УДК 519.2

А. М. Волк
Белорусский государственный технологический университет

АНАЛИЗ СВОЙСТВ СТАТИСТИЧЕСКИХ ОЦЕНОК ПАРАМЕТРОВ ОБОБЩЕННОГО ГАММА-РАСПРЕДЕЛЕНИЯ

Рассмотрено обобщенное гамма-распределение. Данное распределение обобщает распределения класса гамма и имеет широкое применение в статистических методах исследования физических процессов, дистанционном зондировании, теории надежности, при описании дисперсного состава частиц дробления. Исследованы его свойства и найдены числовые характеристики. Методом наибольшего правдоподобия получены уравнения для статистической оценки параметров данного распределения. Для полученных оценок найдена матрица информации Фишера, показана ее знакоположительность, что доказывает их состоятельность, ассимптотическую эффективность и единственность.

Ключевые слова: обобщенное гамма-распределение, физические процессы, теория надежности, свойства, числовые характеристики, статистическая оценка параметров, метод наибольшего правдоподобия, матрица информации Фишера, асимптотическая эффективность, единственность.

Для цитирования: Волк А. М. Анализ свойств статистических оценок параметров обобщенного гамма-распределения // Труды БГТУ. Сер. 3, Физико-математические науки и информатика. 2023. № 1 (266). С. 10–14. DOI: 10.52065/2520-6141-2023-266-1-2.

A. M. Volk
Belarusian State Technological University

ANALYSIS OF THE PROPERTIES OF STATISTICAL ESTIMATES OF PARAMETERS OF THE GENERALIZED GAMMA DISTRIBUTION

A generalized gamma distribution is considered. This distribution generalizes Gamma class distributions and has wide application in statistical methods of investigation of physical processes, in remote sensing, in reliability theory, in description of disperse composition of crushing particles. Its properties have been investigated and numerical characteristics have been found. Equations for statistical estimation of the parameters of this distribution have been obtained by the method of greatest likelihood method. The Fisher information matrix was found for the obtained estimations, its sign-positivity was shown, which proves their consistency, asymptotically-efficiency and uniqueness.

Keywords: generalized Gamma distribution, physical processes, reliability theory, properties, numerical characteristics, statistical estimation of parameters, best likelihood method, Fisher information matrix, asymptotic efficiency, uniqueness.

For citation: Volk A. M. Analysis of properties of statistical estimates of parameters of the generalized gamma distribution. *Proceedings of BSTU, issue 3, Physics and Mathematics. Informatics*, 2023, no. 1 (266), pp. 10–14. DOI: 10.52065/2520-6141-2023-266-1-2 (In Russian).

Введение. Обобщенное гамма-распределение, имеющее функцию плотности

$$f(x)=\frac{|c|}{\theta\Gamma\left(\frac{b}{c}\right)}\left(\frac{x}{\theta}\right)^{b-1}\exp\left(-\left(\frac{x}{\theta}\right)^{c}\right), \qquad (1)$$

было рассмотрено в работе [1]. Данное распределение изучалось ранее и было переоткрыто позднее другими исследователями [2].

Гамма-распределения более полутора столетий используются при моделировании реальных процессов и явлений. Обобщенное гамма-распределение используется в теории надежности, при прогнозировании продолжительности лечения и затрат на медицинское обслуживание, в расчетах инженерных рисков и рисков катастроф (землетрясений и наводнений), при обработке изображений и дистанционном зондировании, в качестве моделей распределения доходов [3].

Данное распределение включает в себя большинство известных законов: экспоненциальное распределение; χ^2-распределение; распределение Эрланга; гамма-распределение; полунормальное распределение, или распределение максимума процесса броуновского движения; распределение Рэлея; распределение Максвелла – Больцмана; χ-распределение; *m*-распределение Накагами; распределение Вильсона – Хильферти; распределение Вейбулла – Гнеденко; обобщенное распределение Вейбулла;

псевдовейбулловское распределение; распределение Пирсона третьего и пятого типа; распределение Леви; распределение Фреше, а также их обратные и масштабированные аналоги [3, 4].

Популярность рассматриваемого распределения обусловливается как гибкостью и многообразием параметров, так и возможностью использовать его в качестве адекватных асимптотических аппроксимаций во многих предельных схемах [5].

