МЕТОДИКА КЛАССИФИКАЦИИ ОБЪЕКТОВ В УСЛОВИЯХ СТОХАСТИЧЕСКОЙ НЕОПРЕДЕЛЕННОСТИ

Ручкин С.В.[1]
(ВА РВСН *имени Петра Великого*, *Балашиха*)

В настоящей статье представлена постановка и методика решения актуальной задачи классификации вероятностно распределённых объектов в условиях стохастической неопределенности об истинных значениях их характеристик. Методика основана на проверке статистической гипотезы о принадлежности математического ожидания случайной величины заданному диапазону методом Неймана-Пирсона. Отличительной особенностью представленной методики является уточненный порядок расчета критических точек, на базе которых формируется более эффективное решающее правило классификации объектов, а также возможность оценивания ожидаемых значений вероятностей ошибок, которые могут быть допущены в результате проведения процедуры распознавания по отношению ко всем предъявленным объектам.

Ключевые слова: классификация, ошибка первого и второго рода, распознавание, статистическая гипотеза

1. *Введение*

Одними из приоритетных целей развития государства, обозначенных для правительства Президентом Российской Федерации, являются ускорение технологического развития страны и обеспечение ускоренного внедрения цифровых технологий в экономике и социальной сфере [14].

В настоящее время цифровые технологии находят свое применение практически во всех сферах человеческой деятельности, на их основе создаются передовые технические разработки, сложные организационно-технические системы, обосновываются рациональные способы их применения. Необходимость ускоренного внедрения цифровых технологий в конечном итоге обуславливает актуальность постановки и решения при-

[1]*Ручкин Сергей Викторович, к.т.н.,* (*pochtampt07@mail.ru*).

кладных задач, имеющих базовое значение при выполнении научных исследований и поиске ответов на конкретные практические вопросы. К числу таких можно отнести задачу, связанную с процедурами классификации объектов в условиях, характеризуемых различной степенью неопределенности.

Важность этой задачи объясняется двумя причинами. Первая связана с тем, что процедура классификации является первичной формой анализа, систематизации знаний, фундаментальным методом проведения научных исследований, в соответствии с которым вся область изучаемых объектов представляется в виде системы классов, или групп, по которым эти объекты распределены на основании их сходства в определенных свойствах [13]. Вторая причина заключается в том, что неопределенность является объективным свойством внешней среды, связанным с отсутствием или недостатком информации о ее состоянии [16]. И именно в этих условиях человек чаще всего осуществляет свою деятельность, а также вынужден принимать важные для себя решения.

Неопределённость как свойство может быть стохастической, поведенческой и природной [1]. Для каждого вида неопределенности постановка задачи классификации будет иметь свои существенные особенности. В дальнейшем будем рассматривать постановку задачи в условиях стохастической неопределенности, в качестве которой будем понимать неопределенность, связанную с результатами протекания случайных явлений, действие которых может быть описано какими либо известными вероятностными распределениями.

Степень проявления определенного свойства в объекте, которая выражается значением связанного с ним показателя (характеристики), является одним из оснований для его дальнейшей классификации. Если результат проявления свойства объекта является случайной величиной (X), то показателем, по которому может быть осуществлена классификация объекта в условиях стохастической неопределённости, может быть выбрано математическое ожидание этой случайной величины $M[X]$, вычисление статистической оценки которого $M^*[X]$

возможно на основе выборочного метода исследования генеральной совокупности предъявляемых объектов [3].

Таким образом, классификацию объектов в условиях стохастической неопределенности можно рассматривать в рамках постановки и решения задачи по проверке статистической гипотезы о принадлежности математического ожидания случайной величины некоторому, наперед заданному количественному или качественному диапазону значений.

Постановка и решение этой задачи может быть выполнена на базе известных статистических методов путем их несложной доработки. Например, на основе методов проверки простой статистической гипотезы о превосходстве (не превосходстве) математического ожидания случайной величины некоторой заданной постоянной. К числу таких методов, предполагающих статистическую обработку выборки фиксированного объема, относятся [8,9,15,17,18]:

метод на основе использования критерия Байеса, позволяющий минимизировать средний риск от принятого решения, при наличии априорной информации о распределении случайной величины и о величине потерь для каждой возможной пары: «начальное условие – принимаемое решение»;

метод на основе использования минимаксного критерия, гарантирующий получение среднего риска от принятого решения не выше некоторой величины, при отсутствии априорной информации о распределении случайной величины и при наличии информации о величине потерь для каждой возможной пары: «начальное условие – принимаемое решение»;

метод на основе использования критерия Неймана-Пирсона, обеспечивающий величину ошибок от принятого решения на заданном уровне, при наличии априорной информации о распределении случайной величины.

Важными моментами в этой «доработке» является переход от простой гипотезы к сложной и необходимость учета вероятностного распределения исходного множества классифицируемых объектов для обоснованного расчета ожидаемых вероятностей, характеризующих величину ошибок, которые могут быть допущены при проведении процедуры классификации всего множества объектов.

2. *Постановка задачи классификации объектов в условиях стохастической неопределенности*

Рассматривается исходное множество U, состоящее из объектов u, которые могут быть охарактеризованы по одному единственному параметру x. Примем, что множество U является бесконечным. Вероятностное распределение объектов $f(u/x)$ по значению параметра x задано, причем

$$\int_{x_{\min}}^{x_{\max}} f(u/x)dx = 1, \tag{1}$$

где $x_{\min}$, $x_{\max}$ – границы интервала возможных значений параметра x (рис. 1).

Все множество объектов разделено на два подмножества U_0 и U_1 по заданному критерию K:

$$K: \begin{cases} u \in U_0 : x_1 \le x_u \le x_2, \\ u \in U_1 : x_{\min} \le x_u < x_1 - \Delta_1 \vee x_2 + \Delta_2 < x_u \le x_{\max}, \end{cases} \tag{2}$$

где Δ_1, Δ_2 – параметры, характеризующие величину области неразличимости, в которой расположены объекты по отношению к которым безразлично куда их можно отнести, в подмножество U_0 или U_1;

x_u – значение параметра x u-го объекта.

Значение величины Δ, определяет строгость критерия (2) и не превышает длины интервалов, соответствующих подмножеству U_1: $0 \le \Delta_2 < (x_{\max} - x_2)$ и $0 \le \Delta_1 < (x_1 - x_{\min})$.

Из множества U случайным образом выбирается один объект u и определяется значение его параметра x путем проведения серии измерений. Результат многократного измерения X_u есть статистическая оценка математического ожидания (выборочное среднее арифметическое) параметра x. Результат измерения X_u является случайной величиной. По объективным причинам, связанными, например, с возможностями измерителя и предельными характеристиками анализируемых объектов, результаты измерения X_u параметра x могут принадлежать только интервалу $X_u \in [\underline{x}; \overline{x}]$, причем: $\underline{x} \le x_{\min}$, $\overline{x} \ge x_{\max}$. В этой связи, принимается, что результат измерения X_u описываются нормальным усеченным законом распределения с математическим ожиданием x_u и средним квадратическим отклонением (СКО) равным σ: $X_u := N_{\text{Ус}}(x_u; \sigma)$. Величина СКО σ зависит от числа и точности проведенных измерений.

