

ISSN 1561-2430 (Print)
ISSN 2524-2415 (Online)

ИНФОРМАТИКА INFORMATICS

УДК 519.2,519.7
<https://doi.org/10.29235/1561-2430-2026-62-2-143-148>

Поступила в редакцию 04.06.2025
Received 04.06.2025

М. С. Баркетов

*Объединенный институт проблем информатики Национальной академии наук Беларуси,
Минск, Республика Беларусь*

ПРИБЛИЖЕННЫЕ АЛГОРИТМЫ ПОСТРОЕНИЯ ОЦЕНОК МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ ПАРАМЕТРОВ ТРАПЕЦИЕВИДНОГО РАСПРЕДЕЛЕНИЯ

Аннотация. Рассмотрена задача оценивания параметров трапециевидного распределения. Построен доверительный интервал, в котором находятся возможные значения случайной величины с трапециевидным распределением. Разработаны два приближенных алгоритма построения оценок максимального правдоподобия, основанные на ограниченном переборе возможных значений параметров. Трапециевидное распределение может быть использовано в задачах моделирования вместо нормального распределения, когда есть нижние и верхние границы на значение случайной величины.

Ключевые слова: трапециевидное распределение, метод максимального правдоподобия, математическая статистика

Для цитирования. Баркетов, М. С. Приближенные алгоритмы построения оценок максимального правдоподобия параметров трапециевидного распределения / М. С. Баркетов // Весті Нацыянальнай акадэміі навук Беларусі. Серыя фізіка-матэматычных навук. – 2026. – Т. 62, № 2. – С. 143–148. <https://doi.org/10.29235/1561-2430-2026-62-2-143-148>

Maksim S. Barketau

United Institute for Informatics Problems of the National Academy of Sciences of Belarus, Minsk, Republic of Belarus

APPROXIMATION ALGORITHMS FOR THE MAXIMUM LIKELIHOOD ESTIMATION OF THE PARAMETERS OF THE TRAPEZOIDAL DISTRIBUTION

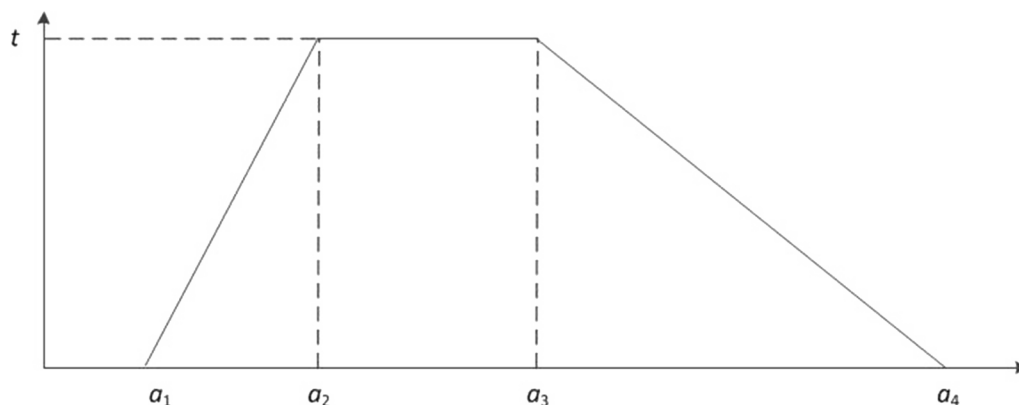
Abstract. We considered the problem of estimation of the parameters of the trapezoidal distribution. We built a confidence interval. This confidence interval with high probability contains the value of the random variable with a trapezoidal distribution. We further developed two approximation algorithms for the problem that are based on the maximum likelihood principle. Algorithm uses the enumeration principle of the values of the parameters. The trapezoidal distribution may be used instead of normal distribution when there are upper and lower bounds on the value of the random variable.

Keywords: trapezoidal distribution, method of maximum likelihood, mathematical statistics

For citation. Barketau M. S. Approximation algorithms for the maximum likelihood estimation of the parameters of the trapezoidal distribution. *Vestsi Natsyyanal'nai akademii navuk Belarusi. Seryya fizika-matematychnykh navuk = Proceedings of the National Academy of Sciences of Belarus. Physics and Mathematics series*, 2026, vol. 62, no. 2, pp. 143–148 (in Russian). <https://doi.org/10.29235/1561-2430-2026-62-2-143-148>

Введение. В данной работе рассматривается задача оценивания параметров трапециевидного распределения с помощью метода максимального правдоподобия.