Обобщенное гамма-распределение удобно применять при описании дисперсного состава частиц дробления [6, 7]. При этом сложной остается задача статистической оценки его параметров. Метод моментов [7] не выявляет свойств оценок и не гарантирует их единственность. Автор данной работы независимо от других авторов определился с названием распределения (1), рассмотрел его свойства и предложил статистическую оценку параметров методом максимального правдоподобия [8].

Исследованию свойств и применению обобщенного гамма-распределения посвящено большое количество работ [9]. Тем не менее и в настоящее время актуальной остается задача статистической оценки параметров обобщенного гамма-распределения и исследование их свойств.

Свойства обобщенного гамма-распределения. Рассмотрим обобщенное гамма-распределение некоторой случайной величины ξ, заданное функцией плотности (1).

Отметим, что параметр θ является параметром масштаба, а b и c есть параметры формы.

Выполним переход к безразмерной случайной величине $\eta = \xi / \theta$ и получим функцию плотности:

$$f(t) = \frac{|c|}{\theta\Gamma\left(\frac{b}{c}\right)} t^{b-1} \exp\left(-t^{c}\right). \quad (2)$$

Функция распределения непрерывной случайной величины η

$$F(t) = \frac{|c|}{\Gamma\left(\frac{b}{c}\right)} \int_0^t \tau^{b-1} \exp\left(-\tau^{c}\right) d\tau \quad (3)$$

сводится к неполной гамма-функции [10].

Если $c > 0$, то

$$F(t) = \frac{1}{\Gamma\left(\frac{b}{c}\right)} \gamma\left(\frac{b}{c}, t^{c}\right), \quad (4)$$

а при $c < 0$

$$F(t) = 1 - \frac{1}{\Gamma\left(\frac{b}{c}\right)} \gamma\left(\frac{b}{c}, t^{c}\right). \quad (5)$$

Если $\frac{b-1}{c} > 0$, то исследуемые распределения имеют моду:

$$\eta_{\text{mod}} = \left(\frac{b-1}{c}\right)^{\frac{1}{c}}, \quad \xi_{\text{mod}} = \theta\, \eta_{\text{mod}} = \theta\left(\frac{b-1}{c}\right)^{\frac{1}{c}}. \quad (6)$$

Для функции распределения мода будет точкой перегиба. На рис. 1, 2 приведены графики функции плотности и функции распределения случайной величины η для различных значений параметров b и c.

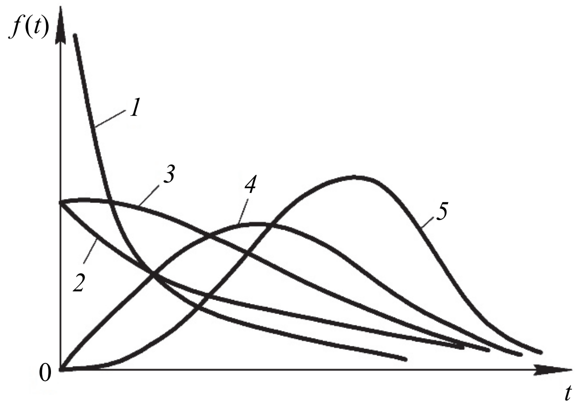

Рис. 1. Функция плотности распределения (2) случайной величины η:
1 – $0 < b < 1$, $c > 0$ или $b < 0$, $c < 0$;
2 – $b = 1$, $0 < c \le 1$; *3* – $b = 1$, $c > 1$;
4 – $1 < b \le 2$, $c > 0$; *5* – $b > 2$, $c > 0$

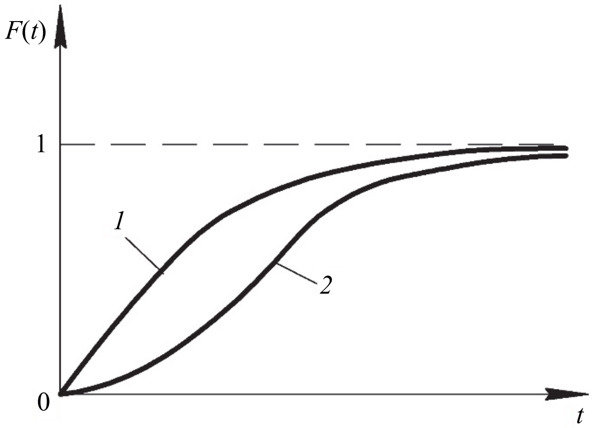

Рис. 2. Функция распределения случайной величины η:
1 – $0 < b \le 1$, $c > 0$ или $b < 0$, $c < 0$; *2* – $b > 1$, $c > 0$

