

© 2019 г. К.В. СЕМЕНИХИН, д-р физ.-мат. наук (siemenkv@rambler.ru)
(Московский авиационный институт)

ДВУСТОРОННЯЯ ВЕРОЯТНОСТНАЯ ГРАНИЦА ДЛЯ СИММЕТРИЧНОЙ УНИМОДАЛЬНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ¹

Для симметричной унимодальной случайной величины с известными модой и дисперсией построена неупрощаемая верхняя граница вероятности попадания вне произвольно заданного интервала. Описано распределение, на котором достигается полученная граница. В частном случае, когда расстояние между модой и серединой интервала меньше 29% его радиуса, найденная граница определяется неравенством Гаусса. Полученный результат применен к построению робастных доверительных интервалов.

Ключевые слова: унимодальное (одновершинное) распределение, неравенство Гаусса, вероятностная граница, наилучшее распределение, робастный доверительный интервал.

DOI: 10.1134/S0005231019030073

1. Введение

Понятие унимодальности (или одновершинности) распределения отражает естественное требование монотонного убывания плотности случайной величины от ее наиболее вероятного значения [1]. При соблюдении этого условия вероятность отклонения случайной величины от своей моды можно оценить сверху с помощью неравенства Гаусса [2]. Если мода совпадает с математическим ожиданием, то это неравенство описывает верхнюю границу, в $9/4$ раз меньшую того, что дает неравенство Чебышёва. Если же мода неизвестна, то вероятность отклонения унимодальной величины от своего среднего оценивается с помощью неравенства Высочанского — Петунина [3]. По форме оно совпадает с неравенством Гаусса при условии, что порог отклонения составляет не меньше $\sqrt{8/3}$ средних квадратичных отклонений. В результате правило “трех сигм” при наилучшем выборе унимодального распределения выполняется с надежностью $77/81 \approx 0,9506$. В этом случае доверительный интервал, основанный на знании первых двух моментов и предположении об унимодальности, оказывается лишь в полтора раза шире интервала, построенного согласно гипотезе о нормальном распределении [4].

Благодаря этому факту стохастические модели с неопределенным унимодальным распределением находят свое применение в различных прикладных задачах, таких как: анализ импульсной активности нейронов [5]; определение границ ошибок при обработке геологических данных [6]; оценка надежности при неопределенной интенсивности отказов [7]; определение технологического допуска с заданной вероятностью [8]. Кроме того, описание неопределенности с помощью различных классов унимодальных распределений встречается и в теоретических работах. В [9–11] доказано, что для случайного

¹ Работа выполнена при поддержке Российского научного фонда (проект № 16-11-00063).

вектора с ограниченным носителем равномерное распределение доставляет минимум вероятности попадания в выпуклое множество. В [12] рассмотрены задачи оптимизации при наличии вероятностных ограничений, которые должны соблюдаться при произвольном выборе распределения альфа-унимодального вектора с фиксированными математическим ожиданием и ковариационной матрицей. В [13] для проблем моментов Маркова, рассматриваемых на классе многомерных унимодальных распределений, разработан метод их преобразования к задачам полуопределенного программирования.

Помимо [3] нахождением неравенств для вероятности выхода унимодальной величины из заданного интервала занимались многие авторы: в [14] задача максимизации указанной вероятности решена для фиксированных значений моды, среднего и дисперсии; в [15] решение этой задачи определено для односторонней вероятности (т.е. вероятности превышения случайной величины заданного порога); в [16] получено обобщение неравенства Высочанского — Петунина с использованием (вместо дисперсии) центрального момента любого положительного порядка; в [17] максимизация односторонней вероятности для симметричной величины с известными моментами сведена к эквивалентной задаче полуопределенного программирования; в [18] определен минимум вероятности попадания положительной величины в интервал с центром в моде, совпадающим с математическим ожиданием.

Таким образом, за рамками изучения осталась задача нахождения верхней границы двусторонней вероятности при условии, что унимодальная случайная величина является симметричной относительно известной моды, не совпадающей с центром интервала. Решение данной задачи планируется применить к построению менее консервативных минимаксных оценок ограниченных параметров в линейных регрессионных моделях при использовании вероятностных критериев качества [19, 20]. Кроме того, результаты статьи могут найти свое применение при разработке робастных методов принятия решений в задачах проверки статистических гипотез, планирования эксперимента и доверительного оценивания.

Точная формулировка рассматриваемой задачи дана в разделе 2, вид искомой вероятностной границы описан в разделе 3, численный анализ характеристик доверительных интервалов, построенных на основе полученного решения, представлен в разделе 4.

2. Постановка задачи

Условие унимодальности случайной величины ξ , имеющей нулевую моду, означает, что функция распределения $F(x) = P\{\xi \leq x\}$ является выпуклой на $(-\infty, 0)$ и вогнутой на $(0, \infty)$. Это равносильно тому, что вероятность попадания в любое борелевское множество B можно записать в виде

$$(1) \quad P\{\xi \in B\} = (1 - q)\delta_0(B) + q \int_B f(x) dx,$$

где q — число из отрезка $[0, 1]$, δ_0 — мера Дирака, сосредоточенная в нуле, $f(x)$ — плотность вероятности, монотонно невозрастающая справа от нуля и монотонно неубывающая слева от нуля.

Предположение о том, что ξ имеет *унимодальное распределение, симметричное относительно нуля*, означает, что имеет место представление (1), в котором функция $f(x)$ может быть выбрана четной. Если дополнительно известна дисперсия $D\xi = d$, то указанное предположение о величине ξ будем записывать в виде $\xi \sim \mathcal{U}(d)$, а также $\xi \sim \mathcal{U}$ при $d = 1$.

Важным представителем класса $\mathcal{U}(d)$ является смесь двух распределений: вырожденного δ_0 и равномерного $\mathcal{R}(-u, u)$. Смесь двух распределений P_0 и P_1 будем обозначать $(1 - q)P_0 + qP_1$.

Цель работы — определить точную верхнюю грань вероятности того, что симметричная унимодальная величина с фиксированными модой и дисперсией находится вне заданного интервала. Такую границу принято называть *двусторонней*, поскольку в ее определении участвуют вероятности попадания величины справа и слева от интервала. Данное событие обычно воспринимается как нежелательное, поэтому распределение, на котором достигается искомая граница, называется *наихудшим*.

Если считать, что мода равна нулю, а дисперсия равна единице, то эту задачу можно сформулировать следующим образом.

Задача. Для любых $h > 0$ и $m \in \mathbb{R}$ определить вероятностную границу

$$(2) \quad \mathbf{u}_h(m) = \sup_{\xi \sim \mathcal{U}} \mathbb{P}\{|\xi - m| \geq h\}$$

и описать распределение, на котором достигается данный супремум.

Отметим, что граница (2) представляет собой функцию двух переменных: середины m и радиуса h промежутка $(m - h, m + h)$. Указанные переменные удобно называть *сдвигом* и *порогом*, поскольку (2) — это наилучшее значение вероятности превышения случайной величиной порога h с учетом сдвига m . Если сдвиг нулевой, т.е. $m = 0$, то искомая граница совпадает с правой частью неравенства Гаусса [2]:

$$(3) \quad \mathbb{P}\{|\xi| \geq h\} \leq \mathbf{g}_h = \begin{cases} 1 - \frac{h}{\sqrt{3}}, & h \leq 2/\sqrt{3}, \\ \frac{4}{9h^2}, & h \geq 2/\sqrt{3}. \end{cases}$$

Если же $m \neq 0$, то промежуток $(m - h, m + h)$ расположен несимметрично относительно моды, и в этом случае задача (2) не имеет известного решения.

Отметим, что помимо (3) известно неравенство Высочанского — Петунина [3, 4]:

$$(4) \quad \mathbb{P}\{|\xi - m| \geq h\} \leq \mathbf{v}_h(m) = \begin{cases} 1, & h \leq \sqrt{m^2 + 1}, \\ \frac{4(m^2 + 1)}{3h^2} - \frac{1}{3}, & \sqrt{m^2 + 1} \leq h \leq \sqrt{8(m^2 + 1)/3}, \\ \frac{4(m^2 + 1)}{9h^2}, & h \geq \sqrt{8(m^2 + 1)/3}, \end{cases}$$

которое справедливо для любой унимодальной величины ξ без требования симметричности распределения и без ограничения на моду, но при тех же условиях на моментные характеристики $E\xi = 0$ и $D\xi = 1$.