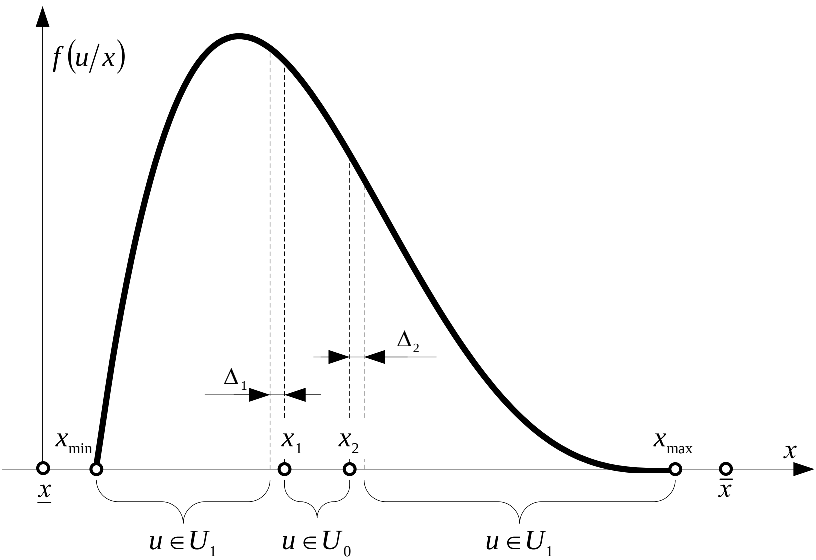

Рис. 1. Распределение объектов множества U по значению параметра x (*форма закона распределения представлена для примера*)

По результату измерения X_u проверятся основная гипотеза H_0 о принадлежности u-го объекта U_0-у подмножеству. Если основная гипотеза не подтверждается, то принимается альтернативная гипотеза H_1 о принадлежности u-го объекта U_1-у подмножеству. В силу стохастического характера результата измерения X_u при проверке основной гипотезы могут быть допущены ошибки двух типов:

$u/U_0 \to U_1$ – u-й объект, фактически принадлежащий подмножеству U_0, был классифицирован как объект принадлежащий подмножеству U_1 (ошибка первого рода:) [7];

$u/U_1 \to U_0$ – u-й объект, фактически принадлежащий подмножеству U_1, был классифицирован как объект принадлежащий подмножеству U_0 (ошибка второго рода) [7].

Далее эта процедура проводится для остальных объектов множества U.

Для рассмотренного комплекса условий необходимо составить математические формулы позволяющие:

по результату измерения X_u проверить основную гипотезу H_0 о принадлежности u-го объекта U_0-у подмножеству, при условии, что $P(u/U_0 \to U_1) \le \alpha$, где α – наперед заданное предельно допустимое значение ошибки первого рода (частная задача 1, которую в дальнейшем будем называть задачей классификации объектов);

определить ожидаемое значение ошибки второго рода $M[\beta]$, которая будет достигнута при проверке всех объектов множества U, при условии, что для u-го объекта принадлежащего U_1-у подмножеству ошибка второго рода β_u численно равна вероятности события $P(u/U_1 \to U_0)$ (частная задача 2);

определить ожидаемое значение ошибки первого рода $M[\alpha]$, которая будет достигнута при проверке всех объектов множества U (частная задача 3).

В кратком, формализованном виде поставленная задача может быть представлена кортежем:

$$(3)\ \left\langle \begin{array}{c} U,\ f(u/x), x \in [x_{\min};\ x_{\max}], X_u \in [\underline{x}; \bar{x}], X_u := N_{Yc}(x_u;\sigma), K, \alpha; \\ H_0, M[\beta], M[\alpha] \end{array} \right\rangle$$

Разработанное решение взаимосвязанных между собой частных задач необходимо представить в виде методики, которую назовем методикой классификации множества объектов в условиях стохастической неопределенности.

3. *Содержание методики классификации объектов в условиях стохастической неопределенности*

Необходимость выполнения требования по недопущению превышения допустимого значения вероятности ошибки первого рода при классификации отдельного объекта определяет метод решения частной задачи 1. Данный метод основан на использовании критерия Неймана-Пирсона [9], который позволяет провести проверку статистической гипотезы с учетом заданного значения ошибки первого, а в некоторых случаях и заданного значения ошибки второго рода (которые могут быть обоснованы на основе информации о возможных потерях от принятия неверных решений, и прибыли от принятия верных решений).

В соответствии с данным методом сложная гипотеза H_0 для u-го объекта принимается при выполнении неравенства

(4) $$H_0 : x_1^{\text{кр}} \le X_u \le x_2^{\text{кр}},$$

если это неравенство не выполняется, то принимается альтернативная гипотеза H_1, т.е:

(5) $$H_1 : x_{\min} \le X_u < x_1^{\text{кр}} \vee x_2^{\text{кр}} < X_u \le x_{\max}.$$

В неравенствах (4) и (5) величины $x_1^{\text{кр}}$ и $x_2^{\text{кр}}$ есть критические значения параметра x, которые с учетом стохастической неопределенности результата его измерения, разделяют все множество возможных значений параметра на две непересекающиеся области, принадлежащих объектам подмножеств U_0 и U_1.

Критические значения $x_1^{\text{кр}}$ и $x_2^{\text{кр}}$ параметра x определяются с помощью квантилей нормального усеченного закона распределения результата его измерения, вычисленных на границах интервала $[x_1; x_2]$ (рис. 2):

(6) $$x_1^{\text{кр}} = F_{\text{Ус}}^{(-1)}(\alpha_1; x_1; \sigma),\ x_2^{\text{кр}} = F_{\text{Ус}}^{(-1)}(1-\alpha_2; x_2; \sigma),$$

где $F_{\text{Ус}}^{(-1)}(\alpha_1; x_1; \sigma)$ и $F_{\text{Ус}}^{(-1)}(1-\alpha_2; x_2; \sigma)$ – квантили нормального усеченного закона распределения с математическим ожиданием равным x_1 и x_2 и СКО равным σ, определенные, соответственно, для вероятностей α_1 и $(1-\alpha_2)$;

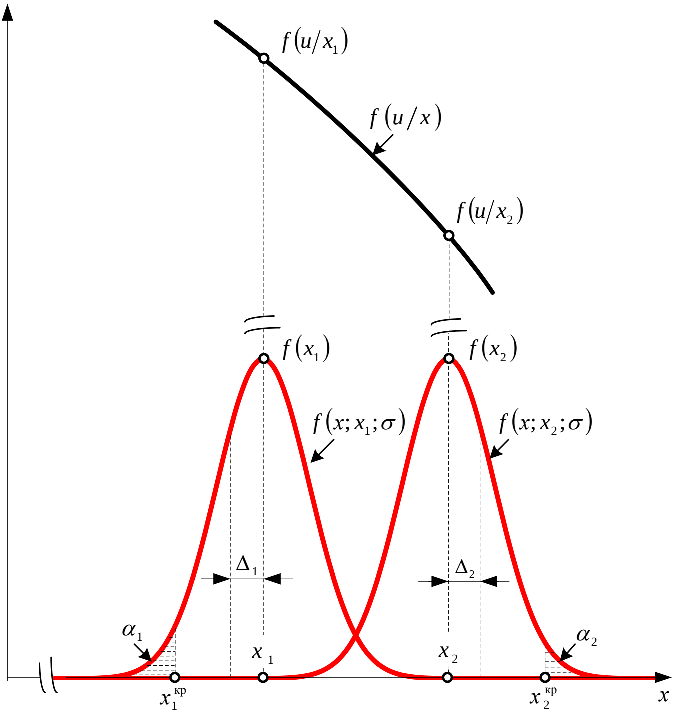

Рис. 2. К определению критических значений $x_1^{кр}$ *и* $x_2^{кр}$ *параметра x (на рисунке* $f(x;x_1;\sigma)$ *и* $f(x;x_2;\sigma)$[1] *– плотности нормального закона распределения результата измерения значений* x_1 *и* x_2 *с СКО равным* σ*)*

$\alpha_1 = k_1\alpha$ и $\alpha_2 = k_2\alpha$ – слагаемые вероятности ошибки первого рода, $\alpha_1 + \alpha_2 = \alpha$;

k_1, k_2 – коэффициенты, учитывающие неравномерность распределения объектов множества U по значению параметра x ($0 \le k_1, k_2 \le 1$, $k_1 + k_2 = 1$).