Для распределения (1) при $c > 0$ найдены начальные моменты порядка ν, удовлетворяющего условию $b + \nu > 0$, причем

$$\alpha_{\nu}(\eta) = \Gamma\left(\frac{b+\nu}{c}\right) / \Gamma\left(\frac{b}{c}\right); \quad (7)$$

$$\alpha_{\nu}(\xi) = \theta^{\nu}\Gamma\left(\frac{b+\nu}{c}\right) / \Gamma\left(\frac{b}{c}\right). \quad (8)$$

Действительно,

$$\alpha_\nu(\xi) = \frac{c}{\theta\Gamma\left(\frac{b}{c}\right)} \int_0^{+\infty} x^\nu \left(\frac{x}{\theta}\right)^{b-1} \exp\left\{-\left(\frac{x}{\theta}\right)^c\right\} dx =$$

$$= \frac{c\theta^\nu}{\theta\Gamma\left(\frac{p}{c}\right)} \int_0^{+\infty} \left(\frac{x}{\theta}\right)^{b+\nu-1} \exp\left\{-\left(\frac{x}{\theta}\right)^c\right\} dx =$$

$$= \frac{c\theta^\nu}{\Gamma\left(\frac{b}{c}\right)} \int_0^{+\infty} t^{b+\nu-1} \exp(-t^c) dt = \theta^\nu \frac{\Gamma\left(\frac{b+\nu}{c}\right)}{\Gamma\left(\frac{b}{c}\right)}.$$

Статистическая оценка параметров. Статистическая оценка параметров распределения может быть выполнена методом моментов, при котором параметры распределения находятся из условия равенства теоретических и статистических моментов. Метод моментов требует решения достаточно сложной системы уравнений, а также проверки соответствия полученной функции статистическим данным с помощью критериев согласия и не гарантирует единственности решения.

Выполним статистическую оценку параметров распределения (1) методом наибольшего правдоподобия [11].

Пусть дана выборка $X = \; = (x_1, x_2, ..., x_n)$ генеральной совокупности случайной величины ξ, имеющей функцию плотности распределения (1). Рассмотрим функцию правдоподобия:

$$L = \prod_{i=1}^{n} \frac{|c|}{\theta\Gamma\left(\frac{b}{c}\right)} \left(\frac{x_i}{\theta}\right)^{b-1} \exp\left\{-\left(\frac{x_i}{\theta}\right)^c\right\}. \quad (9)$$

Прологарифмируем данную функцию, получим функцию

$$L_n = LnL = \sum_{i=1}^{n} \left[\ln\frac{|c|}{\theta} - \ln\Gamma\left(\frac{b}{c}\right) + (b-1)\ln\frac{x_i}{\theta} - \left(\frac{x_i}{\theta}\right)^c\right] =$$

$$= n\left[\ln\frac{|c|}{\theta} - \ln\Gamma\left(\frac{b}{c}\right) + (b-1)\frac{1}{n}\sum_{i=1}^{n}\ln\frac{x_i}{\theta} - \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i}{\theta}\right)^c\right]$$

и найдем ее частные производные:

$$\frac{\partial L_n}{\partial \theta} = -\frac{nb}{\theta} + \frac{c}{\theta}\sum_{i=1}^{n}\left(\frac{x_i}{\theta}\right)^c; \quad (10)$$

$$\frac{\partial L_n}{\partial b} = -\frac{n}{c}\psi\left(\frac{b}{c}\right) + \sum_{i=1}^{n}\ln\frac{x_i}{\theta}; \quad (11)$$

$$\frac{\partial L_n}{\partial c} = \frac{n}{c} + \frac{nb}{c^2}\psi\left(\frac{b}{c}\right) - \sum_{i=1}^{n}\left(\frac{x_i}{\theta}\right)^c \ln\frac{x_i}{\theta}. \quad (12)$$

Применим необходимое условие экстремума функции многих переменных, приравняем найденные частные производные к нулю и получим уравнения правдоподобия для определения статистических оценок параметров распределения:

$$b - \frac{c}{n}\sum_{i=1}^{n}\left(\frac{x_i}{\theta}\right)^c = 0; \quad (13)$$

$$\psi\left(\frac{b}{c}\right) - \frac{c}{n}\sum_{i=1}^{n}\ln\frac{x_i}{\theta} = 0; \quad (14)$$

$$\frac{1}{c} + \frac{b}{c^2}\psi\left(\frac{b}{c}\right) - \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i}{\theta}\right)^c \ln\frac{x_i}{\theta} = 0. \quad (15)$$

Решение уравнений (13)–(15) дает статистическую оценку параметров исследуемого распределения (1).