3. Решение задачи

Сначала опишем элементарные свойства вероятности, которая подлежит максимизации в задаче (2).

Лемма. Если $\xi \sim \mathcal{U}(d)$ и $h > 0$, то вероятность

$$\pi_h(m) = P\{|\xi - m| \geq h\},$$

как функция параметра $m \in \mathbb{R}$: четна; выпукла на отрезке $[-h, h]$; монотонно не убывает на $[0, \infty)$ и не возрастает на $(-\infty, 0]$; принимает наименьшее значение в нуле; непрерывна всюду кроме, быть может, точек $\pm h$, но $\pi_h(m) \rightarrow \pi_h(h)$ при $m \downarrow h$ или $m \uparrow -h$.

Из леммы следует, что искомая функция $\mathbf{u}_h(m)$ тоже обладает свойством симметрии: $\mathbf{u}_h(m) = \mathbf{u}_h(-m)$. Поэтому не ограничивая общности, можно считать, что $m \geq 0$, т.е. середина интервала смещена вправо от моды.

Решение поставленной задачи приведено в следующей теореме.

Теорема. Для $h > 0$ и $m \geq 0$ граница (2) имеет вид

$$\mathbf{u}_h(m) = \begin{cases} 1 - \frac{h}{\sqrt{3}}, & m \leq \left(1 - \frac{1}{\sqrt{2}}\right)h, \quad h \leq \frac{2}{\sqrt{3}}, \quad (\text{i}) \\ \frac{4}{9h^2}, & \text{---} \text{---} \text{---} \text{---} \quad h \geq \frac{2}{\sqrt{3}}, \quad (\text{ii}) \\ 1 - \frac{h}{\sqrt{3}}, & \left(1 - \frac{1}{\sqrt{2}}\right)h \leq m < h, \quad q_1 \geq 1, \quad (\text{iii}) \\ \frac{1}{2} \left\{ \frac{a-2}{a-1} + \frac{1}{9h^2} \left(\frac{a}{a-1}\right)^3 \right\}, & \text{---} \text{---} \text{---} \text{---} \quad q_1 \leq 1 \leq q_2, \quad (\text{iv}) \\ \frac{1}{2} \left\{ 1 - \frac{h-m}{\sqrt{3}} \right\}, & \text{---} \text{---} \text{---} \text{---} \quad q_2 \leq 1 \leq q_3, \quad (\text{v}) \\ \frac{2}{9(h-m)^2}, & \text{---} \text{---} \text{---} \text{---} \quad q_3 \leq 1, \quad (\text{vi}) \\ 1, & m \geq h, \quad (\text{vii}) \end{cases}$$

где обозначено

$$a = \left(\frac{2h}{h-m} \right)^{2/3},$$

$$q_1 = \frac{1}{3h^2} \left(\frac{a}{a-1} \right)^2,$$

$$q_2 = \frac{4}{3(h-m)^2(a-1)^2},$$

$$q_3 = \frac{4}{3(h-m)^2},$$

причем указанные параметры удовлетворяют условиям

$$\begin{cases} a > 2, & \frac{4}{3h^2} < q_1 < q_2 < q_3, & \text{если } \left(1 - \frac{1}{\sqrt{2}}\right)h < m < h, \\ a = 2, & q_1 = \frac{4}{3h^2} < q_2 = q_3 = \frac{8}{3h^2}, & \text{если } m = \left(1 - \frac{1}{\sqrt{2}}\right)h. \end{cases}$$

В зависимости от выбора случая из (i)–(vi) граница (2) будет достигаться на одном из следующих распределений:

(i), (iii), (v) $\mathcal{R}(-\sqrt{3}, \sqrt{3})$;

(ii) $(1-q)\delta_0 + q\mathcal{R}\left(-\frac{3h}{2}, \frac{3h}{2}\right)$ при $q = \frac{4}{3h^2}$;

(iv) $(1-q)\mathcal{R}(-u_0, u_0) + q\mathcal{R}(-u_1, u_1)$, где

$$u_0 = \frac{3h(a-1)}{a^{3/2}} < u_1 = \frac{3h(a-1)}{a} \quad \text{и} \quad q = \frac{aq_1 - 1}{a-1};$$

(vi) $(1-q)\delta_0 + q\mathcal{R}\left(-\frac{3(h-m)}{2}, \frac{3(h-m)}{2}\right)$ при $q = q_3$.

В случае (vii) граница (2) достигается на последовательности распределений $(1-q_n)\delta_0 + q_n\mathcal{R}(-\sqrt{3/q_n}, \sqrt{3/q_n})$, где $q_n \downarrow 0$.

Доказательство теоремы дано в Приложении.

Для сдвига произвольного знака во всех приведенных формулах необходимо заменить m на модуль $|m|$.

Важнейшим следствием сформулированной теоремы является тот факт, что граница \mathbf{g}_h из неравенства Гаусса (3) остается в силе для сравнительно малых значений сдвига.

Следствие 1. Неравенство Гаусса

$$(5) \quad \mathbf{P}\{|\xi - m| \geq h\} \leq \mathbf{g}_h = \begin{cases} 1 - \frac{h}{\sqrt{3}}, & h \leq \frac{2}{\sqrt{3}}, \\ \frac{4}{9h^2}, & h \geq \frac{2}{\sqrt{3}}, \end{cases}$$

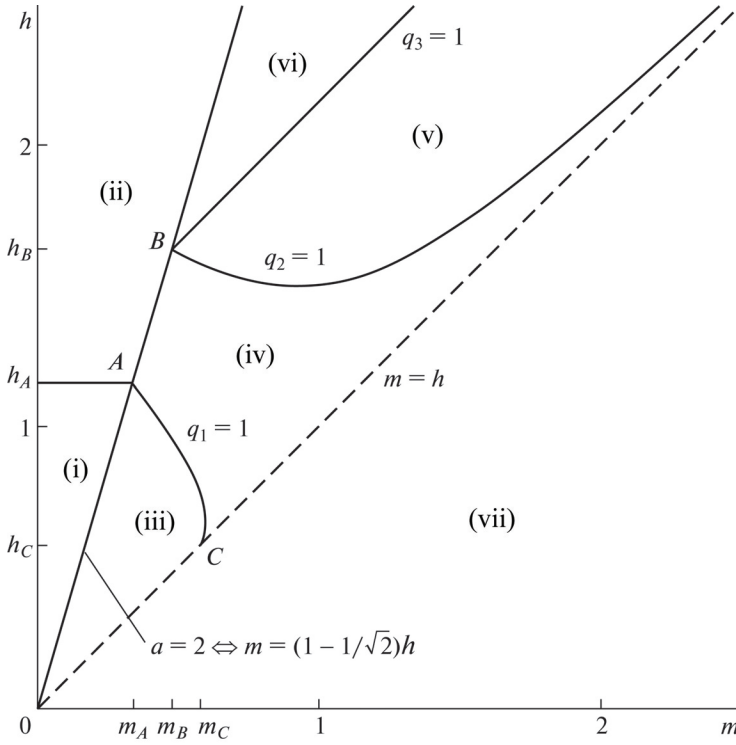


Рис. 1. Области (i)–(vii).

описывает неумлучшаемую верхнюю границу вероятности выхода случайной величины $\xi \sim \mathcal{U}$ за пределы интервала, смещение которого t относительно моды не превышает $1 - 1/\sqrt{2} \approx 0,2928$ доли радиуса h .