Функция плотности трапециевидного распределения изображена на рисунке, где $a_1 < a_2 \leq a_3 < a_4$, а высота трапеции t вычисляется, исходя из свойства функции плотности распределения или, другими словами, из того, что площадь нарисованной трапеции должна быть равна единице:



Плотность трапециевидного распределения

Density of the trapezoidal distribution

$$t = \frac{2}{a_3 + a_4 - a_1 - a_2}. \quad (1)$$

Предполагается, что трапециевидное распределение не вырождено (знаки строгого равенства между параметрами). Если не делать этого предположения, то в случае, если все элементы выборки, полученные из трапециевидного распределения, равны между собой, можно за оценку распределения принимать постоянное значение, равное элементу выборки. В случае, если есть хотя бы два разных значения в выборке, мы предлагаем приближенный подход к вычислению параметров трапециевидного распределения с помощью метода максимального правдоподобия.

В работах [1; 2] рассматривались различные свойства трапециевидного распределения. В частности, были подсчитаны математическое ожидание и дисперсия такого распределения. В статье [3] был предложен подход к сравнению оценок на основе минимизации их среднеквадратичного отклонения. Ключевые рассматриваемые оценки – это выборочное среднее для нормального распределения и двухкомпонентные оценки для симметричного трапециевидного распределения. В данной работе мы рассматриваем задачу определения всех параметров произвольного трапециевидного распределения с помощью метода максимального правдоподобия, а не исследуем оценки с точки зрения среднеквадратичного отклонения.

Нам неизвестно о других исследованиях, в которых бы рассматривалась задача приближенного или неприменного оценивания всех параметров трапециевидного распределения. На наш взгляд, это связано с тем, что распределение зависит от четырех параметров и функция плотности, которую необходимо максимизировать в методе максимального правдоподобия, имеет нетривиальный нелинейный вид.

Метод максимального правдоподобия. Рассмотрим функцию плотности трапециевидного распределения $p_{a_1, a_2, a_3, a_4}(x)$ (иногда далее будем просто обозначать ее $p(x)$). И пусть дана выборка независимых значений из этого трапециевидного распределения $x_1, x_2, \dots, x_n$. Тогда многомерная функция плотности в данной точке равна

$$p(x_1)p(x_2)\dots p(x_n). \quad (2)$$

Метод максимального правдоподобия предполагает выбор таких допустимых значений параметров a_1, a_2, a_3, a_4 , которые бы максимизировали эту многомерную функцию плотности. (Подробнее о методе максимального правдоподобия см. в [4].)

Доверительный интервал. Рассмотрим два подхода к построению доверительного интервала, основанных на неравенстве Чебышева и на оценках для субгауссовской случайной величины.

Теорема [5, с. 81]. Для произвольной случайной величины ζ , имеющей математическое ожидание, справедливо неравенство

$$P(|\zeta - E\zeta| \geq \varepsilon) \leq \frac{D\zeta}{\varepsilon^2}, \quad (3)$$

где $E\zeta$ – математическое ожидание случайной величины ζ ; $D\zeta$ – дисперсия случайной величины ζ .
Таким образом, если полагаем

$$\varepsilon = \sqrt{\frac{D\zeta}{0,01}},$$

то с вероятностью 0,99 значение случайной величины ζ находится в интервале

$$\left[E\zeta - \sqrt{\frac{D\zeta}{0,01}}, E\zeta + \sqrt{\frac{D\zeta}{0,01}} \right].$$

Осталось отметить, что случайная величина с трапецевидным распределением имеет математическое ожидание (см. [1]). Тогда мы можем подсчитать несмещенные оценки математического ожидания

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

и дисперсии

$$\bar{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

и подставить их в выражение для доверительного интервала. Обозначим границы этого доверительного интервала:

$$\bar{a}_{\min}^1 = \bar{x} - \sqrt{\frac{\bar{\sigma}^2}{0,01}}, \quad \bar{a}_{\max}^1 = \bar{x} + \sqrt{\frac{\bar{\sigma}^2}{0,01}}.$$

Далее обратимся к результатам, изложенным в [6].