Оценка наибольшего правдоподобия для параметра масштаба θ выражается в явном виде при известных b и c:

$$\theta = \left(\frac{c}{nb}\sum_{i=1}^{n} x_i^{\,c}\right)^{\frac{1}{c}}. \quad (16)$$

Условия существования и свойства статистических оценок определяются выполнением условий регулярности [12], основным из которых является знакоположительность матрицы информации Фишера для параметров b и c:

$$I_n(b,c) = -E\begin{bmatrix} \frac{\partial^2 L_n}{\partial b^2} & \frac{\partial^2 L_n}{\partial b \partial c} \\ \frac{\partial^2 L_n}{\partial c \partial b} & \frac{\partial^2 L_n}{\partial c^2} \end{bmatrix}. \quad (17)$$

Вычислим производные второго порядка прологарифмированной функции правдоподобия L_n, обозначим $b / c = k$ и получим:

$$\frac{\partial^2 L_n}{\partial b^2} = -\frac{n}{c^2}\psi'\left(\frac{b}{c}\right) = -\frac{n}{c^2}\psi'(k);$$

$$\frac{\partial^2 F}{\partial b \partial c} = \frac{n}{c^2}\psi\left(\frac{b}{c}\right) + \frac{nb}{c^3}\psi'\left(\frac{b}{c}\right) = \frac{n}{c^2}[\psi(k) + k\psi'(k)];$$

$$\frac{\partial^2 F}{\partial c \partial b} = \frac{\partial^2 F}{\partial b \partial c} = \frac{n}{c^2}[\psi(k) + k\psi'(k)];$$

$$\frac{\partial^2 F}{\partial c^2} = -\frac{n}{c^2} - \frac{2nb}{c^3}\psi\left(\frac{b}{c}\right) - \frac{nb^2}{c^4}\psi'\left(\frac{b}{c}\right) - \sum_{i=1}^{n}\left(\frac{x_i}{\theta}\right)^c \ln^2\frac{x_i}{\theta} =$$

$$= -\frac{n}{c^2} - \frac{2nk}{c^2}\psi(k) - \frac{nk^2}{c^2}\psi'(k) - \sum_{i=1}^{n}\left(\frac{x_i}{\theta}\right)^c \ln^2\frac{x_i}{\theta}.$$

Математическое ожидание величин, не зависящих от переменных x_i, равно этим величинам. Оценки максимального правдоподобия удовлетворяют условиям регулярности. При этих условиях элементы матрицы информации Фишера равны для независимых случайных величин.

$$E\left(\sum_{i=1}^{n}\left(\frac{x_i}{\theta}\right)^c \ln^2\frac{x_i}{\theta}\right) = nE\left(\left(\frac{x}{\theta}\right)^c \ln^2\frac{x}{\theta}\right) =$$

$$= \frac{nc}{\Gamma\left(\frac{b}{c}\right)}\int_0^{+\infty}\left(\frac{x}{\theta}\right)^c \ln^2\left(\frac{x}{\theta}\right)\left(\frac{x}{\theta}\right)^{p-1}\exp\left\{-\left(\frac{x}{\theta}\right)^c\right\}d\left(\frac{x}{\theta}\right) =$$

$$= \frac{nc}{\Gamma(k)}\int_0^{+\infty} t^{c+b-1}\ln^2 t\exp\left\{-t^c\right\}dt = \left|t^c = z\right| =$$

$$= \frac{n}{\Gamma(k)}\int_0^{+\infty} z^k \ln^2 z\exp(-z)dt = \frac{n}{\Gamma(k)}\frac{\partial^2\Gamma(k+1)}{\partial b^2} =$$

$$= \frac{n\Gamma''(k+1)}{c^2\Gamma(k)} = \frac{nk\Gamma''(k+1)}{c^2\Gamma(k+1)} =$$

$$= \frac{nk}{c^2}[\psi'(k+1) + \psi^2(k+1)].$$

$$I_n(k,c) = \frac{n}{c^2}\begin{bmatrix}\psi'(k) & -[\psi(k)+k\psi'(k)] \\ -[\psi(k)+k\psi'(k)] & D\end{bmatrix},$$

где $D = 1 + 2k\psi(k) + k^2\psi'(k) + k[\psi'(k+1) + \psi^2(k+1)]$.