Каждый из случаев (i)–(vii) соответствует своей области, изображенной на рис. 1. Точки пересечения границ нескольких областей имеют следующие координаты:

$$m_A = (2 - \sqrt{2})/\sqrt{3} \approx 0,3382, \quad m_B = \sqrt{2} m_A \approx 0,4783, \quad m_C = 1/\sqrt{3} \approx 0,5774,$$

$$h_A = 2/\sqrt{3} \approx 1,1547, \quad h_B = \sqrt{2} h_A \approx 1,6330, \quad h_C = 1/\sqrt{3} \approx 0,5774.$$

На границе между любой из двух областей (i)–(vi) описанные наихудшие распределения совпадают, кроме двух пар: (ii), (iv) и (ii), (vi). Это означает, что на указанных границах в качестве наихудшего распределения можно взять смесь распределений, соответствующих внутренностям двух областей. Поэтому искомое распределение на общей границе имеет вид:

$$(1 - \varepsilon)(1 - q)\delta_0 + 2\varepsilon(1 - q)\mathcal{R}(-u_0, u_0) + (q - \varepsilon(1 - q))\mathcal{R}(-u_1, u_1) \quad \text{для (ii) и (iv),}$$

$$(1 - q - \varepsilon q)\delta_0 + 2\varepsilon q\mathcal{R}(-u_0, u_0) + (1 - \varepsilon)q\mathcal{R}(-u_1, u_1) \quad \text{для (ii) и (vi),}$$

где $q = 4/(3h^2)$, $u_0 = 3h/(2\sqrt{2})$, $u_1 = 3h/2$, а ε — произвольное число из отрезка $[0, 1]$.

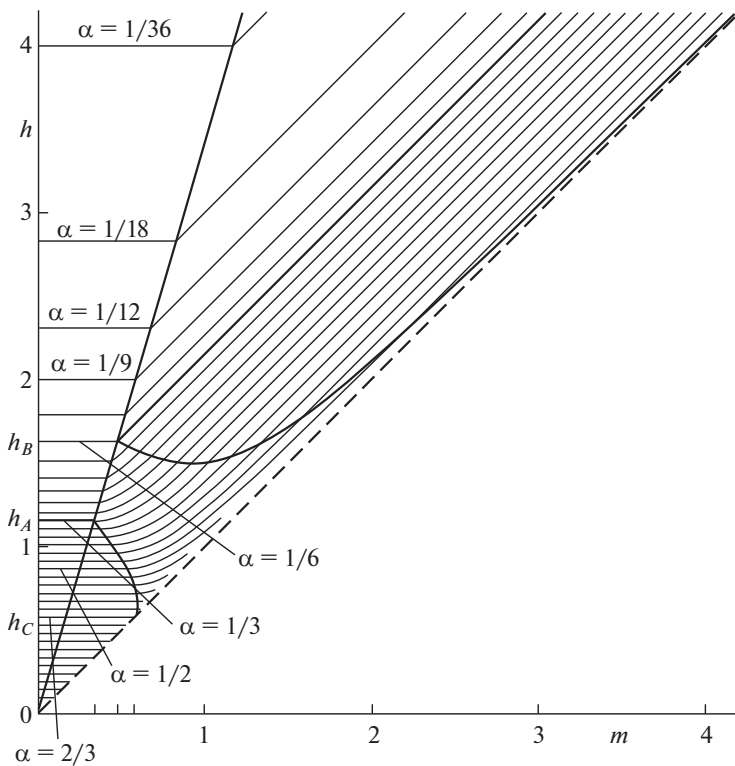


Рис. 2. Линии уровня $u_h(m) = \alpha$ для $\alpha = k/36$ при $k = 1, \dots, 35$.

При доказательстве теоремы был использован метод преобразования задачи (2) к проблеме моментов Маркова, т.е. к задаче оптимизации на классе неотрицательных мер без требований унимодальности, но при специальных моментных ограничениях. Анализ этой задачи показывает, что ее решение определено единственным образом при любом сочетании параметров m и h , взятых из области $|m| < h$ за исключением двух полупрямых: $m = \pm(1 - 1/\sqrt{2})h$, $h > h_A$. При положительных m это соответствует границам областей (ii), (iv) и (ii), (vi) кроме точки A (см. рис. 1).

Вид линий уровня функции $u_h(m)$, рассматриваемой на плоскости переменных (m, h) , представлен на рис. 2.

Из теоремы следует, что для нахождения границы $u_h(m)$ достаточно рассматривать только распределения, представляющие собой смесь двух равномерных. Чтобы получить явные соотношения, следующие из этого факта, определим вероятность выхода величины $\xi \sim \mathcal{R}(-u, u)$ из интервала $(m - h, m + h)$ при $0 \leq m < h$:

$$(6) \quad P\{|\xi - m| \geq h\} = \begin{cases} 0, & u \leq h - m, \\ \{1 - (h - m)/u\}/2, & h - m \leq u \leq h + m, \\ 1 - h/u, & u \geq h + m. \end{cases}$$

Если три выражения в (6) обозначить как $\rho_k(m, h; u)$, $k = 0, 1, 2$, заменив всюду m на $|m|$, то равенство (6) при $|m| < h$ можно записать в виде

$$P\{|\xi - m| \geq h\} = \max_{k=0,1,2} \rho_k(m, h; u).$$

Теперь сформулируем полученное утверждение.

Следствие 2. В определении границы (2) класс \mathcal{U} можно заменить на семейство распределений вида $(1 - q)\mathcal{R}(-u_0, u_0) + q\mathcal{R}(-u_1, u_1)$, где параметры подчиняются ограничениям

$$(7) \quad q \in [0, 1], \quad u_0 > 0, \quad u_1 > 0 \quad \text{и} \quad (1 - q)u_0^2 + qu_1^2 = 3.$$

Граница $\mathbf{u}_h(m)$ представляет собой непрерывную выпуклую функцию пары переменных (m, h) , таких что $|m| < h$, и может быть записана в виде

$$(8) \quad \mathbf{u}_h(m) = \sup\{(1 - q)\rho_k(m, h; u_0) + q\rho_k(m, h; u_1)\},$$

где супремум берется по всевозможным наборам (k, q, u_0, u_1) , в которых $k = 0, 1, 2$, а q, u_0, u_1 удовлетворяют (7). Но на диагоналях $m = \pm h$ имеет место разрыв:

$$\lim_{|m| \uparrow h} \mathbf{u}_h(m) = \begin{cases} 1 - \frac{h}{\sqrt{3}}, & h \leq \frac{1}{\sqrt{3}}, \\ \frac{1}{2} \left(1 + \frac{1}{9h^2} \right), & h \geq \frac{1}{\sqrt{3}}, \end{cases} \quad \lim_{|m| \downarrow h} \mathbf{u}_h(m) = 1.$$

Теперь перейдем к рассмотрению найденной границы как функции одной переменной.

На рис. 3 изображена зависимость вероятности $\alpha = \mathbf{u}_h(m)$ от сдвига $m \geq 0$ для нескольких значений порога $h > 0$. Полученные функции монотонно не убывают, выпуклы на промежутке $[0, h]$, терпят разрыв в точке $m = h$ и постоянны при $m \leq (1 - 1/\sqrt{2})h \approx 0,2928h$. Поэтому при относительно малых значениях сдвига неравенство Гаусса остается в силе. Например, при $m = 1$ и $h = 3$ неравенство Гаусса дает $4/81$, что лишь на 11 % меньше истинной границы $\mathbf{u}_h(m) = 1/18$. Если же $h = 2$, то при том же значении сдвига отличие между двумя границами становится существенным: $1/9 \approx 0,1111$ и $(1 - 1/\sqrt{3})/2 \approx 0,2113$ соответственно.

На рис. 3 представлены также границы областей (i)–(vii), описанных в теореме (см. также рис. 1 и 2). Общая граница для областей (i) и (ii) опущена, так как она представляет собой отрезок, совпавший с частью графика функции $\mathbf{u}_{h_A}(m)$.