Определение [6, р. 23]. Случайная переменная ζ со случайным математическим ожиданием $E\zeta$ является субгауссовской, если существует положительное число σ , такое, что $E[e^{\lambda(\zeta - E\zeta)}] \leq e^{\sigma^2 \lambda^2 / 2}$ для всех $\lambda \in R$.

Из примера 2.4 [6, р. 24] и упражнения 2.4 [6, р. 51] следует, что ограниченная случайная величина является субгауссовской с параметром $\sigma = (b - a)/2$, где a – нижняя граница этой случайной величины, b – верхняя граница этой случайной величины. Так как трапецевидная случайная величина ограничена, то она является субгауссовской с данным параметром.

Далее, известно концентрационное неравенство для субгауссовских случайных величин (см. неравенство (2.9) [6, р. 23]): $P(|\zeta - E\zeta| \geq \varepsilon) \leq 2e^{-\frac{\varepsilon^2}{\sigma^2}}$ для всех $\varepsilon \in R$.

Исходя из этого неравенства и неравенства $2e^{-\frac{\varepsilon^2}{\sigma^2}} \leq 0,01$, можем положить, что $\varepsilon = 3,25\sigma$. Этим гарантируем, что значение случайной величины ζ находится в доверительном интервале $[E\zeta - \varepsilon, E\zeta + \varepsilon]$ с вероятностью 0,99.

Далее воспользуемся оценкой $\bar{x}$ для математического ожидания и оценкой $(\max_{1 \leq i \leq n} x_i - \min_{1 \leq i \leq n} x_i) / 2$ для параметра σ . При этом получим границы доверительного интервала

$$\bar{a}_{\min}^2 = \bar{x} - 3,25 \frac{(\max_{1 \leq i \leq n} x_i - \min_{1 \leq i \leq n} x_i)}{2}, \quad \bar{a}_{\max}^2 = \bar{x} + 3,25 \frac{(\max_{1 \leq i \leq n} x_i - \min_{1 \leq i \leq n} x_i)}{2}.$$

В качестве результирующего доверительного интервала выберем доверительный интервал с наименьшей длиной и обозначим границы этого интервала как $\bar{a}_{\min}$, $\bar{a}_{\max}$.

Приближенные алгоритмы. Сначала выберем допустимый интервал значений для трапециевидного распределения. Пусть $x_{\min} = \min_{1 \leq i \leq n} x_i$, $x_{\max} = \max_{1 \leq i \leq n} x_i$. Тогда $a_{\min} = \min\{x_{\min}, \bar{a}_{\min}\}$, $a_{\max} = \max\{x_{\max}, \bar{a}_{\max}\}$. В случае, если даны дополнительные предположения, что значения трапециевидного распределения неотрицательные и если $a_{\min}$ меньше нуля, положим $a_{\min} = 0$.

Далее разобьем интервал $[a_{\min}, a_{\max}]$ на g равных пересекающихся только по границам интервалов, и пусть $a_{\min} = b_1 < b_2 < \dots < b_{g+1} = a_{\max}$ – возрастающая последовательность границ этих интервалов. Добавим в это множество значения выборки, и пусть в этом множестве теперь содержится r элементов, т. е. $b_1 < b_2 < \dots < b_r$.

Алгоритм 1.

Шаг 1. Для всех четверок параметров $a_1 < a_2 \leq a_3 < a_4$, где возможное множество значений каждого параметра $\{b_1, b_2, \dots, b_r\}$, подсчитать с помощью (1) значение высоты трапеции t и с помощью (2) значение многомерной функции плотности.

Шаг 2. Выбрать четверку параметров с максимальным значением многомерной функции плотности.

Вычислительная сложность алгоритма – $O(r^4 n)$.

Основа алгоритма 1 – это дискретизация пространства параметров и перебор по этим значениям параметров. Алгоритм строит приближенное значение оптимальных параметров трапециевидного распределения и его точность тем выше, чем больше число интервалов r . Однако с ростом числа интервалов значительно повышается и вычислительная сложность алгоритма.

Далее предложим другой подход к определению оптимальных значений параметров. Для этого необходимо оценить максимальное возможное значение высоты трапеции t .