Рассмотрим матрицу

$$I(k) = \begin{bmatrix}\psi'(k) & -[\psi(k)+k\psi'(k)] \\ -[\psi(k)+k\psi'(k)] & D\end{bmatrix}. \quad (18)$$

Логарифмическая производная $\psi(k)$ гамма-функции монотонно возрастает, непрерывна на интервале $(0, +\infty)$ и принимает значения в пределах $(-\infty, +\infty)$ [10]. Поэтому $\psi'(k)$ принимает положительные значения на интервале $(0; +\infty)$.

Определитель матрицы (18) в общем виде будет

$$\det I(k) = \psi'(k)D - [\psi'(k+1) + \psi^2(k+1)]^2. \quad (19)$$

График значений определителя матрицы (19), полученный численными методами (рис. 3), показывает, что он принимает положительные значения при $k > 0$.

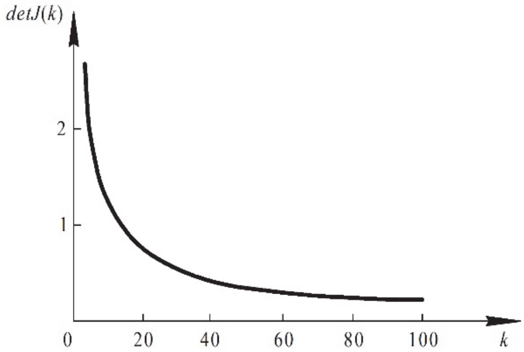

Рис. 3. График значений определителя матрицы (18)

Матрица (18) будет знакоположительной.

Матрица информации Фишера $I_n(k,c)$ отличается от матрицы (18) на положительный множитель и также будет знакоположительной.

Решение уравнений (13)–(15) дает статистическую оценку параметров распределения (1), для которых будут выполнятся условия регулярности.

Данные оценки являются состоятельными, асимптотически-несмещенными, эффективными, асимптотически-нормальными и ассимптотически-эффективными [12]. При условии эффективности оценок система (13)–(15) имеет единственное решение [11].

Заключение. Исследованное обобщенное гамма-распределение имеет широкую область применения в силу своей универсальности. Но его использование ограничивалось отсутствием способов достоверной оценки параметров на основании статистических данных.

Предложенный метод наибольшего правдоподобия, полученные уравнения (13)–(15), знакоположительность матрицы информации Фишера позволяют получить состоятельные, ассимптотически-эффективные статистические оценки параметров распределения, что доказывают их единственность.

Список литературы

1. Stacy E. W. A generalization of the gamma distribution // Ann. Math. Statistics. 1962. Vol. 33. P. 1187–1192.

2. Кудрявцев А. А. О представлении гамма-экспоненциального и обобщенного отрицательного биномиального распределений // Информатика и ее применения. 2019. Т. 13, вып. 4. С. 76–80.

3. Королев В. Ю., Крылов В. А., Кузьмин В. Ю. Устойчивость конечных смесей обобщенных гамма-распределений относительно возмущений параметров // Информатика и ее применения. 2011. Т. 5, вып. 1. С. 31–38.

4. Кудрявцев А. А. Априорное обобщенное гамма-распределение в байесовских моделях баланса // Информатика и ее применения. 2019. Т. 13, вып. 3. С. 27–33.

5. Закс Л. М., Королев В. Ю. Обобщенные дисперсионные гамма-распределения как предельные для случайных сумм // Информатика и её применения. 2013. Т. 7, вып. 1. С. 105–115.

6. Коузов П. А. Основы анализа дисперсионного состава промышленных пылей и измельченных материалов. Л.: Химия, 1987. 264 с.

7. Левданский Э. И., Волк А. М., Плехов И. М. О законе распределения частиц при дроблении // ТОХТ. 1986. № 5. С. 672–677.

8. Волк А. М. Обобщенное гамма-распределенние // Актуальные проблемы информатики: сб. тр. VI Междунар. науч. конф., 26–30 окт. 1998 г. В 3 ч. Ч. 2. Минск: БГУ, 1998. С. 426–432.