Зависимость вероятностной границы $\alpha = \mathbf{u}_h(m)$ от величины порога $h \geq 0$ для нескольких значений m представлена на рис. 4. Полученные функции монотонно не возрастают, равны единице при $h \leq m$, имеют разрыв при $h = m$ и выпуклы при $h \geq m$. Случай $m = 0$ соответствует правой части неравенства Гаусса, т.е. $\mathbf{u}_h(0) = g_h$. График этой функции оказывается самой нижней кривой, на которую ложатся все графики $\alpha = \mathbf{u}_h(m)$, соответствующие

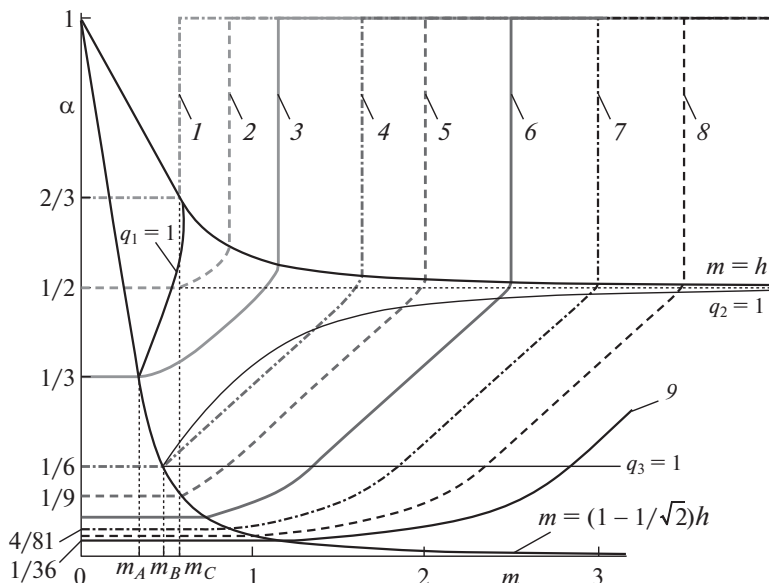


Рис. 3. Графики зависимости $\alpha = u_h(m)$ от m для нескольких значений h (толстые кривые 1-9: 1 — $h = h_C$; 2 — $h = 1,5h_C$; 3 — $h = h_A$; 4 — $h = h_B$; 5 — $h = 2$; 6 — $h = 2,5$; 7 — $h = 3$; 8 — $h = 3,5$; 9 — $h = 4$) и границы областей (тонкие сплошные кривые).

значениям параметра $m \leq m_A$ и рассматриваемые при $h > m$. С указанной кривой совпадает проекция прямой $m = (1 - 1/\sqrt{2})h$, которая служит границей нескольким областям (см. рис. 1). Часть этой кривой при $\alpha \in [1/3, 2/3]$ представляет собой проекцию общей границы областей (iii) и (iv). Граница между (i) и (ii) при проектировании на плоскость переменных (h, α) превратилась в точку $(h_A, 1/3)$.

Отсутствие непрерывности $u_h(m)$ на всей плоскости параметров (m, h) объясняется тем, что класс распределений \mathcal{U} не образует множества, компактного в слабой топологии. Если же носитель распределений P_ξ считать равномерно ограниченным, то аналогичная вероятностная граница будет непрерывной при любом соотношении между переменными m и h . Однако полное решение соответствующей задачи выходит за рамки данной статьи.

4. Сравнительный анализ

Найденную вероятностную границу можно использовать для анализа надежности доверительных интервалов и повышения их робастности.

Допустим, что для неизвестного параметра $\theta \in \mathbb{R}$ используется точечная оценка $\tilde{\theta}$, которая имеет фиксированную дисперсию $D\tilde{\theta} = 1$, неизвестное, но ограниченное смещение $m \in [-r, r]$ и неопределенное симметричное унимодальное распределение. Тогда $(\tilde{\theta} - h, \tilde{\theta} + h)$ называется *доверительным интервалом гарантированной надежности* $1 - \alpha$, если

$$(9) \quad P\{\theta \in (\tilde{\theta} - h, \tilde{\theta} + h)\} \geq 1 - \alpha.$$

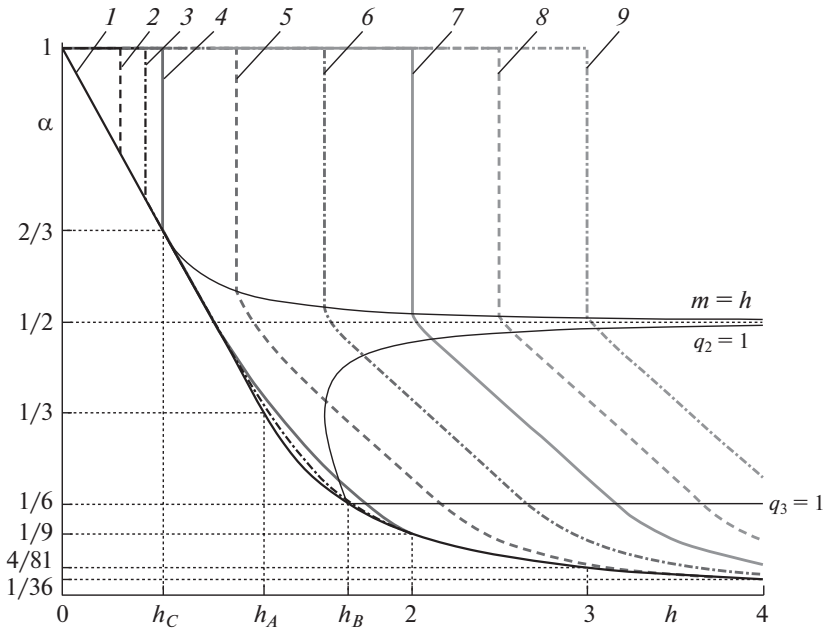


Рис. 4. Графики зависимости $\alpha = u_h(m)$ от h для нескольких значений m (толстые кривые 1–9: 1 – $m = 0$; 2 – $m = m_A$; 3 – $m = m_B$; 4 – $m = m_C$; 5 – $m = 1$; 6 – $m = 1,5$; 7 – $m = 2$; 8 – $m = 2,5$; 9 – $m = 3$) и границы областей (тонкие сплошные кривые).

Определение числа $1 - \alpha$ как неулучшаемой границы в этом неравенстве равносильно решению поставленной задачи. Чтобы в этом убедиться, введем в рассмотрение *центрированную ошибку* $\xi = \Delta\tilde{\theta} - m$ и запишем неравенство для противоположной вероятности $P\{|\xi + m| \geq h\} \leq \alpha$. Тогда для нахождения α необходимо вычислить точную верхнюю грань левой части по всем $\xi \sim \mathcal{U}$ и $|m| \leq r$. В силу леммы указанная вероятность монотонно зависит от $|m|$, поэтому достаточно взять $m = r$. Теперь искомый *уровень значимости* α будет определяться как вероятностная граница (2):

$$(10) \quad u_h(r) = \alpha.$$

Если выполнено (10), то будем говорить, что построен *робастный доверительный интервал* $(\tilde{\theta} - h, \tilde{\theta} + h)$ на уровне значимости α .

При этом интервал $(\tilde{\theta} - h, \tilde{\theta} + h)$ будет наименьшим среди интервалов вида $(\tilde{\theta} - h_1, \tilde{\theta} + h_2)$, для которых условие $P\{\theta \in (\tilde{\theta} - h_1, \tilde{\theta} + h_2)\} \geq 1 - \alpha$ выполнено при произвольном выборе ограниченного смещения $m \in [-r, r]$ и распределения центрированной ошибки из класса \mathcal{U} .

Описанные предположения возникают при использовании линейных смещенных оценок $\tilde{\theta} = \langle \ell, Y \rangle$ параметрической функции $\theta = \langle b, v \rangle$ в модели линейной регрессии $Y = Av + \eta$ в присутствии ограниченного вектора неизвестных параметров v и центрированной помехи η с неопределенным распределением симметричного и унимодального вида. Чтобы определить гарантирован-

ную надежность $1 - \alpha$ соответствующей интервальной оценки $(\tilde{\theta} - h, \tilde{\theta} + h)$ необходимо положить $\alpha = \mathbf{u}_{h/\sigma}(r/\sigma)$, где σ^2 — дисперсия оценки $\tilde{\theta}$, а r — наибольшее значение смещения $\langle A^* \ell - b, v \rangle$ с учетом априорных ограничений на вектор v .

Важность изучения смещенных оценок объясняется тем, что отказ от несмещенности позволяет снизить среднеквадратическую ошибку оценивания за счет использования априорных ограничений на неизвестные параметры. Более того, в моделях с большим числом факторов и в мультиколлинеарных регрессиях построение несмещенных оценок наталкивается на серьезные численные проблемы, связанные с точным соблюдением равенства $A^* \ell = b$.