[1] - *далее по тексту записи вида* $f(x; m_x; \sigma)$ *и* $F(x; m_x; \sigma)$ *будут соответственно обозначать плотность и функцию нормального закона распределения с математическим ожиданием и СКО равным* m_x *и* σ*, определенные для аргумента x.*

Определение квантилей нормального усеченного закона может быть установлено через соответствующие квантили нормального закона [4,10]:

(7) $$F_{\text{Ус}}^{(-1)}(\alpha_1; x_1; \sigma) = F^{(-1)}(\alpha_1^0; x_1; \sigma),$$

(8) $$F_{\text{Ус}}^{(-1)}(1-\alpha_2; x_2; \sigma) = F^{(-1)}(1-\alpha_2^0; x_2; \sigma),$$

(9) $$\alpha_1^0 = \frac{\alpha_1}{c_1} + F(\underline{x}; x_1; \sigma),$$

(10) $$\alpha_2^0 = \frac{\alpha_2}{c_2} + F(\underline{x}; x_2; \sigma),$$

где c_1, c_2 – нормирующие коэффициенты, устанавливающие связь между усеченным и не усеченным нормальным законом распределения.

Нормирующие коэффициенты определяются по формулам [4]:

(11) $$c_1 = \frac{1}{F(\bar{x}; x_1; \sigma) - F(\underline{x}; x_1; \sigma)},$$

(12) $$c_2 = \frac{1}{F(\bar{x}; x_2; \sigma) - F(\underline{x}; x_2; \sigma)}.$$

Следует отметить, что при выполнении условий $|\bar{x} - x_{\max}| > 3\sigma$ и $|x_{\min} - \underline{x}| > 3\sigma$, можно считать, что $\alpha_1^0 = \alpha_1$, $\alpha_2^0 = \alpha_2$.

Заканчивается решение задачи классификации объектов определением коэффициентов k_1 и k_2. Очевидно, что критическая область должна быть больше с той стороны, относительно x_1 и x_2, с которой вероятность появления анализируемых объектов выше. В этой связи в качестве исходного положения установлено, что значения коэффициентов k_1 и k_2 пропорциональны распределению объектов множества U на δ-й части интервалов $[x_{\min}; x_1[$ и $]x_2; x_{\max}]$, в пределах которой возможно допущение ошибки, связанной с классификацией объектов. В связи с этим положением:

$$k_1 = \frac{\int\limits_{x_1-\delta_1}^{x_1} f(u/x)dx}{\int\limits_{x_1-\delta_1}^{x_1} f(u/x)dx + \int\limits_{x_2}^{x_2+\delta_2} f(u/x)dx}, \tag{13}$$

$$k_2 = \frac{\int\limits_{x_2}^{x_2+\delta_2} f(u/x)dx}{\int\limits_{x_1-\delta_1}^{x_1} f(u/x)dx + \int\limits_{x_2}^{x_2+\delta_2} f(u/x)dx}. \tag{14}$$

При обосновании значений величин δ_1 и δ_2 было принято, что допущение ошибки связанной с классификацией объектов в принципе возможно только для тех объектов, значения параметра x у которых принадлежат диапазонам $[x_1 - 3\sigma; x_1]$ и $[x_2; x_2 + 3\sigma]$[1]. С учетом данного допущения и располагаемого расстояния между границами интервалов $[x_{\min}; x_1[$ и $]x_2; x_{\max}]$, величины δ_1 и δ_2 должны принимать следующие значения (см. таблицу).

условие	$\delta_1 =$	$\delta_2 =$
$x_1 - x_{\min} < 3\sigma$, $x_{\max} - x_2 < 3\sigma$	$x_1 - x_{\min}$	$x_{\max} - x_2$
$x_1 - x_{\min} < 3\sigma$, $x_{\max} - x_2 \geq 3\sigma$		3σ
$x_1 - x_{\min} \geq 3\sigma$, $x_{\max} - x_2 \geq 3\sigma$	3σ	
$x_1 - x_{\min} \geq 3\sigma$, $x_{\max} - x_2 < 3\sigma$		$x_{\max} - x_2$

Ввод указанных коэффициентов позволяет избежать ошибок, которые возникнут, если расположение критических точек будет осуществляться так как это принято для двусторонней критической области, построенной для $x_1 = x_2$. В этом случае принято, что левая и правая критические точки определяются для значения вероятности равной $\alpha/2$ [6]. Такой подход будет безошибочным для случая $x_1 \neq x_2$, если распределение объектов на диапазоне $[x_{\min}; x_{\max}]$ будет равномерным, в остальных слу-

---

[1] - *в некоторых случаях указанный диапазон может быть основан на величине более чем 3σ.*

чаях он может быть причиной ошибки, тем более существенной, чем более разница в распределении объектов на указанном диапазоне.

Первоначально, для расчета коэффициентов k_1 и k_2 предполагалось использовать выражения:

$$k_1 = \frac{f(u/x_1)}{f(u/x_1) + f(u/x_2)},\ k_2 = \frac{f(u/x_2)}{f(u/x_1) + f(u/x_2)}, \tag{15}$$

но их дальнейшее практическое применение показало, что они не обеспечивают ожидаемого эффекта для решения задачи классификации объектов, так как учитывают пропорцию распределения объектов только в окрестностях точек x_1 и x_2, а этого оказалось недостаточно. Поэтому, впоследствии были составлены выражения, которые учитывали характер распределения анализируемых объектов на всей длине интервалов $[x_{min}; x_1[$ и $]x_2; x_{max}]$:

$$k_1 = \frac{\int_{x_{min}}^{x_1} f(u/x)dx}{1 - \int_{x_1}^{x_2} f(u/x)dx},\ k_2 = \frac{\int_{x_2}^{x_{max}} f(u/x)dx}{1 - \int_{x_1}^{x_2} f(u/x)dx}. \tag{16}$$

Эффект от использования выражений (16) вырос, но не достиг ожидаемого уровня. Анализ показал, что учет характера распределения на всей длине интервалов $[x_{min}; x_1[$ и $]x_2; x_{max}]$ является избыточным и приводит при некоторых начальных условиях к результату решения задачи хуже ожидаемого, так как для определения данных коэффициентов необходимо знать характер распределения только тех объектов в указанных интервалах, для которых совершение ошибки является в принципе возможным. Поэтому для расчета коэффициентов k_1 и k_2 на практике предлагается использовать выражения (13, 14), но в некоторых условиях для этого могут быть успешно использованы выражения (15) и (16).

Для решения частной задачи 2 необходимо учесть, что основанием ошибки второго рода является неверная классификация объектов, фактически принадлежащих подмножеству U_1 и распределенных по закону $f(u/x)$, у которых значение параметра x может находиться в любой точке диапазонов $[x_{min}; x_1\text{-}\Delta_1]$ или $[x_2\text{+}\Delta_2; x_{max}]$. На рис.3 изображены плотности распределения результатов измерения для случая, когда истинное значение параметра u-го объекта принадлежащего подмножеству U_1, расположено в окрестностях точек $(x_1 - \Delta_1)$ или $(x_2 + \Delta_2)$, для которых значение ошибки второго рода достигает максимального значения (заштрихованная область на графиках). С приближением истинного значения параметра x к границам его возможных значений – точкам x_{min} и x_{max}, значение вероятности ошибки второго будет становиться меньше.