9. Джонсон Н. Л., Коц С., Балакришнан Н. Одномерные непрерывные распределения. В 2 ч. Ч. 1. М.: БИНОМ. Лаборатория знаний, 2010. 703 с.

10. Янке Е., Эмдэ Ф., Леш Ф. Специальные функции: формулы, графики, таблицы. М.: Наука, 1977. 458 с.

11. Крамер Г. Математические методы статистики: Основы моделирования и первичная обработка данных. М.: Мир, 1975. 648 с.

12. Леман Э. Теория точечного оценивания. М.: Наука. Гл. ред. физ.-мат. лит., 1991. 448 с.

References

1. Stacy E. W. A generalization of the gamma distribution. *Ann. Math. Statistics*, 1962, vol. 33, pp. 1187–1192.

2. Kudryavtsev A. A. On the representation of the gamma exponential and generalized negative binomial distributions. *Informatika i yeye primeneniya* [Computer science and its applications], 2019, vol. 13, issue 4, pp. 76–80 (In Russian).

3. Korolev V. Yu., Krylov V. A., Kuzmin V. Yu. Stability of finite mixtures of generalized gamma distributions with respect to perturbations of parameters. *Informatika i yeye primeneniya* [Computer science and its applications], 2011, vol. 5, issue 1, pp. 31–38 (In Russian).

4. Kudryavtsev A. A. A priori generalized gamma distribution in Bayesian balance models. *Informatika i yeye primeneniya* [Computer science and its applications], 2019, vol. 13, issue 3, pp. 27–33 (In Russian).

5. Zaks L. M., Korolev V. Yu. Generalized dispersion gamma distributions as the limit for random sums. *Informatika i yeye primeneniya* [Computer science and its applications], 2013, vol. 7, issue 1, pp. 105–115 (In Russian).

6. Kouzov P. A. *Osnovy analiza dispersionnogo sostava promyshlennykh pyley i izmel'chennykh materialov* [Principles of analysis of variance and composition of industrial dust from grinding materials]. Leningrad, Khimiya Publ., 1987. 264 p. (In Russian).

7. Levdanskiy E. I., Volk A. M., Plekhov I. M. On the particle distribution law in crushing. *Tekhnicheskiye osnovy khimicheskoy tekhnologii* [Technical Foundations of Chemical Engineering], 1986, no 5, pp. 672–677 (In Russian).

8. Volk A. M. Generalized Gamma-distribution. *Aktual'nyye problemy informatiki*: *sb. trudov VI Mezhdunoy nauch. konf., 26–30 okt. 1998. V 3 ch. Ch. 2* [Actual problems of informatics: collection of works of the VI International scientific conference 26–30 October 1998. In 3 parts. Part 2]. Minsk, BGU Publ., 1998, pp. 426–432 (In Russian).

9. Johnson N. L., Kotz S., Balakrishnan N. *Odnomernyy'e neprery'vnyy'e raspredeleniya. V 2 ch. Ch. 1* [One-dimensional continuous distributions. In 2 parts. Part 1]. Moscow, BINOM. Laboratoriya znaniy Publ., 2010. 703 p. (In Russian).

10. Yanke E., Emde F., Lesh F. *Spetsial'nyye funktsii: Formuly, grafiki, tablitsy* [Special functions: Formulas, graphs, tables]. Moscow, Nauka Publ., 1977. 458 p. (In Russian).

11. Kramer G. *Matematicheskiye metody statistiki: Osnovy modelirovaniya i pervichnaya obrabotka dannykh* [Mathematical Methods of Statistics: Basics of modeling and primary data processing]. Moscow, Mir Publ., 1975. 648 p. (In Russian).

12. Lehman E. *Teoriya tochechnogo otsenivaniya* [The theory of point estimation]. Moscow, Nauka. Gl. red. fiz.-mat. lit. Publ., 1991. 448 p. (In Russian).

Информация об авторе

Волк Анатолий Матвеевич − кандидат технических наук, доцент, доцент кафедры высшей математики. Белорусский государственный технологический университет (220006, г. Минск, ул. Свердлова, 13а, Республика Беларусь). E-mail: volk@belstu.by

Information about the author

Volk Anatoliy Matveevich − PhD (Engineering), Associate Professor, Assistant Professor, the Department of Higher Mathematics. Belarusian State Technological University (13a, Sverdlova str., 220006, Minsk, Republic of Belarus). E-mail: volk@belstu.by

Поступила после доработки 06.02.2023