Сравним характеристики интервальных оценок для нескольких гипотез о распределении централизованной ошибки ξ :

1) ξ — нормальная случайная величина, т.е. $\xi \sim \mathcal{N}(0, 1)$;

2) ξ имеет распределение Тьюки $\mathcal{T} = (1 - \varepsilon)\mathcal{N}(0, \sigma_0^2) + \varepsilon\mathcal{N}(0, \sigma_1^2)$, где доля выбросов ε равна 10%, а соотношение между номинальными значениями и выбросами задается дробью $\sigma_1/\sigma_0 = 5$ с учетом $D\xi = (1 - \varepsilon)\sigma_0^2 + \varepsilon\sigma_1^2 = 1$;

3) ξ определяет решение задачи (2), т.е. P_ξ — наихудший представитель класса симметричных унимодальных распределений \mathcal{U} ;

4) P_ξ — распределение, на котором достигается граница $\nu_h(m)$ из неравенства Высочанского — Петунина (4).

Во всех четырех случаях речь идет об унимодальной величине с единственными моментными характеристиками: $E\xi = 0$ и $D\xi = 1$. Первые три распределения симметричны относительно нуля, который одновременно выполняет роль единственной моды. Четвертое распределение имеет неопределенную моду и может быть несимметричным.

Теперь для каждого из четырех вариантов рассмотрим вероятность

$$(11) \quad P\left\{\theta \notin (\tilde{\theta} - h, \tilde{\theta} + h)\right\} = P\{|\xi + m| \geq h\}.$$

При $\xi \sim \mathcal{N}(0, 1)$ указанная вероятность равна

$$\nu_h(m) = \Psi(h - m) + \Psi(h + m),$$

где

$$\Psi(x) = \int_x^\infty \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt.$$

Для случая $\xi \sim \mathcal{T}$ вероятность (11) принимает вид

$$\begin{aligned} \tau_h(m) = (1 - \varepsilon) & \left[\Psi\left(\frac{h - m}{\sigma_0}\right) + \Psi\left(\frac{h + m}{\sigma_0}\right) \right] + \\ & + \varepsilon \left[\Psi\left(\frac{h - m}{\sigma_1}\right) + \Psi\left(\frac{h + m}{\sigma_1}\right) \right]. \end{aligned}$$

Таблица 1. Надежность доверительного интервала

m	h	$1 - \nu_h(m)$	$1 - \tau_h(m)$	$1 - \mathbf{u}_h(m)$	$1 - \mathbf{v}_h(m)$
0	0,866	0,6135	0,8258	0,5	0
0	1,282	0,8	0,9200	0,7294	0,5215
0	1,633	0,8975	0,9430	0,8333	0,8333
0	2	0,9545	0,9537	0,8889	0,8889
0	3	0,9973	0,9731	0,9506	0,9506
0	4	0,9999	0,9860	0,9722	0,9722
0,5	1,282	0,7453	0,8680	0,7074	0,3185
0,5	2	0,9270	0,9506	0,8889	0,8611
0,5	3	0,9936	0,9723	0,9506	0,9383
0,5	4	0,9998	0,9853	0,9722	0,9653
1	2	0,8400	0,9216	0,7887	0,6667
1	3	0,9772	0,9699	0,9444	0,9012
1	4	0,9986	0,9833	0,9722	0,9444
2	3	0,8413	0,9318	0,7887	0,5926
2	4	0,9772	0,9755	0,9444	0,8611
3	4	0,8413	0,9346	0,7887	0,5

Для третьего и четвертого вариантов распределений вероятность (11) совпадает со своим наихудшим значением: $\mathbf{u}_h(m)$ и $\mathbf{v}_h(m)$ соответственно (см. теорему и неравенство (4)). В отличие от первых двух случаев здесь распределение не является фиксированным и само зависит от параметров m и h .

В табл. 1 для каждой из четырех гипотез приведены данные о надежности доверительного интервала в зависимости от его радиуса h и величины смещения m . Первые три значения h выбраны из следующих соображений: во-первых, при $h = 1,5h_A \approx 0,866$ неравенство Гаусса дает $g_h = 0,5$; во-вторых, $h \approx 1,282$ — решение уравнения $\Psi(h) = 0,1$; в третьих, $h_B = \sqrt{8/3} \approx 1,633$ — наименьшее значение порога h , при котором границы Гаусса и Высочанского — Петунина совпадают, т.е. $g_h = \mathbf{v}_h(0)$.

В целом, можно отметить, что граница $\mathbf{u}_h(m)$ остается приемлемой при значительном смещении. Например, если пользоваться правилом “трех сигм”, то при наличии смещения в “одну сигму” надежность доверительного интервала будет: примерно 0,98 в гауссовском случае; почти 0,97 в модели Тьюки; не меньше 0,94 в случае симметричного унимодального распределения; около 0,9 при наихудшем выборе моды.

Теперь найдем радиус h робастного доверительного интервала $(\tilde{\theta} - h, \tilde{\theta} + h)$, построенного на уровне значимости α . Для этого необходимо определить наибольшую вероятность события $\{|\xi + m| \geq h\}$ с учетом ограниченности смещения m и априорной информации о распределении центрированной ошибки ξ . Максимум по $m \in [-r, r]$ будет достигаться при $m = \pm r$ в силу симметричной и монотонной зависимости от m (см. лемму и выражения для границ $\mathbf{u}_h(m)$ и $\mathbf{v}_h(m)$). Следовательно, искомый радиус h находится из уравнения, в одной части которого стоит α , а в другой — одна из четырех вероятностей: $\nu_h(r)$, $\tau_h(r)$, $\mathbf{u}_h(r)$ или $\mathbf{v}_h(r)$.

Таблица 2. Радиус робастного доверительного интервала

α	r	$h: \nu_h(r) = \alpha$	$h: \tau_h(r) = \alpha$	$h: \mathbf{u}_h(r) = \alpha$	$h: \mathbf{v}_h(r) = \alpha$
0,2	0	1,282	0,806	1,491	1,581
0,2	0,5	1,439	1,078	1,542	1,768
0,1	0	1,645	1,137	2,108	2,108
0,1	0,5	1,839	1,427	2,108	2,357
0,1	1	2,284	1,874	2,491	2,981
0,05	0	1,960	1,853	2,981	2,981
0,05	0,5	2,181	1,984	2,981	3,333
0,05	1	2,646	2,303	3,108	4,216
0,05	1,5	3,145	2,715	3,608	5,375
0,025	0	2,241	3,119	4,216	4,216
0,025	0,5	2,485	3,172	4,216	4,714
0,025	1	2,961	3,332	4,216	5,963
0,025	1,5	3,460	3,599	4,481	7,601
0,025	2	3,960	3,962	4,981	9,428

В табл. 2 для каждой из четырех гипотез указаны значения радиуса h робастного доверительного интервала в зависимости от уровня значимости α и границы смещения r . Стоит отметить, что использование границы Высочанского — Петунина в ситуации большого смещения приводит к неоправданно широким интервалам. Вместе с тем граница $\mathbf{u}_h(r)$ в этой ситуации дает возможность построить робастный доверительный интервал приемлемой длины. Например, при наличии смещения в “одну сигму” такой интервал лишь на четверть длиннее интервала, построенного для объединенной гипотезы: $N(0, 1)$ или \mathcal{T} .

Итак, при использовании симметричных унимодальных статистик с неопределенным распределением и ограниченным смещением новая вероятностная граница $\mathbf{u}_h(m)$ позволяет получать робастные статистические решения, существенно менее консервативные, чем известные ранее.

5. Заключение

В статье найдена точная верхняя грань вероятности попадания вне конечного интервала для случайной величины с неопределенным симметричным унимодальным распределением и фиксированными модой и дисперсией. Специфика этой задачи состоит в том, что мода смещена относительно середины интервала. Однако если смещение составляет меньше 29 % радиуса интервала, то найденная вероятностная граница определяется неравенством Гаусса, т.е. оказывается такой же, как в случае нулевого смещения. Полученный результат позволил построить робастные версии интервальных оценок, которые сохраняют приемлемую точность даже при наличии неизвестного ограниченного смещения центральной статистики. Решенная в статье задача расширяет инструментарий робастных методов анализа данных и идентификации систем.