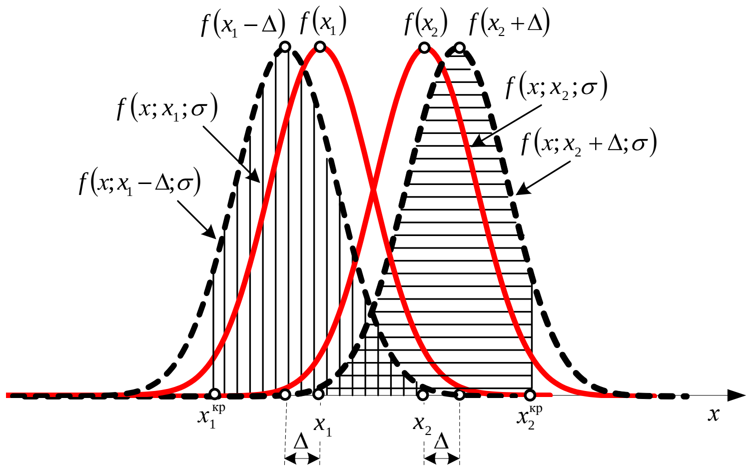

Рис. 3. Расположение области ошибки второго рода на графиках плотности распределения результатов измерения (область выделена штрихованием)

В этой связи возникает необходимость составления выражения, которое позволит определить значение ошибки второго рода безотносительно от того в каком конкретно месте диапазонов $[x_{min}; x_1 - \Delta_1]$ и $[x_2\text{+}\Delta_2; x_{max}]$ находится истинное значение параметра x исследуемого u-го объекта, т.е. позволит определить ожидаемое значение ошибки второго рода $M[\beta]$. Основой такого выражения является формула математического ожидания непрерывной случайной величины [7], которая с учетом принятого комплекса условий представляется в следующем виде:

$$(17) \qquad M[\beta] = \int_{x_{\min}}^{x_{\max}} \beta(x) f(u/x) dx,$$

где $f(u/x)\,dx$ – элемент вероятности того, что истинное значение оцениваемого параметра u-го объекта находится в окрестностях точки x;

$\beta(x)$ – вероятность ошибки второго рода, определенная для условий расположения истинного значения оцениваемого параметра в окрестностях точки x:

$$(18) \qquad \beta(x) = \begin{cases} 1 - F_{\text{Ус}}\left(x_1^{\text{кр}}; x; \sigma\right), \underline{x} \le x < x_1 - \Delta, \\ 0, x_1 - \Delta \le x < x_2 + \Delta, \\ F_{\text{Ус}}\left(x_2^{\text{кр}}; x; \sigma\right), x_2 + \Delta \le x < \bar{x}, \end{cases}$$

где $F_{\text{Ус}}\left(x_1^{\text{кр}}; x; \sigma\right) = P\left(X < x_1^{\text{кр}}\right)$, $F_{\text{Ус}}\left(x_2^{\text{кр}}; x; \sigma\right) = P\left(X < x_2^{\text{кр}}\right)$ – функция распределения усеченного нормального закона распределения.

Необходимость применения функции распределения усеченного нормального закона обусловлена тем, что как было оговорено при постановке задачи, результаты измерения $X=X_u$ ограничены слева и справа значениями $\underline{x}$ и $\bar{x}$.

Для расчета функции распределения усеченного нормального закона воспользуемся выражением (на примере $F_{\text{Ус}}\left(x_1^{\text{кр}}; x; \sigma\right)$) [4]:

$$(19) \qquad F_{\text{Ус}}\left(x_1^{\text{кр}}\right) = c\left(F\left(x_1^{\text{кр}}\right) - F\left(\underline{x}\right)\right),$$

где $F\left(x_1^{\text{кр}}\right) = F\left(x_1^{\text{кр}}; x; \sigma\right)$, $F\left(\underline{x}\right) = F\left(\underline{x}; x; \sigma\right)$ – функция нормального закона распределения;

$c = \dfrac{1}{F(\bar{x}; x; \sigma) - F(\underline{x}; x; \sigma)}$ – нормирующий коэффициент, позволяющий перевести закон распределения нормального закона к усеченному виду.

При составлении выражения (19) учтено, что критические точки всегда будут располагаться в пределах границ интервала $\left[\underline{x};\bar{x}\right]$. Если бы это условие не соблюдалось, то для аналогичных расчетов, необходимо было бы воспользоваться выражением:

$$F_{\text{Ус}}\left(x_1^{\text{кр}}\right)=\begin{cases}0, x_1^{\text{кр}}<\underline{x},\\ c\left(F\left(x_1^{\text{кр}}\right)-F\left(\underline{x}\right)\right), \underline{x}\le x_1^{\text{кр}}<\bar{x},\\ 1, \bar{x}\le x_1^{\text{кр}}.\end{cases} \tag{20}$$

Аналогично 2-й задаче решается и 3-я частная задача. При составлении математического выражения, необходимого для определения ожидаемого значения $M[\alpha]$, также необходимо учесть, что причиной ошибки первого рода является неверная классификация объектов, фактически принадлежащих подмножеству U_0 и распределенных по закону $f(u/x)$, у которых значение параметра x может находиться в любой точке диапазона $[x_1; x_2]$. Как и во второй частной задаче, основой такого выражения является формула для определения математического ожидания непрерывной случайной величины:

$$M[\alpha]=\int_{x_1}^{x_2}\left(\alpha_1(x)+\alpha_2(x)\right)f(u/x)dx, \tag{21}$$

где $\alpha_1(x)$ и $\alpha_2(x)$ – слагаемые вероятности ошибки первого рода, определенные для условий расположения истинного значения оцениваемого параметра u-го объекта в окрестностях точки x:

$$\alpha_1(x)=\begin{cases}F_{\text{Ус}}\left(x_1^{\text{кр}}; x; \sigma\right), x_1\le x\le x_2,\\ 0, x<x_1\vee x_2<x,\end{cases} \tag{22}$$

$$\alpha_2(x)=\begin{cases}1-F_{\text{Ус}}\left(x_2^{\text{кр}}; x; \sigma\right), x_1\le x\le x_2,\\ 0, x<x_1\vee x_2<x;\end{cases} \tag{23}$$

где $F_{\text{Ус}}\left(x_1^{\text{кр}}; x; \sigma\right)=P\left(X_u<x_1^{\text{кр}}\right)$, $F_{\text{Ус}}\left(x_2^{\text{кр}}; x; \sigma\right)=P\left(X_u<x_2^{\text{кр}}\right)$ – функция распределения усеченного нормального закона распределения.

Таким образом, решение поставленных частных задач проведено. Представленные решения дополняют известный порядок проверки статистических гипотез на основе критерия Неймана-Пирсона, для комплекса условий, имеющего более широкое прикладное значение за счет расширения постановочной части задачи следующими условиями:

рассмотрения диапазона принадлежности исследуемых объектов заданному подмножеству $[x_1; x_2]$, относительно которого проверяется основная статистическая гипотеза (в классическом варианте постановки задачи вместо этого диапазона указывается одно единственное значение параметра, характеризующее то или иное свойство объекта);

рассмотрения диапазона возможных значений параметра x $[x_{\min}; x_{\max}]$ и вероятностного распределения объектов множества U на этом диапазоне $f(u/x)$ (в классическом варианте этот элемент постановки задачи отсутствует);

ввода величины области неразличимости Δ у границ диапазона $[x_1; x_2]$, в которой расположены объекты по отношению, к которым безразлично, куда их можно отнести: в подмножество U_0 или U_1 (аналогичный по смыслу параметр рассматривается А. Вальдом в методе последовательного анализа, который используется при решении задачи проверки гипотезы о равенстве математического ожидания случайной величины некоторой постоянной [2,19]);

рассмотрения диапазона возможных результатов измерения параметра x $\left[\underline{x}; \bar{x}\right]$ (в классическом варианте этот элемент постановки задачи также отсутствует).

Кроме расширения области, в которой рассмотренная задача может найти свое прикладное значение, ввод дополнительных параметров в ее условие позволяет повысить надежность принимаемого решения за счет более тонкого расчета положения критических точек, и, кроме этого, позволяет вычислить ожидаемый уровень ошибок первого и второго рода в результате классификации всего исходного множества объектов, что также имеет важный практический смысл.

Важной особенностью представленной постановки является возможность ее сведения к классической постановке, например, как это представлено в [11,12]. В этом случае для проверки гипотезы о равенстве математического ожидания случайной величины, которая может себя проявить на всем диапазоне действительных значений, некоторому, наперед заданному значению, необходимо ввести следующие изменения: $x_1 = x_2 = x_0$, $\Delta_1 = \Delta_2 = 0$; в качестве плотности распределения $f(u/x)$ принять любое распределение симметричное относительно своего математического ожидания (например, равномерное распределение), границами диапазона $[x_{\min}; x_{\max}]$ выбрать точки удовлетворяющие неравенству $x_{\min} + K\sigma < x_0 < x_{\max} - K\sigma$, где $K \geq 3$, принять $\underline{x} = x_{\min}$ и $\bar{x} = x_{\max}$.

