *Advances in Mathematics for Applied Sciences*. V. 86). DOI: 10.1142/9789814678629_0014.
2. Вошинин А.П. Интервальный анализ данных: развитие и перспективы // *Заводская лаборатория*. 2002. Т. 68, № 1. С. 118–126.
3. Вошинин А.П. Метод анализа данных с интервальными ошибками в задачах проверки гипотез и оценивания параметров неявных линейно параметризованных функций // *Заводская лаборатория*. 2000. Т. 66, № 3. С. 51–64.
4. Орлов А.И. Основные идеи статистики интервальных данных // *Научный журнал КубГАУ*. 2013. № 94 (10). С. 1–26.
5. Зенкова Ж.Н., Краковецкая И.В. Непараметрическая оценка Тёрнбулла для интервально-цензурированных данных в маркетинговом исследовании спроса на биоэнергетические напитки // *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*. 2013. № 3 (24). С. 64–69.
6. Лемешко Б.Ю., Постовалов С.Н. Об оценивании параметров распределений по интервальным наблюдениям // *Вычислительные технологии*. 1998. Т. 3, № 2. С. 31–38.
7. Лемешко Б.Ю., Постовалов С.Н. О решении задач статистического анализа интервальных наблюдений // *Вычислительные технологии*. 1997. Т. 2, № 1. С. 28–36.
8. Лемешко Б.Ю., Постовалов С.Н. Статистический анализ наблюдений, имеющих интервальное представление // *Сборник научных трудов НГТУ*. Новосибирск : Изд-во НГТУ, 1996. № 1. С. 3–12.
9. Turnbull B.W. Nonparametric estimation of a survivorship function with doubly-censored data // *J. Am. Statist. Assoc.* 1974. V. 69. P. 169–73.
10. Вожов С.С. Исследование свойств непараметрической оценки функции распределения по интервальным данным // *Сборник научных трудов НГТУ*. Новосибирск : Изд-во НГТУ, 2015. № 1 (79). С. 33–44.
11. Vozhov S., Chimitova E. Investigation of Maximum Likelihood Estimates and Goodness-of-Fit Tests for Data with Known Measurement Error // *Applied methods of statistical analysis. Applications in survival analysis, reliability and quality control. AMSA'2015, Novosibirsk, 14–19 Sept. 2015 : proc. of the intern. workshop. Novosibirsk : NSTU publ., 2015. P. 124–130.*
12. Groeneboom P. Asymptotics for interval censored observations // *Technical Report 87-18*. Department of Mathematics, University of Amsterdam, 1987. 69 p.
13. Groeneboom P. Nonparametric maximum likelihood estimation for interval censored data // *Technical Report*, Statistics Department, Stanford University, 1991. 87 p.
14. Groeneboom P., Wellner J.A. *Information Bounds and Nonparametric Maximum Likelihood Estimation*. Basel : Birkhauser Verlag, 1992. 126 p.

Вожов Станислав Сергеевич. E-mail: vss920414@gmail.com

Чимитова Екатерина Владимировна, канд. техн. наук, доцент. E-mail: chimitova@corp.nstu.ru
Новосибирский государственный технический университет

Поступила в редакцию 3 декабря 2015 г.

Vozhov Stanislav S., Chimitova Ekaterina V. (Novosibirsk State Technical University, Russian Federation).

Testing goodness-of-fit with interval data.

Keywords: interval data; nonparametric estimation of distribution function; Turnbull algorithm; ICM-algorithm.

DOI: 10.17223/19988605/34/4

The main terms of interval data analysis was initially founded the measurement theory in metrology where an interval uncertainty is naturally introduced. It is expected that every observation is measured by an instrument with absolute error Δ . Thus, if the precise value of an observed response is \dot{x} , measurement error is $e \in [-\Delta, \Delta]$, then the measurement is equal to $x = \dot{x} + e$. In this case, we deal with a usual complete sample $X_n = \{X_1, \dots, X_n\}$. Nevertheless, the measurement can be represented as an interval $(x - \Delta, x + \Delta) = (L, R)$. In this case, for the sample of n observations we obtain an interval sample of the form

$$\mathcal{I}_n = \{(L_1, R_1), \dots, (L_n, R_n)\}.$$

The main idea of nonparametric estimation of the distribution function with interval data is based on maximization of the loglikelihood function

$$\ln L(\mathcal{I}_n) = \sum_{i=1}^n \ln(F(R_i) - F(L_i))$$

at the boundary points of observations $L_i, R_i, i = \overline{1, n}$, under condition of monotonicity of the distribution function.

The Turnbull and ICM algorithms are used for calculation of the nonparametric estimate of the distribution function with interval data. The accuracy of the estimates calculation is the same for both algorithm, but the computing time is less for the ICM algorithm.