Доказательство леммы. Вероятность $\pi_h(m)$ можно записать через функцию распределения $F(x)$ с учетом симметрии распределения:

$$\pi_h(m) = \mathbf{P}\{\xi \leq m - h\} + \mathbf{P}\{\xi \geq m + h\} = F(m - h) + F(-m - h).$$

Ясно, что $\pi_h(m) = \pi_h(-m)$. Выпуклость по $m \in [-h, h]$ следует из того, что $F(x)$ выпукла при $x \leq 0$. Если $m \downarrow h$, то $\pi_h(m) \rightarrow \pi_h(h)$ в силу того, что $F(x)$ непрерывна справа в точке $x = 0$ и непрерывна при $x < 0$. Величина скачка $\pi_h(m)$ в точке $m = h$ равна неотрицательной величине $1 - q$. Других точек разрыва у этой функции, рассматриваемой при $m \geq 0$, нет. Поэтому для проверки монотонности по $m \geq 0$ остается проверить знак производной. Ее можно записать в виде $q(f(m - h) - f(-m - h))$ в силу (1). Если $m \leq h$, то производная неотрицательна в силу того, что $f(x)$ не убывает при $x \leq 0$. Если же $m \geq h$, то производная равна $q(f(m - h) - f(m + h))$ в силу $f(x) = f(-x)$ и неотрицательна благодаря тому, что $f(x)$ не возрастает при $x \geq 0$. Тем самым лемма доказана.

При доказательстве основного результата понадобится утверждение, позволяющее преобразовать задачу оптимизации на классе унимодальных распределений к проблеме моментов Маркова.

Лемма П.1. Пусть \mathfrak{M} — класс неотрицательных счетно-аддитивных мер μ , заданных на борелевских подмножествах открытого интервала $(0, \infty)$.

Существует взаимно однозначное соответствие между распределениями \mathbf{P}_ξ из класса $\mathcal{U}(d)$ и мерами $\mu \in \mathfrak{M}$, которые подчиняются ограничениям

$$(П.1) \quad \int_0^\infty y \mu(dy) \leq 1 \quad \text{и} \quad \int_0^\infty \frac{y^3}{3} \mu(dy) = d.$$

Указанное соответствие определено равенством

$$(П.2) \quad 2 \int_0^\infty g(x) \mathbf{P}_\xi(dx) = \int_0^\infty \left\{ \int_0^y g(x) dx \right\} \mu(dy),$$

где $g(x)$ — произвольная борелевская функция, такая что $g(0) = 0$ и $\mathbf{E}|g(\xi)| < \infty$.

Доказательство леммы П.1. Для каждого $\mathbf{P}_\xi \in \mathcal{U}(d)$ зададим меру Лебега — Стильеса μ :

$$\mu((a, b]) = 2q(f(a) - f(b)), \quad 0 < a < b < \infty,$$

где число q и функция $f(x)$ взяты из представления (1), причем $f(x)$ выбрана непрерывной справа при $x > 0$. Из теоремы Каратеодори в силу монотонности $f(x)$ следует $\mu \in \mathfrak{M}$. Запишем интеграл в левой части (П.2) с помощью

равенства $2qf(x) = \mu((x, \infty))$:

$$\int_0^{\infty} g(x) \mu((x, \infty)) dx = \int_0^{\infty} g(x) \left\{ \int_0^{\infty} \mathbf{I}\{y > x\} \mu(dy) \right\} dx.$$

Меня местами интегралы, приходим к правой части равенства (П.2).

Построенное соответствие взаимно однозначно, так как

$$f(x) = \mu((x, \infty))/q, \quad q = \mathbf{P}\{\xi \neq 0\} = 2 \int_0^{\infty} \mathbf{I}\{x > 0\} \mathbf{P}_{\xi}(dx) = \int_0^{\infty} y \mu(dy).$$

Остается проверить соотношения (П.1). Первое из них вытекает из условия нормировки, доказанного равенства (П.2) и условия $q \leq 1$:

$$1 = 1 - q + 2 \int_0^{\infty} \mathbf{I}\{x > 0\} \mathbf{P}_{\xi}(dx) = 1 - q + \int_0^{\infty} y \mu(dy).$$

Второе соотношение (П.1) следует из предположения $E\xi^2 = d$ и равенства (П.2), в котором $g(x) = x^2$. Лемма П.1 доказана.

Доказательство теоремы. Начнем с проверки случая (vii): $m \geq h$. В этом случае интервал $(m - h, m + h)$ не содержит нуля, поэтому вероятность попадания в него случайной величины

$$\xi_n \sim (1 - q_n)\delta_0 + q_n\mathcal{R}(-\sqrt{3/q_n}, \sqrt{3/q_n})$$

равна $q_n h \sqrt{q_n/3}$ для всех достаточно малых $q_n > 0$. Если $q_n \downarrow 0$, то противоположная вероятность сходится к единице, что с учетом $\xi_n \sim \mathcal{U}$ дает $\mathbf{u}_h(m) = 1$.

Пусть теперь и далее $0 \leq m < h$. В силу симметрии распределения имеем

$$\pi_h(m) = \mathbf{P}\{\xi \geq h - m\} + \mathbf{P}\{\xi \geq h + m\} = \int_0^{\infty} g(x) \mathbf{P}_{\xi}(dx),$$

где $g(x) = \mathbf{I}\{x \geq h - m\} + \mathbf{I}\{x \geq h + m\}$. Первообразная этой функции имеет вид

$$G(y) = \int_0^y g(x) dx = \max\{0, y - (h - m), 2(y - h)\}.$$

Если в определении искомой границы (2) расширить класс распределений до $\bigcup_{d \leq 1} \mathcal{U}(d)$, то благодаря лемме П.1 получится неравенство

$$\mathbf{u}_h(m) \leq \frac{1}{2} \sup_{\mu \in \mathfrak{M}} \left\{ \int_0^{\infty} G(y) \mu(dy) : \int_0^{\infty} y \mu(dy) \leq 1, \int_0^{\infty} \frac{y^3}{3} \mu(dy) \leq 1 \right\}.$$

Нахождение супремума в правой части составляет проблему моментов Маркова. Для ее решения введем множители λ_1, λ_2 , функцию $H_{\lambda}(y) = \lambda_1 y +$

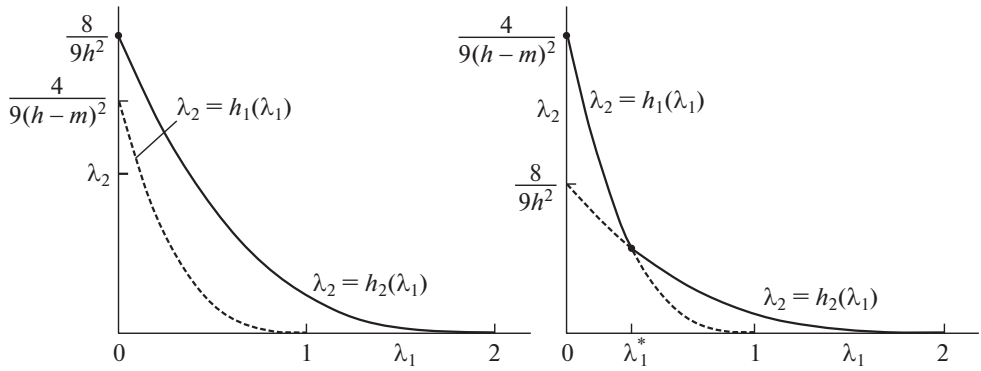


Рис. 5. Зависимости $\lambda_2 = h_1(\lambda_1)$, $\lambda_2 = h_2(\lambda_1)$ и граница множества Λ (сплошная кривая); точки и участки сплошной кривой соответствуют областям (i)–(vi); слева изображен случай $h_1(0) < h_2(0)$, справа — случай $h_1(0) > h_2(0)$.