Следует также отметить, что ввод параметров Δ_1 и Δ_2, обусловлен необходимостью дальнейшего развития комплекса условий, находящегося в основе поставленных частных задач. Данный параметр в полной мере проявит свое предназначение для условий, в которых рассматривается поток объектов, на анализ каждого из которых отводится ограниченное время, а также имеются потери, которые неизбежно настанут в случае их неверной классификации. Кроме этого эти параметры будут необходимы при решении поставленных задач методом последовательного анализа А.Вальда, который обеспечивает получение результата проверки статистических гипотез с заданной степенью надежности для выборки меньшего объема (меньшего числа измерений параметра x), чем это необходимо для решения тех же задач методом, который требует формирование выборки фиксированного объема [2,19].

Таким образом, представленные решения частных задач позволяют составить методику классификации объектов в условиях стохастической неопределенности. Данная методика состоит из следующих, последовательно выполняемых этапов.

1 этап. Ввод значений исходных данных, представленных в кортеже (3).

2 этап. Определение значений критических значений $x_1^{\text{кр}}$, $x_2^{\text{кр}}$ параметра x, путем подстановки в выражения (6-14) соответствующих значений исходных данных.

3 этап. Определяется ожидаемое значение ошибки первого и второго рода ($M[\alpha]$ и $M[\beta]$) с использованием выражений (17-23). Представление полученных результатов.

4 этап. Измерение параметра x очередного классифицируемого объекта – X_u.

5 этап. Проверка статистической гипотезы о принадлежности неизвестного значения математического ожидания x_u случайной величины X_u диапазону $[x_1; x_2]$ на базе критерия, заданного неравенствами (4) и (5).

6 этап. Представление результата классификации очередного объекта.

7 этап. Проверка условия окончания процедуры классификации множества объектов U. Если условие не выполняется, то осуществляется возврат к 4 этапу, иначе выполнение методики завершается. В самом простом виде условием окончания процедуры может быть классификация всех объектов множества U.

4. *Результаты классификации объектов в условиях стохастической неопределенности с использованием разработанной методики*

С целью подтверждения возможности практического использования разработанной методики, проверки решений частных задач на адекватность, а также формулирования выводов прикладного характера был спланирован и проведен комплексный эксперимент.

В ходе проведения эксперимента проведено решение поставленных частных задач для различных комплексов условий (КУ). Выражение плотности распределения $f(u/x)$ для всех КУ составлено на основе бета-распределения, четыре параметра которого позволяют смоделировать практически любой вид распределения [7]. Параметры этого распределения λ и η определяют его форму, а параметры $[x_{\min}; x_{\max}]$ определяют границы его возможных значений. В эксперименте рассмотрены два типа распределения объектов, смоделированных с помощью бета-распределения: равномерное и существенно сдвинутое вправо. Комплексы условий отличались друг от друга типами диапазонов $[x_1; x_2]$ («узкий диапазон», «диапазон средней длины» и «широкий диапазон»), а также тремя уровнями точности проводимых измерений. Тип диапазона определялся величиной отношения длин диапазонов $[x_1; x_2]$ и $[x_{\min}; x_{\max}]$. Так же при проведении эксперимента принято, что для всех комплексов условий при проверке гипотезы H_0 применительно к любому объекту U-го множества ошибка первого рода α не должна превысить 0,1, значения обоих параметров Δ_1 и Δ_2 равно 0, а диапазон возможных результатов измерения параметра x задан границами $\left[\underline{x} = 0; \bar{x} = \infty\right]$. Сочетания представленных условий проведения эксперимента позволили сформировать 18 комплексов условий (см. табл.1).

В ходе проведения эксперимента для каждого комплекса условий по формулам (17-23) двумя способами определены ожидаемые значения ошибок первого и второго рода ($M[\alpha]$, $M[\beta]$), достигнутые при проверке всех объектов множества U. Способы вычисления значений этих ошибок отличались расположением критических точек. В первом случае критические точки вычислялись классическим способом, т.е. принималось, что $\alpha_1 = \alpha/2$ и $\alpha_2 = \alpha/2$ – так как это принято для двухсторонней критической области [6], во втором случае реализован уточненный способ расчета критических точек, для которых $\alpha_1 = k_1\alpha$ и $\alpha_2 = k_2\alpha$, где k_1 и k_2 определяются по формулам (13, 14). Кроме двух аналитических способов расчета $M[\alpha]$ и $M[\beta]$, расчет указанных величин был также проведен на основе метода статистического моделирования. Для этого была разработана статистическая математическая модель, реализующая введенный критерий проверки гипотез (4, 5) по отношению к множеству объектов U заданного объема N. Доля верных результатов проверки всего множества объектов позволяет определить среднюю частоту ошибок первого и второго рода:

(24, 25) $M^*[\alpha] = \dfrac{N(U_0) - n(U_0)}{N}$, $M^*[\beta] = \dfrac{N(U_1) - n(U_1)}{N}$,

где $N(U_0)$, $N(U_1)$ – число проанализированных объектов, принадлежащих подмножествам U_0 и U_1;

$n(U_0)$, $n(U_1)$ – число проанализированных объектов подмножеств U_0 и U_1, по отношению к которым соответственно была принята гипотеза H_0 или H_1;

N – суммарное число проанализированных объектов, принадлежащих исходному множеству U.

Таблица 1. Комплексы условий, сформированные для проведения комплексного эксперимента

№ п/п	№ КУ	Параметры Бета-распределения и его краткая характеристика		Тип диапазона, $[x_1; x_2]$	СКО измерения, σ
		$\lambda; \eta$	$[x_{min}; x_{max}]$		
1.	1.1.1	1,01; 1,01	0; 20 Практически равномерное распределение объектов на диапазоне возможных значений	«узкий диапазон», 9,8; 10,2	0,6
2.	1.1.2				1,2
3.	1.1.3				1,8
4.	1.2.1			«диапазон средней длины», 9; 11	0,6
5.	1.2.2				1,2
6.	1.2.3				1,8
7.	1.3.1			«широкий диапазон», 7,5; 12,5	0,6
8.	1.3.2				1,2
9.	1.3.3				1,8
10.	2.1.1	6; 2	0; 20 Существенно сдвинутое вправо распределение объектов на диапазоне возможных значений	«узкий диапазон», 9,8; 10,2	0,6
11.	2.1.2				1,2
12.	2.1.3				1,8
13.	2.2.1			«диапазон средней длины», 9; 11	0,6
14.	2.2.2				1,2
15.	2.2.3				1,8
16.	2.3.1			«широкий диапазон», 7,5; 12,5	0,6
17.	2.3.2				1,2
18.	2.3.3				1,8

На рис.4 представлена общая схема алгоритма, использованного для разработки статистической математической модели реализующей критерий (4, 5). Рассмотрим содержание отдельных блоков этого алгоритма.