Положим $a_1 = x_{\min}$, $a_4 = x_{\max}$, $a_1 < a_2 = a_3 < a_4$ (для a_2 и a_3 зафиксируем любые подходящие значения). Вычислим с помощью (1) высоту получившейся трапеции t и обозначим эту высоту $t_{\max}$. Разобьем интервал $[t_{\max}/k, t_{\max}]$ на $k - 1$ интервалов, пересекающихся только по границам. Пусть $t_{\max}/k = c_1 < c_2 < \dots < c_k = t_{\max}$ – возрастающая последовательность границ этих интервалов.

Алгоритм 2.

Шаг 1. Для всех значений $t \in \{c_1, c_2, \dots, c_k\}$ выполнить шаги 2–5.

Шаг 2. Для всех $j \in \{1, 2, \dots, r-1\}$, $i < j$, подсчитать $R_1(i, j) = \prod_{x_l \in [b_i, b_j]} p(x_l)$ (если ни одно значение x_l не попадает в интервал $[b_i, b_j]$, полагаем данное произведение равным 1). Отметим, что функция плотности $p(x)$ для всех значений x_l из такого отрезка $[b_i, b_j]$, где $a_1 = b_i$, $a_2 = b_j$ и высота трапеции t фиксирована, определяется однозначно как линейная на отрезке $[b_i, b_j]$ функция, проходящая через точки $[a_1, 0]$ и $[a_2, t]$.

Шаг 3. Для всех $j \in \{1, 2, \dots, r-1\}$, $i \leq j$, подсчитать $R_2(i, j) = \prod_{x_l \in (b_i, b_j]} p(x_l)$ (если ни одно значение x_l не попадает в полуинтервал $(b_i, b_j]$, полагаем данное произведение $\prod_{x_l \in (b_i, b_j]} p(x_l)$ равным 1). Отметим, что функция плотности $p(x)$ для всех значений x_l из такого полуинтервала $(b_i, b_j]$, где $a_2 = b_i$, $a_3 = b_j$ и высота трапеции t фиксирована, определяется однозначно как t .

Шаг 4. Для всех $j \in \{1, 2, \dots, r\}$, $i < j$, подсчитать $R_3(i, j) = \prod_{x_l \in (b_i, b_j]} p(x_l)$ (если ни одно значение x_l не попадает в полуинтервал $(b_i, b_j]$, полагаем данное произведение $\prod_{x_l \in (b_i, b_j]} p(x_l)$ равным 1). Отметим, что функция плотности $p(x)$ для всех значений x_l из такого полуинтервала $(b_i, b_j]$ с фиксированной высотой t определяется однозначно как линейная на отрезке $(b_i, b_j]$ функция, проходящая через точки $[b_i, t]$ и $[b_j, 0]$.

Шаг 5. Для каждого $j_1, j_2, j_3 \in \{1, 2, \dots, r-1\}$, $j_1 < j_2 \leq j_3$, $b_{j_1} \leq x_{\min}$, положим $a_1 = b_{j_1}$, $a_2 = b_{j_2}$, $a_3 = b_{j_3}$. Вычислим $a_4 = \frac{2 - t(a_3 - a_1 - a_2)}{t}$ (это соотношение получено из равен-

ства (1)) и в случае, если $a_4 \geq x_{\max}$ и $a_4 > a_3$, положим $R(j_1, j_2, j_3) = R_1(a_1, a_2)R_2(a_2, a_3)R_3(a_3, a_4)$, иначе $R(j_1, j_2, j_3) = 0$.

Подсчитаем $\max_{j_1, j_2, j_3 \in \{1, 2, \dots, r-1\}, j_1 < j_2 \leq j_3} R(j_1, j_2, j_3)$ и параметры a_1, a_2, a_3, a_4 , на которых достигается этот максимум.

Шаг 6. Выбрать высоту трапеции t , для которой значения выбранных параметров приводят к максимальному значению многомерной функции плотности распределения.

Вычислительная сложность алгоритма – $O(\max\{r^3 k, r^2 n\})$.