+ $\lambda_2 y^3/3$ и воспользуемся стандартным переходом к двойственной задаче (см., например, [2]):

$$(П.3) \quad \mathbf{u}_h(m) \leq \frac{1}{2} \inf_{\lambda_1, \lambda_2} \{ \lambda_1 + \lambda_2 : G(y) \leq H_\lambda(y) \forall y > 0, \lambda_1 \geq 0, \lambda_2 \geq 0 \}.$$

Выполнение неравенства $G \leq H_\lambda$ справа от нуля равносильно тому, что график $z = H_\lambda(y)$ лежит не ниже двух прямых: $z = y - (h - m)$ и $z = 2(y - h)$. Достаточно рассмотреть только те множители λ_1, λ_2 , при которых график $z = H_\lambda(y)$ касается хотя бы одной из двух прямых. Для первой прямой условие касания означает, что равенства $H_\lambda(y) = y - (h - m)$ и $H'_\lambda(y) = 1$ выполнены для некоторого $y > 0$. Для второй прямой аналогично: $H_\lambda(y) = 2(y - h)$ и $H'_\lambda(y) = 2$. После несложных выкладок указанные условия преобразуются в два уравнения: $\lambda_2 = 4(1 - \lambda_1)^3/(9(h - m)^2)$ при $\lambda_1 \in [0, 1]$ и $\lambda_2 = (2 - \lambda_1)^3/(9h^2)$ при $\lambda_1 \in [0, 2]$. Обозначим эти зависимости как $\lambda_2 = h_1(\lambda_1)$ и $\lambda_2 = h_2(\lambda_1)$ соответственно. Тогда множество, образованное множителями λ_1, λ_2 из (П.3), можно записать в виде

$$\Lambda = \{ (\lambda_1, \lambda_2) : \lambda_2 \geq h_1(\lambda_1), \lambda_2 \geq h_2(\lambda_1), \lambda_1 \geq 0, \lambda_2 \geq 0 \}.$$

Вид этого множества зависит от соотношения между числами $h_1(0) = 4/(9(h - m)^2)$ и $h_2(0) = 8/(9h^2)$ (см. рис. 5). Если выполнено $h_1(0) \leq h_2(0)$, т.е. $m \leq (1 - 1/\sqrt{2})h$, то график второй функции лежит выше графика первой. Если же $(1 - 1/\sqrt{2})h \leq m$, то графики пересекаются в точке $\lambda_1 = \lambda_1^*$, где $\lambda_1^* = (a - 2)/(a - 1)$, $a = (2h/(h - m))^{2/3}$, причем $0 \leq \lambda_1^* \leq 1$ и $a \geq 2$.

Для решения (П.3) в случае $m \leq (1 - 1/\sqrt{2})h$ достаточно определить минимум функции $\lambda_1 + h_2(\lambda_1)$ по $\lambda_1 \in [0, 2]$. Так как производная равна $1 - (2 - \lambda_1)^2/(3h^2)$, то минимум на бесконечном промежутке $(-\infty, 2]$ достигается в точке $\lambda_1'' = 2 - h\sqrt{3}$. В зависимости от знака ($\lambda_1'' \geq 0$ или $\lambda_1'' \leq 0$) получаем случаи, описанные в формулировке теоремы: (i) $h \leq 2/\sqrt{3}$ и (ii) $h \geq 2/\sqrt{3}$ соответственно. В случае (i) подстановка $\lambda_1 = \lambda_1''$, $\lambda_2 = h_2(\lambda_1'')$

Таблица 3

(iii)	(iv)	(v)	(vi)
$1 \leq q_1$	$q_1 \leq 1 \leq q_2$	$q_2 \leq 1 \leq q_3$	$q_3 \leq 1$
$\lambda_1^* \leq \lambda_1''$	$\lambda_1'' \leq \lambda_1^* \leq \lambda_1'$	$0 \leq \lambda_1' \leq \lambda_1^*$	$\lambda_1' \leq 0$

в выражение $(\lambda_1 + \lambda_2)/2$ дает $1 - h/\sqrt{3}$, а в случае (ii) минимум достигается при $\lambda_1 = 0$, откуда $\lambda_2 = h_2(0)$ и $(\lambda_1 + \lambda_2)/2 = 4/(9h^2)$.

Теперь предположим, что выполнено $(1 - 1/\sqrt{2})h \leq m < h$. В этом случае точную нижнюю грань в (П.3) можно найти как минимум выражения, определяемого двумя функциями: $\lambda_1 + h_1(\lambda_1)$ при $\lambda_1 \in [0, \lambda_1^*]$ и $\lambda_1 + h_2(\lambda_1)$ при $\lambda_1 \in [\lambda_1^*, 2]$. Для этого понадобятся точки λ_1' и λ_1'' , в которых эти функции достигают минимума на промежутках $(-\infty, 1]$ и $(-\infty, 2]$ соответственно. Точка λ_1'' указана выше, а точка $\lambda_1' = 1 - (h - m)\sqrt{3}/2$ находится из равенства нулю производной $1 - 4(1 - \lambda_1)^2/(3(h - m)^2)$. Нетрудно проверить, что следующие пары неравенств эквивалентны: $\lambda_1' \leq 0$ и $q_3 \leq 1$; $\lambda_1' \leq \lambda_1^*$ и $q_2 \leq 1$; $\lambda_1'' \leq \lambda_1^*$ и $q_1 \leq 1$, причем $q_1 \leq q_2 \leq q_3$. В зависимости от того, как расположена единица относительно чисел q_1, q_2, q_3 , получаются четыре варианта взаимного расположения точек $0, \lambda_1', \lambda_1'', \lambda_1^*$. Эти варианты (см. табл. 3) соответствуют случаям (iii)–(vi), перечисленным в формулировке теоремы.

В случае (iii) минимум в (П.3) достигается при $\lambda_1 = \lambda_1''$, $\lambda_2 = h_2(\lambda_1'')$. Это идентично случаю (i): $(\lambda_1 + \lambda_2)/2 = 1 - h/\sqrt{3}$.

В случае (iv) точкой минимума будет $(\lambda_1^*, \lambda_2^*)$, где

$$\lambda_2^* = h_2(\lambda_1^*) = \frac{1}{9h^2} \left(\frac{a}{a-1} \right)^3,$$

откуда следует, что

$$\frac{\lambda_1^* + \lambda_2^*}{2} = \frac{1}{2} \left\{ \frac{a-2}{a-1} + \frac{1}{9h^2} \left(\frac{a}{a-1} \right)^3 \right\}.$$

В случае (v) минимум достигается при $\lambda_1 = \lambda_1'$, $\lambda_2 = h_1(\lambda_1') = (h - m)\sqrt{3}/6$, поэтому $(\lambda_1 + \lambda_2)/2 = (1 - (h - m)/\sqrt{3})/2$.

Наконец, в случае (vi) точка минимума определяется через $\lambda_1 = 0$ и $\lambda_2 = h_1(0)$. Тогда $(\lambda_1 + \lambda_2)/2 = 2/(9(h - m)^2)$.

Итак, доказано неравенство: искомая граница $u_h(m)$ не превосходит выражений, представленных в теореме для каждого из случаев (i)–(vi). Чтобы установить равенство, остается предъявить распределение, на котором полученная верхняя оценка достигается.

Возьмем $\xi \sim \mathcal{R}(-\sqrt{3}, \sqrt{3})$ и воспользуемся представлением (6). Тогда имеет место равенство $P\{|\xi - m| \geq h\} = 1 - h/\sqrt{3}$ для всех точек (m, h) , лежащих ниже прямой $h + m = \sqrt{3}$. Судя по рис. 1, области (i) и (iii) лежат ниже указанной прямой. Поэтому утверждение теоремы доказано для случаев (i) и (iii).

Теперь возьмем точку (m, h) между прямыми $h + m = \sqrt{3}$ и $h - m = \sqrt{3}$. Тогда в силу (6) справедливо $P\{|\xi - m| \geq h\} = (1 - (h - m)/\sqrt{3})/2$. Так как

область (v) лежит между указанными прямыми, то утверждение теоремы доказано для случая (v).

Предположим, что $\xi \sim (1 - q)\delta_0 + q\mathcal{R}(-u, u)$, где $u = 3h/2$ и $q = 4/(3h^2)$. Из (6) следует $P\{|\xi - m| \geq h\} = q(1 - h/u) = 4/(9h^2)$, если $u \geq h + m$. Последнее неравенство означает, что точка (m, h) лежит выше прямой $h = 2m$. Это заведомо так для точек области (ii). Поэтому утверждение теоремы установлено для случая (ii).