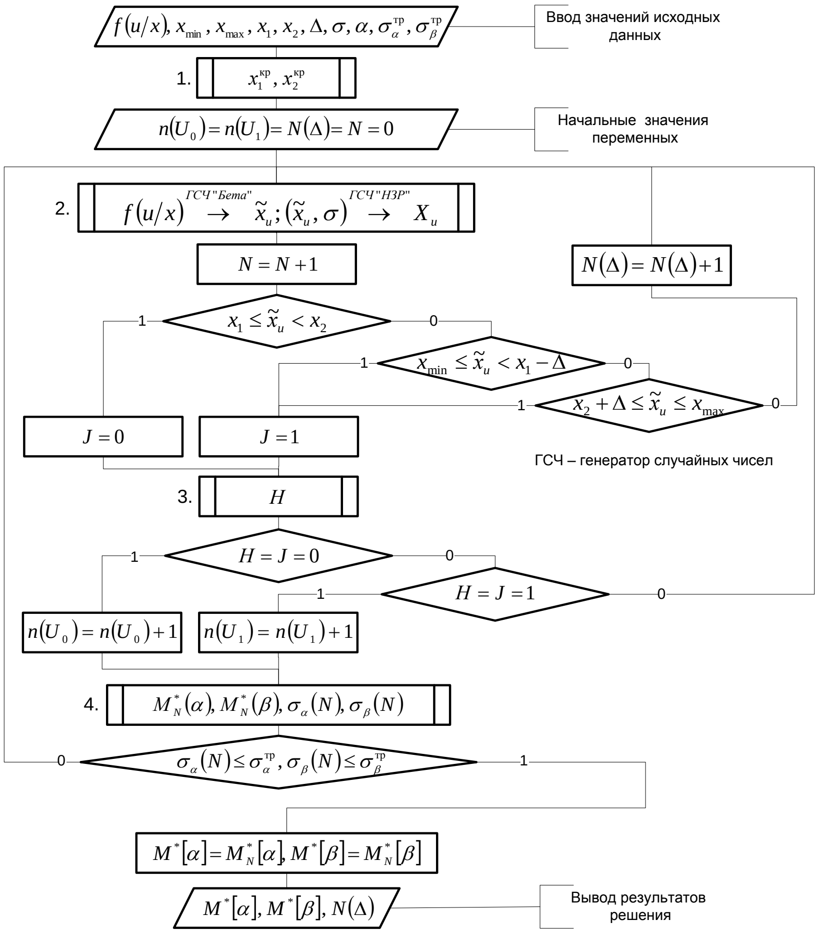

Рис. 4. Схема алгоритма статистической математической модели, реализующей критерий проверки основной гипотезы

В блоке 1 алгоритма проводится расчет значений критических точек. В блоке 2 с использованием генератора случайных чисел бета-распределения для *u*-го случайного объекта моделируется истинное значение его параметра x – $\tilde{x}_u$, а далее моделируется случайный нормально распределенный результат его измерения X_u. Генераторы случайных чисел, которые использованы в данном блоке построены на основе алгоритмов, представленных в [5]. В блоке 3 определяется результат проверки основной гипотезы на основе критерия (4, 5). Если основная гипотеза принята, то индикатор H=0, иначе H=1. В блоке 4, по формулам (24, 25), определяются текущие значения статистических оценок $M_N^*(\alpha)$ и $M_N^*(\beta)$. Также в этом блоке по формулам математической статистики определяются достигнутые характеристики точности статистических оценок, в качестве которых приняты их СКО – $\sigma_\alpha(N)$ и $\sigma_\beta(N)$.

Практически данный алгоритм был программно реализован с помощью табличного редактора «MS Excel» и встроенных в него средств программирования VBA.

Перед проведением КУ предполагалось, что при равномерном распределении объектов на диапазоне возможных значений результаты аналитического и статистического решения частных задач должны совпасть для обоих способов расчета критических точек. При неравномерном распределении объектов на диапазоне возможных значений результаты аналитического и статистического решения частных задач должны совпасть, но при этом результаты решения частных задач для обоих способов расчета критических точек будут разные.

В таблице 2 представлены результаты проведенного комплексного эксперимента. При проведении расчетов с использованием статистической модели объем выборки N был таким, что СКО статистических оценок было не больше 1/50 от значений оценок $M^*[\alpha]$ и $M^*[\beta]$, т.е.: $\sigma_\alpha^{\text{тр}} = M^*[\alpha]/50$ и $\sigma_\beta^{\text{тр}} = M^*[\beta]/50$.

Таблица 2. Результаты, полученные в ходе проведения комплексного эксперимента (значения округлены до 5-го знака после запятой)

№ п/п	№ КУ	1 способ расчета (α1=α/2, α2=α/2)			
		$M[\alpha]$	$M[\beta]$	$M^*[\alpha]$	$M^*[\beta]$
1.	1.1.1	0,00105	0,10055	0,00105	0,10268
2.	1.1.2	0,00144	0,20104	0,00143	0,19302
3.	1.1.3	0,00160	0,30158	0,00160	0,29567
4.	1.2.1	0,00126	0,10053	0,00126	0,10217
5.	1.2.2	0,00251	0,20098	0,00251	0,20100
6.	1.2.3	0,00362	0,30168	0,00365	0,30158
7.	1.3.1	0,00126	0,10045	0,00126	0,09934
8.	1.3.2	0,00252	0,20077	0,00253	0,19785
9.	1.3.3	0,00378	0,30259	0,00378	0,30478
10.	2.1.1	0,00068	0,06872	0,00068	0,06767
11.	2.1.2	0,00094	0,15087	0,00094	0,15094
12.	2.1.3	0,00105	0,25219	0,00105	0,24879
13.	2.2.1	0,00085	0,07501	0,00085	0,07522
14.	2.2.2	0,00167	0,16813	0,00166	0,16449
15.	2.2.3	0,00241	0,28051	0,00241	0,27324
16.	2.3.1	0,00103	0,09528	0,00104	0,09601
17.	2.3.2	0,00199	0,21193	0,00198	0,20224
18.	2.3.3	0,00290	0,33867	0,00285	0,33128

№ п/п	№ КУ	2 способ расчета (α_1=$k_1\alpha$, α_2=$k_2\alpha$)			
		$M[\alpha]$	$M[\beta]$	$M^*[\alpha]$	$M^*[\beta]$
	1.1.1	0,00105	0,10055	0,00107	0,10168
2.	1.1.2	0,00144	0,20104	0,00149	0,20257
3.	1.1.3	0,00160	0,30158	0,00164	0,31208
4.	1.2.1	0,00126	0,10053	0,00127	0,09649
5.	1.2.2	0,00251	0,20098	0,00247	0,20428
6.	1.2.3	0,00362	0,30168	0,00366	0,30824
7.	1.3.1	0,00126	0,10045	0,00128	0,10089
8.	1.3.2	0,00252	0,20077	0,00255	0,19962
9.	1.3.3	0,00378	0,30259	0,00379	0,30171
10.	2.1.1	0,00070	0,06634	0,00070	0,06543
11.	2.1.2	0,00096	0,13676	0,00096	0,13825
12.	2.1.3	0,00107	0,21557	0,00107	0,21720
13.	2.2.1	0,00107	0,06891	0,00107	0,06861

№ п/п	№ КУ	2 способ расчета ($\alpha_1=k_1\alpha$, $\alpha_2=k_2\alpha$)			
		$M[\alpha]$	$M[\beta]$	$M^*[\alpha]$	$M^*[\beta]$
14.	2.2.2	0,00211	0,14461	0,00209	0,14455
15.	2.2.3	0,00294	0,23070	0,00295	0,22543
16.	2.3.1	0,00187	0,08013	0,00183	0,08024
17.	2.3.2	0,00360	0,17126	0,00363	0,16975
18.	2.3.3	0,00511	0,27066	0,00516	0,26791

Анализ полученных результатов позволяет сделать следующие выводы:

ожидаемые результаты комплексного эксперимента после его проведения подтвердились;

при равномерном распределении анализируемых объектов ожидаемые значения ошибок первого и второго рода ($M[\alpha]$ и $M[\beta]$), вычисленные на основе классического и уточненного способа расчета критических точек, одинаковые;

при неравномерном распределении анализируемых объектов ожидаемое значение ошибки первого рода ($M[\alpha]$) будет ниже при использовании классического способа расчета критических точек;

при неравномерном распределении анализируемых объектов ожидаемое значение ошибки второго рода ($M[\beta]$) будет ниже при использовании уточненного способа расчета критических точек;

при неравномерном распределении анализируемых объектов ожидаемое значение суммарной ошибки ($M[\alpha]+M[\beta]$) будет ниже при использовании уточненного способа расчета критических точек;

сравнение результатов расчета полученных аналитическим и статистическим способом показывает, что между ними имеется очевидная сходимость.