Алгоритм 2 основан на дискретизации высоты трапеции и пространства значений параметров и последовательном вычислении рекуррентных соотношений для каждого возможного значения высоты трапеции:

- на шаге 2 вычисляется вклад в многомерную функцию плотности (2) значений, находящихся на отрезке $[a_1, a_2]$;
- на шаге 3 – вклад в многомерную функцию плотности (2) значений, находящихся на отрезке $[a_2, a_3]$;
- на шаге 4 – вклад в многомерную функцию плотности (2) значений, находящихся на отрезке $[a_3, a_4]$;
- на шаге 5 – максимум многомерной функции плотности по всем параметрам для фиксированной высоты трапеции;
- на шаге 6 – максимум многомерной функции плотности по всем возможным значениям высоты трапеции.

Заключение. Разработаны два приближенных алгоритма вычисления оценок параметров трапециевидного распределения с помощью метода максимального правдоподобия. Трапециевидное распределение может быть использовано в задачах моделирования вместо нормального распределения, когда есть нижние и верхние границы на значение случайной величины.

Благодарности. Работа выполнена при частичной поддержке Белорусского республиканского фонда фундаментальных исследований (проект № Ф26КИ-009).

Acknowledgements. The research is partially supported by the Belarusian Republican Foundation for Fundamental Research (project no. Ф26КИ-009).

Список использованных источников

1. Kacker, R. N. Trapezoidal and triangular distributions for Type B evaluation of standard uncertainty / R. N. Kacker, J. F. Lawrence // *Metrologia*. – 2007. – Vol. 44, № 2. – P. 117–127. <https://doi.org/10.1088/0026-1394/44/2/003>
2. Dorp van, J. R. Generalized trapezoidal distributions / J. R. van Dorp, S. Kotz // *Metrika*. – 2003. – Vol. 58. – P. 85–97. <https://doi.org/10.1007/s001840200230>
3. Zygmunt, L. W. Effective measurand estimators for samples of trapezoidal PDFs / L. W. Zygmunt // *Journal of Automation, Mobile Robotics and Intelligent Systems*. – 2012. – Vol. 6, № 1. – P. 8–14.
4. Боровков, А. А. Математическая статистика: учебник / А. А. Боровков. – СПб.: Лань, 2010. – 704 с.
5. Боровков, А. А. Теория вероятностей / А. А. Боровков. – М.: Эдиториал УРСС, 1999. – 472 с.
6. Wainwright, M. J. High-Dimensional Statistics: A Non-Asymptotic Viewpoint / M. J. Wainwright. – Cambridge University Press, 2019. – 571 p. <https://doi.org/10.1017/9781108627771>

References

1. Kacker R. N., Lawrence J. F. Trapezoidal and triangular distributions for Type B evaluation of standard uncertainty. *Metrologia*, 2007, vol. 44, no. 2, pp. 117–127. <https://doi.org/10.1088/0026-1394/44/2/003>
2. Dorp van J. R., Kotz S. Generalized trapezoidal distributions. *Metrika*, 2003, vol. 58, pp. 85–97. <https://doi.org/10.1007/s001840200230>
3. Zygmunt L. W. Effective measurand estimators for samples of trapezoidal PDFs. *Journal of Automation, Mobile Robotics and Intelligent Systems*, 2012, vol. 6, no. 1, pp. 8–14.
4. Borovkov A. A. *Mathematical Statistics*. Saint Petersburg, Lan' Publ., 2010. 704 p. (in Russian).
5. Borovkov A. A. *Probability Theory*. Moscow, Editorial URSS Publ., 1999. 472 p. (in Russian).
6. Wainwright M. J. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge University Press, 2019. 571 p. <https://doi.org/10.1017/9781108627771>

Информация об авторе

Баркетов Максим Сергеевич – кандидат физико-математических наук, доцент, старший научный сотрудник, Объединенный институт проблем информатики Национальной академии наук Беларуси (ул. Сурганова, 6, 220012, Минск, Республика Беларусь). E-mail: barketau@mail.ru. <https://orcid.org/0000-0001-5209-0149>

Information about the author

Maksim S. Barketau – Ph. D. (Physics and Mathematics), Associate Professor, Senior Researcher, United Institute of Informatics Problems of the National Academy of Sciences of Belarus (6, Surganov Str., 220012, Minsk, Republic of Belarus). E-mail: barketau@mail.ru. <https://orcid.org/0000-0001-5209-0149>