Проверим, что в случае (iv) справедливы неравенства

$$(II.4) \quad h - m < u_0 < h + m < u_1,$$

где $u_0 = 3h(a - 1)/a^{3/2}$, $u_1 = 3h(a - 1)/a$ и $a^{3/2} = 2h/(h - m)$. Запишем последнее равенство как $m/h = 1 - 2/a^{3/2}$ и подставим его в (II.4):

$$2/a^{3/2} < 3(a - 1)/a^{3/2} < 2 - 2/a^{3/2} < 3(a - 1)/a.$$

Первое неравенство равносильно условию $a > 5/3$, которое заведомо выполнено, так как $a \geq 2$ в случае (iv). Если во втором неравенстве обозначить $x = a$, а в третьем положить $x = 1/a$, то оба неравенства превращаются в одно и то же соотношение $2x^{3/2} - 3x + 1 > 0$, где $x \in (0, 1/2] \cup [2, \infty)$. Это соотношение верно, так как при $x > 0$ минимум этой функции равен нулю и достигается в единственной точке $x = 1$.

Теперь рассмотрим

$$\xi \sim (1 - q)\mathcal{R}(-u_0, u_0) + q\mathcal{R}(-u_1, u_1),$$

где $q = (aq_1 - 1)/(a - 1)$, $q_1 = a^2/(3h^2(a - 1)^2)$. В силу (II.4) с помощью (6) получаем

$$P\{|\xi - m| \geq h\} = (1 - q) \left(\frac{1}{2} - \frac{h - m}{2u_0} \right) + q \left(1 - \frac{h}{u_1} \right).$$

Если обозначить $p = (a/(a - 1))^3/(3h^2)$, то веса можно записать как $q = p - 1/(a - 1)$ и $1 - q = a/(a - 1) - p$. Тогда искомая вероятность принимает вид

$$\begin{aligned} & \left(\frac{a}{a - 1} - p \right) \left(\frac{1}{2} - \frac{1}{3(a - 1)} \right) + \left(p - \frac{1}{a - 1} \right) \left(1 - \frac{a}{3(a - 1)} \right) = \\ & = p \left(\frac{1}{2} + \frac{1 - a}{3(a - 1)} \right) + \frac{1}{a - 1} \left(\frac{a}{2} - 1 \right) = \frac{p}{6} + \frac{a - 2}{2(a - 1)}, \end{aligned}$$

что совпадает с выражением, указанным в теореме для случая (iv).

Остается рассмотреть только случай (vi). Пусть величина ξ распределена по закону $(1 - q)\delta_0 + q\mathcal{R}(-u, u)$, где $u = 3(h - m)/2$, $q = 4/(3(h - m)^2)$. Чтобы применить формулу (6), проверим два неравенства $h - m < u < h + m$. Первое из них очевидно, а второе можно записать в виде соотношения $m > 0,2h$, которое выполнено на области (vi), так как $m \geq (1 - 1/\sqrt{2})h > 0,29h$. Теперь вероятность $P\{|\xi - m| \geq h\}$ равна $q(1 - (h - m)/u)/2 = q/6 = 2/(9(h - m)^2)$, что и требовалось.

Итак, теорема полностью доказана.

СПИСОК ЛИТЕРАТУРЫ

1. *Dharmadhikari S., Joag-Dev K.* Unimodality, Convexity, and Applications. San Diego: Academic, 1988.
2. *Карлин С., Стадден В.* Чебышевские системы и их применение в анализе и статистике. М.: Наука, 1976.
3. *Высочанский Д.Ф., Петунин Ю.И.* Об одном неравенстве Гаусса для одновершинных распределений // Теория вероятн. и ее примен. 1982. Т. 27. № 2. С. 339–341.
Vysochanskii D.F., Petunin Yu.I. On a Gauss Inequality for Unimodal Distributions // Theory Probab. Appl. 1983. V. 27. No. 2. P. 359–361.
4. *Pukelsheim F.* The Three Sigma Rule // The American Statistician. 1994. V. 48. P. 88–91.
5. *Петунин Ю.И.* Приложение теории случайных процессов в биологии и медицине. Киев: Наук. думка, 1981.
6. *Chiles J.-P., Delfiner P.* Geostatistics. Modeling Spatial Uncertainty. N.Y.: J. Wiley & Sons, 1999.
7. *Barlow P., Proschan F.* Mathematical Theory of Reliability. Philadelphia: SIAM, 1996.
8. *Чернышев К.Р.* Вероятностные неравенства чебышевского типа в одной задаче робастного управления // Информационно-управляющие системы. 2010. № 3. С. 9–12.
9. *Barmish B.R., Lagoa C.M.* The Uniform Distribution: A Rigorous Justification for Its Use in Robustness Analysis // Math. Control Signal. Syst. 1997. V. 10. P. 203–222.
10. *Кибзун А.И.* О наихудшем распределении в задачах стохастической оптимизации с функцией вероятности // АиТ. 1998. № 11. С. 104–116.
Kibzun A.I. On the Worst-Case Distribution in Stochastic Optimization Problems with Probability Function // Autom. Remote Control. 1998. V. 59. No. 11. P. 1587–1597.
11. *Кан Ю.С.* Об обосновании принципа равномерности в задаче оптимизации вероятностного показателя качества // АиТ. 2000. № 1. С. 54–70.
Kan Yu.S. On the Justification of the Uniformity Principle in the Optimization of a Probability Performance Index // Autom. Remote Control. 2000. V. 61. No. 1. P. 50–64.
12. *Stellato B.* Data-Driven Chance Constrained Optimization: Master thesis / Swiss Federal Institute of Technology (ETH) Zürich. Zürich, 2014.
<https://doi.org/10.3929/ethz-a-010266857>
13. *Van Parys B.P.G., Goulart P.J., Kuhn D.* Generalized Gauss Inequalities via Semidefinite Programming // Math. Program. 2016. V. 156. P. 271–302.
14. *Ulin B.* An Extremal Problem in Mathematical Statistics // Scandinavian Actuarial J. 1953. No. 1. P. 158–167.
15. *Clarkson E., Denny J.L., Shepp L.* ROC and the Bounds on Tail Probabilities via Theorems of Dubins and F. Riesz // Ann. Appl. Probab. 2009. V. 19. No. 1. P. 467–476.
16. *Dharmadhikari S.W., Joag-Dev K.* The Gauss–Tchebyshev Inequality for Unimodal Distributions // Теория вероятн. и ее примен. 1985. Т. 30. № 4. С. 817–820.
Dharmadhikari S.W., Joag-Dev K. The Gauss–Tchebyshev Inequality for Unimodal Distributions // Theor. Probab. Appl. 1986. V. 30. No. 4. P. 867–871.

17. *Popescu I.* A Semidefinite Programming Approach to Optimal Moment Bounds for Convex Classes of Distributions // *Math. Oper. Res.* 2005. V. 30. No. 3. P. 632–657.
18. *Стойкова Л.С.* Наибольшая точная нижняя граница вероятности отказа системы в специальном интервале времени при неполной информации о функции распределения времени до отказа системы // *Кибернетика и системный анализ.* 2017. Т. 53. № 2. С. 65–73.
Stoikova L.S. Greatest Lower Bound of System Failure Probability on a Special Time Interval under Incomplete Information about the Distribution Function of the Time to Failure of the System // *Cybern. Syst. Anal.* 2017. V. 53. No. 2. P. 217–224.
19. *Панков А.Р., Семенikhин К.В.* О минимаксном оценивании по вероятностному критерию // *АиТ.* 2007. № 3. С. 66–82.
Pankov A.R., Semenikhin K.V. Minimax Estimation by Probabilistic Criterion // *Autom. Remote Control.* 2007. V. 68. No. 3. P. 430–445.
20. *Семенikhин К.В.* Минимаксность линейных оценок неопределенно-стохастического вектора по обобщенным вероятностным критериям // *АиТ.* 2007. № 11. С. 88–104.
Semenikhin K.V. Minimax Nature of the Linear Estimates of the Indefinite Stochastic Vector from the Generalized Probabilistic Criteria // *Autom. Remote Control.* 2007. V. 68. No. 11. P. 1970–1985.

Статья представлена к публикации членом редколлегии А.И. Кибзуном.

Поступила в редакцию 05.09.2018

После доработки 05.10.2018

Принята к публикации 08.11.2018