Таким образом, анализ полученных данных показал, что составленная методика классификации объектов в условиях стохастической неопределенности на основе метода проверки статистической гипотезы о принадлежности математического ожидания случайной величины заданному диапазону позволяет получать адекватные результаты и может быть реализована на практике на основе предложенного алгоритма статистической математической модели. Кроме того, вывод о том, что ожидаемое значение суммарной ошибки ($M[\alpha] + M[\beta]$) будет меньше при использовании уточненного способа расчета критических точек является предпосылкой к тому, что проверка статистической гипотезы о принадлежности математического ожидания случайной величины заданному диапазону при использовании уточненного способа будет более целесообразным по сравнению с классическим способом расчета критических точек.

В качестве показателя, характеризующего целесообразность применения того или иного способа, примем отношение ожидаемых потерь от принятия ошибочных решений основанных на различных способах расчета критических точек

(26) $\quad \nu = c_\kappa / c_y$,

где c_κ и c_y – суммарные потери от принятия неверных решений при использовании классического и уточненного способа расчета критических точек в которых исходное значение ошибки первого рода (α) принято таким, чтобы эти потери были минимальными из возможных.

Если в результате расчетов показатель $\nu > 1$, то значит наблюдается преимущество уточненного способа расчета критических точек. Если в результате расчетов показатель $\nu < 1$, то значит наблюдается преимущество классического способа расчета критических точек. Иначе, при $\nu = 1$, оба способы по отношению друг к другу по выбранному показателю являются равноценными.

С учетом принятого условия о минимизации возможных потерь, расчет величин c_κ и c_y осуществляется с помощью выражения

(27) $\quad c = c_\alpha M[\alpha] + c_\beta M[\beta],$

где c_α и c_β – потери от принятия неверных решений, связанных с допущением ошибок первого и второго рода.

На рис. 5 представлены графики зависимости показателя (26) от аргумента c_α / c_β , построенные на основе уточненного и классического способа расчета критических точек. Построение графиков проведено для комплексов условий №№10-18 (см.

таблицу 1), при этом принято, что возможные потери от совершения ошибки второго рода не превышают возможные потери от совершения ошибки первого рода: $c_\alpha > c_\beta$. В связи с тем, что подынтегральное выражение плотности бета-распределения не имеет первообразной, расчеты, необходимые для построения графиков, были проведены численным методом.

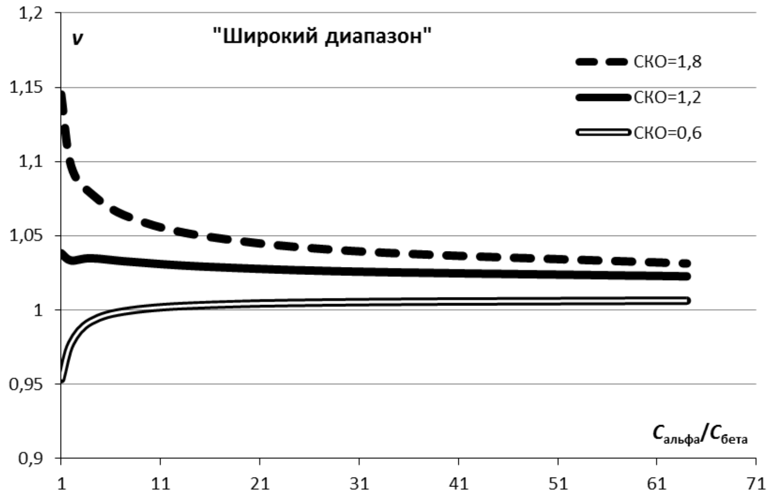

Рис. 5. Графики отношений ожидаемых потерь от принятия ошибочных решений, основанных на различных способах расчета критических точек

Анализ построенных графиков позволяет сделать следующие выводы, имеющие важное прикладное значение.

1) Чем уже диапазон, определяющий принадлежность u-го объекта U_0-у подмножеству, и чем ниже точность проводимых измерений параметра x, тем более целесообразно применение уточненного способа расчета критических точек. Для принятого комплекса условий преимущество уточненного способа расчета критических точек над классическим для «узкого диапазона» с СКО равным 1,8 составило от 10 до 15%.

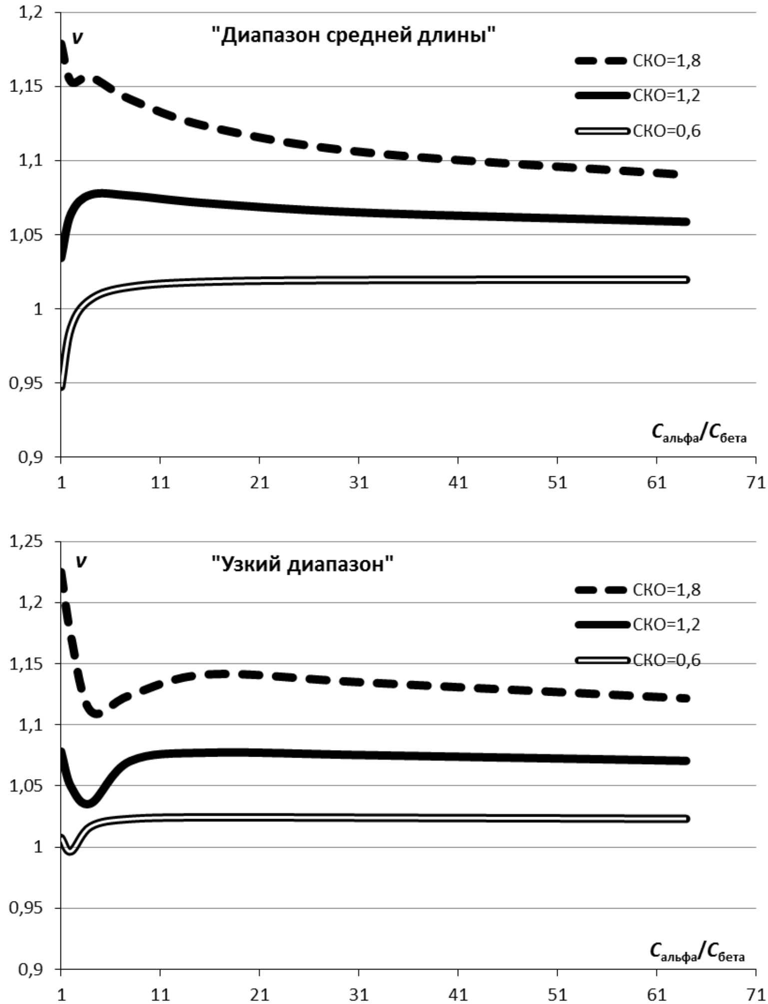

Рис. 5 (продолжение). Графики отношений ожидаемых потерь от принятия ошибочных решений, основанных на различных способах расчета критических точек

2) Чем выше точность проводимых измерений и меньше отношение ожидаемых потерь от принятия ошибочных решений, тем более целесообразно применение классического способа расчета критических точек. Для принятого комплекса условий преимущество классического способа расчета критических точек над уточненным способом для «широкого диапазона» и «диапазона средней длины» с СКО равным 0,6 составило 5%.

3) Область факторного пространства, в которой применение уточненного способа расчета критических точек целесообразнее классического способа, является доминирующей. Для принятого комплекса условий 5%-е преимущество классического способа наблюдается только в той области, где потери от совершения ошибок первого и второго рода приблизительно сопоставимы. В области, где потери от совершения ошибки первого рода многократно выше, чем потери от совершения ошибок второго рода в среднем наблюдается следующее преимущество:

в 0,5-2,5%, при измерениях, проведенных с высокой точностью;

в 3-7%, при измерениях, проведенных со средней точностью;

в 5-14%, при измерениях, проведенных с низкой точностью.