Unknown distribution parameters can be estimated by the maximum likelihood method, which is based on the maximization of likelihood function by parameter θ :

$$L(\mathcal{I}_n | \theta) = \prod_{i=1}^n (F(R_i | \theta) - F(L_i | \theta)).$$

Thus, the maximum likelihood estimates can be written as

$$\hat{\theta} = \arg \max_{\theta \in \Theta} \ln L(\mathcal{I}_n | \theta).$$

In this paper, the modifications of the classical goodness-of-fit tests for composite hypothesis $H_0: F(t) \in \{F_0(t; \theta), \theta \in \Theta\}$ have been proposed. The main idea of this modification is based on the usage of nonparametric estimate of the distribution function, obtained by the ICM algorithm, instead of the empirical distribution function. In this case, we have the test statistic of the Kolmogorov type as

$$D_n = \sup_{0 < t < \tau_m} |\hat{F}_n(t) - F_0(t, \hat{\theta})|,$$

the statistic of Cramer-von Mises-Smirnov type test as

$$S_{\omega^2} = \int_0^{\tau_m} (\hat{F}_n(t) - F_0(t, \hat{\theta}))^2 dF_0(t, \hat{\theta}),$$

and the statistic of Anderson-Darling type test as

$$S_W = \int_0^{\tau_m} (\hat{F}_n(t) - F_0(t, \hat{\theta}))^2 \frac{dF_0(t, \hat{\theta})}{F_0(t, \hat{\theta})(1 - F_0(t, \hat{\theta}))},$$

where $\hat{F}_n(t)$ is the nonparametric estimate of the distribution function by the interval data, $0 < \tau_0 < \tau_1 < \dots < \tau_m$ are ordered different values L_i and $R_i, i = \overline{1, n}$.

We have formulated the sequence of steps for estimation of the p -value for the proposed tests. The hypothesis is not rejected if the obtained p -value is larger than the significance level α .

As an example, we have tested the normality hypothesis by the interval sample of consumer demand prices for bio-energy drink SPC "SAVA". It has been shown that there is no reason for rejecting the hypothesis of normality of consumer demand prices.

REFERENCES

1. Kreinovich, V. (2015) Interval computations and interval-related statistical techniques: estimating uncertainty of the results of data processing and indirect measurements. In: Pavese, P., Bremser, W., Chunovkina, A., Fischer, N. & Forbes, A. (eds) *Advanced Mathematical and Computational Tools in Metrology and Testing X*. Singapore: World Scientific, pp. 38–49. (Book series: Advances in Mathematics for Applied Sciences, vol. 86). DOI 10.1142/9789814678629_0014.
2. Voshchinin, A.P. (2002) Interval'nyy analiz dannykh: razvitie i perspektivy [Interval data analysis: development and prospects]. *Zavodskaya laboratoriya*. 68(1). pp. 118-126.

3. Voshchinin, A.P. (2000) Metod analiza dannykh s interval'nymi oshibkami v zadachakh proverki gipotez i otsenivaniya parametrov neyavnykh lineynno parametrizovannykh funktsiy [A method for data analysis with interval errors in problems of hypothesis testing and parameter estimation of fussy linear parameterized functions]. *Zavodskaya laboratoriya*. 66(3). pp. 51–64.
4. Orlov, A.I. (2013) Osnovnye idei statistiki interval'nykh dannykh [Basic ideas of interval data statistics]. *Nauchnyy zhurnal KubGAU – Scientific Journal of KubSAU*. 94(10). pp. 1-26.
5. Zenkova, Zh.N. & Krakovetskaya, I.V. (2013) Nonparametric Turnbull estimator for interval censored data in the marketing research of the demand of bio-energy drinks. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 3(24). pp. 64-69. (In Russian).
6. Lemeshko, B.Yu. & Postovalov, S.N. (1998) Ob otsenivanii parametrov raspredeleniy po interval'nym nablyudeniya [On estimation of distribution parameters by interval observations]. *Vychislitel'nye tekhnologii*. 3(2). pp. 31-38.
7. Lemeshko, B.Yu. & Postovalov, S.N. (1997) O reshenii zadach statisticheskogo analiza interval'nykh nablyudeniya [On solving the problems of statistical analysis of interval data]. *Vychislitel'nye tekhnologii*. 2(1). pp. 28-36.
8. Lemeshko, B.Yu. & Postovalov, S.N. (1996) Statisticheskii analiz nablyudeniya, imeyushchikh interval'noe predstavlenie [Statistical analysis of interval observations]. *Sbornik nauchnykh trudov NGTU*. 1. pp. 3-12.
9. Turnbull, B.W. (1974) Nonparametric estimation of a survivorship function with doubly-censored data. *J. Am. Statist. Assoc.* 69. pp. 169-73. DOI: 10.1080/01621459.1974.10480146
10. Vozhov, S.S. (2015) Issledovanie svoystv neparametricheskoy otsenki funktsii raspredeleniya po interval'nym dannym [Investigation of the properties of nonparametric estimate for distribution function with interval data]. *Sbornik nauchnykh trudov NGTU*. 1(79). pp. 33-44.
11. Vozhov, S. & Chimitova, E. (2015) Investigation of Maximum Likelihood Estimates and Goodness-of-Fit Tests for Data with Known Measurement Error. *Applied methods of statistical analysis. Applications in survival analysis, reliability and quality control*. AMSA'2015, Novosibirsk, 14–19 Sept. 2015. Proc. of the Intern. Workshop. Novosibirsk: NSTU. pp. 124-130.
12. Groeneboom, P. (1987) *Asymptotics for interval censored observations*. Technical Report 87-18. Department of Mathematics, University of Amsterdam.
13. Groeneboom, P. (1991) *Nonparametric maximum likelihood estimation for interval censored data*. Technical Report, Statistics Department, Stanford University, 1991. 87 p.
14. Groeneboom, P. & Wellner, J.A. (1992) *Information Bounds and Nonparametric Maximum Likelihood Estimation*. Basel: Birkhauser Verlag.