4) Можно предположить, что при $c_\alpha/c_\beta \to \infty$, будет наблюдаться, что $v \to 1$. Дополнительные расчеты показали, что при $c_\alpha / c_\beta = 1000$ и $c_\alpha / c_\beta = 10\,000$, данное предположение подтверждается, но, тем не менее, даже при таких высоких отношениях продолжает наблюдаться преимущество уточненного способа расчета критических точек (см. таблицу 3).

Таблица 3. Отношение ожидаемых потерь от принятия ошибочных решений, основанных на различных способах расчета критических точек (результаты округлены до 4-го знака после запятой)

№ п/п	Тип диапазона	СКО	$c_\alpha / c_\beta = 1000$	$c_\alpha / c_\beta = 10\,000$
1.	Широкий диапазон	0,6	1,0068	1,0062
2.		1,2	1,0143	1,0101
3.		1,8	1,0133	1,0060

№ п/п	Тип диапазона	СКО	$c_\alpha / c_\beta = 1000$	$c_\alpha / c_\beta = 10\,000$
4.	Диапазон средней длины	0,6	1,0175	1,0154
5.		1,2	1,0406	1,0310
6.		1,8	1,0503	1,0307
7.	Узкий диапазон	0,6	1,0192	1,0166
8.		1,2	1,0523	1,0416
9.		1,8	1,0762	1,0516

Важно отметить, что представленные результаты расчетов могут существенно поменяться при изменении комплекса условий, при этом характер выводов вероятнее всего существенно не изменится, что, впрочем, требует своего подтверждения.

5. *Заключение*

Таким образом, сущность и содержание методики классификации объектов в условиях стохастической неопределенности на основе метода проверки статистической гипотезы о принадлежности математического ожидания случайной величины заданному диапазону рассмотрена в полном объеме:

проведена постановка задачи для комплекса условий, в котором учтено большее число факторов, оказывающих существенное влияние на результат классификации объектов в условиях стохастической неопределенности, за счет чего было усилено ее прикладное значение;

разработана методика решения поставленной задачи на основе проверки сложной статистической гипотезы методом Неймана-Пирсона, в котором используется новый способ определения критических точек и составлены выражения для расчета ожидаемых значений ошибок первого и второго рода, которые будут получены при анализе множества объектов, характеризуемых вероятностно распределенным параметром;

показана возможность сведения поставленной задачи к классической проверке гипотезы о равенстве математического ожидания случайной величины заданной величине;

по результатам проведения комплексного эксперимента подтверждена адекватность и возможность практического использования разработанной методики;

разработана статистическая математическая модель, реализующая разработанное решение на практике, которая может быть использована в качестве прототипа отдельных компонент «передовых технических разработок»;

проведен сравнительный анализ двух способов расчета критических точек, который показал существенное преимущество уточненного способа расчета над классическим способом.

Вместе с тем, ряд вопросов требуют дальнейшего исследования. В первую очередь к их числу следует отнести:

детальное исследование влияния параметров комплекса условий на качество результатов решения задачи классификации объектов на основе применения разработанной методики;

обоснование значений параметров, характеризующих область неразличимости Δ_1 и Δ_2, для условий, в которых рассматривается поток объектов на анализ каждого из которых отводится ограниченное время, а также имеются потери, которые неизбежно возникнут в случае их неверной классификации;

составление аналитического выражения, предназначенного для определения значения ошибки первого рода α, при которой обеспечивается минимум ожидаемых суммарных потерь от принятия неверных решений (c);

составление аналитических выражений для расчета отношений ожидаемых потерь от принятия ошибочных решений, основанных на различных способах расчета критических точек и др.

Литература

1. БАЛДИН К.В., ВОРОБЬЕВ С.Н., УТКИН В.Б. *Управленческие решения*: Учебник / 8-е изд. – М.: Дашков и К, 2018. - 496 с.
2. ВАЛЬД. А. *Последовательный анализ.* – М.: Государственное издательство физико-математической литературы, 1960. - 328 с.

3. ВИНОГРАДОВ И.М. *Математическая энциклопедия. В 5 т.т./ 1 т. А-Г.*– М.: Советская энциклопедия. 1977г. - 576с.
4. ВИНОГРАДОВ И.М. *Математическая энциклопедия. В 5 т.т./ 5 т. Слу-Я.* – М.: Советская энциклопедия. 1985г. - 624с.
5. КЕЛЬТОН В., ЛОУ А. *Имитационное моделирование. Классика CS.* 3-е изд. – СПб.: Питер; Киев: Издательская группа DHV, 2004. - 847 с.
6. КОБЗАРЬ А.И. *Прикладная математическая статистика. Справочник для инженеров и научных работников.* – М.: Физматлит, 2006. - 816 с.
7. КОРОЛЮК В.С., ПОРТЕНКО Н.И.,СКОРОХОД А.В. ТУРБИН А.Ф. *Справочник по теории вероятностей и математической статистике.* – М.: Наука, 1985. - 640 с.
8. ЛЕМАН Э. *Проверка статистических гипотез.* – М.: Наука. 1964. - 500 с.
9. ПЕРЕСАДА В.П. *Автоматическое распознавание образов.* – Л .: Энергия, 1970. – 92 с.
10. РУЧКИН С.В. *Моделирование действия фактора стохастического характера на основе использования усеченного закона распределения случайной величины. Стр. 94-96* // Труды XXXVI Всероссийской НТК. Проблемы безопасности функционирования сложных технических и информационных систем. Часть 7. Серпухов.: 2016. - 96 с.
11. СЕВАСТЬЯНОВ Б.А. *Введение в статистическую теорию распознавания образов.* – М.: Наука, Главная редакция физико-математической литературы. 1979. - 368 с.
12. СЕВАСТЬЯНОВ Б.А. *Курс теории вероятностей и математической статистики.* – М.: Наука. Главная редакция физико-математической литературы. 1982. - 256 с.
13. СТЕПИН В.С., ГУСЕЙНОВ А.А., СЕМИГИН Г.Ю., ОГУРЦОВ А.П. и др. *Новая философская энциклопедия. В 4-х тт./ 2 т. Е-М* – М.: Мысль. 2010. - 640 с.
14. Указ Президента Российской Федерации «О национальных целях и стратегических задачах развития РФ на период до 2024 года» от 7.05.2018г. №204. Сайт Президента России –

URL: http://www.kremlin.ru/events/president/news/57425 (дата обращения: 02.01.2020).
15. ФУКУНАГА К. *Введение в статистическую теорию распознавания образов.* – М.: Наука. Главная редакция физико-математической литературы. 1979. - 368 с.
16. https://ru.wikipedia.org/wiki/Неопределенность (дата обращения: 03.01.2020)
17. KEINOSUKE FUKUNAGA. *Introduction to Statistical Pattern Recognition. 2nd Edition. Academic Press, 1990, 592 p.*
18. LEXMANN E.L., ROMANO J.P. *Testing statistical hypotheses.* – New York: Springer. 2005. - 786 p.
19. WALD, ABRAHAM. *Sequential Analysis.* – Dover Publications. 1974. - 224 p.

METHOD OF OBJECT CLASSIFICATION IN CONDITIONS OF STOCHASTIC UNCERTAINTY

Sergey Ruchkin, The Military Academy of Strategic Rocket Troops after Peter the Great, Moscow region, Balashikha, Cand.Sc., senior lecturer (pochtampt07@mail.ru).

Abstract: This publication presents the formulation and method of solving the problem of classification of probabilistically distributed objects under stochastic uncertainty about the true values of their characteristics. The method is based on testing the statistical hypothesis that the mathematical expectation of a random variable belongs to a given range by the Neyman-Pearson method. The peculiarity of the method is a refined procedure for calculating critical points, on the basis of which a more effective decisive rule of object classification is formed, and the ability to estimate the expected values of the probability of errors that may be made as a result of the recognition procedure in relation to all the presented objects.

Keywords: classification, type I error, type II error, recognition, statistical hypothesis

УДК 004.9, 519.2
ББК 22.172

Статья представлена к публикации членом редакционной коллегии ...заполняется редактором...

Поступила в редакцию ...заполняется редактором...
Опубликована ...заполняется редактором...