УДК 519.2; 316.4 DOI: 10.36871/2618-9976.2023.07-2.002

ОПРЕДЕЛЕНИЕ АНАЛИТИЧЕСКИХ ЗАВИСИМОСТЕЙ ДЛЯ ДИСПЕРСИЙ ВЫБОРОЧНЫХ КОЭФФИЦИЕНТОВ АСИММЕТРИИ И ЭКСЦЕССА ТИПОВЫХ АППРОКСИМИРУЮЩИХ РАСПРЕДЕЛЕНИЙ

Светлана Васильевна Прокопчина[1]

[1] *Доктор технических наук, профессор, профессор кафедры "Системный анализ в экономике", Финансовый университет при Правительстве Российской Федерации, Москва, Россия, e-mail: svprokopchina@mail.ru*

ИНФОРМАЦИЯ

Ключевые слова:
плотность распределения вероятностей
коэффициент асимметрии
коэффициент эксцесса

АННОТАЦИЯ

Статья посвящена вопросам определения аналитических зависимостей для дисперсий выборочных характеристик статистик коэффициентов асимметрии и эксцесса типовых распределений. Показано, что значения параметров точных выборочных распределений коэффициентов асимметрии и эксцесса сильно коррелированы с видом основного распределения случайной величины или случайного процесса. Приводятся схемы статистических выводов аналитических решений. Приводятся графические решения. Приведены конкретные зависимости и графики для дисперсий статистик коэффициентов асимметрии и эксцесса типовых унимодальных распределений.

ДЛЯ ЦИТИРОВАНИЯ: *Прокопчина С.В.* Определение аналитических зависимостей для дисперсий выборочных коэффициентов асимметрии и эксцесса типовых аппроксимирующих распределений // Мягкие измерения и вычисления. 2023. № 7-2. Т. 68(2). С. 14–22; https://doi.org/10.36871/2618-9976.2023.07-2.002.

DETERMINATION OF ANALYTICAL DEPENDENCIES FOR VARIANCES OF SAMPLE COEFFICIENTS OF ASYMMETRY AND KURTOSIS OF TYPICAL APPROXIMATING DISTRIBUTIONS

Svetlana V. Prokopchina[1]

[1] *Doctor of Technical Sciences, Professor, Professor of the Department "System Analysis in Economics", Financial University under the Government of the Russian Federation, Moscow, Russia, e-mail: svprokopchina@mail.ru*

ARTICLE INFO

Keywords:
Probability distribution density
Asymmetry coefficient
Kurtosis coefficient

ABSTRACT

The article is devoted to the issues of determining analytical dependencies for variances of sample characteristics of statistics of coefficients of asymmetry and kurtosis of typical distributions. It is shown that the values of the parameters of the exact sample distributions of the coefficients of asymmetry and kurtosis are strongly correlated with the type of the main distribution of a random variable or a random process. Schemes of statistical conclusions of analytical solutions are given. Graphical solutions are given. Specific dependences and graphs for the variances of the statistics of the coefficients of asymmetry and kurtosis of typical unimodal distributions are given.

FOR CITATION: *Prokopchina S.V.* (2023) Determination of analytical dependencies for variances of sample coefficients of asymmetry and kurtosis of typical approximating distributions. *Soft measurements and computing,* vol. 68(2), no. 7-2, pp. 14–22 (In Russ.); https://doi.org/10.36871/2618-9976.2023.07-2.002.

Введение

Важными вопросами в задачах статистического оценивания являются вопросы точности и надежности определения статистических характеристик по экспериментальным данным. К часто используемым в практических задачах вероятностным характеристикам относятся коэффициенты асимметрии и эксцесса выборочных распределений. Являясь статистическими оценками их вероятностных аналогов, они могут быть максимально информативно представлены своими точными выборочными распределениями. Как правило, объемы выборочных данных, в условиях которых определяются эти оценки, являются весьма значительными. Как показано в [1, 2, 5] они должны быть свыше 500 элементов. В таких условиях точные выборочные распределения этих статистик аппроксимируются нормальным распределением. Однако некоторым сложным аспектом применения таких выборочных распределений является необходимость определения параметров этих нормальных законов, в частности дисперсий указанных статистик. В данной статье рассматриваются методологические основы решения данных вопросов и определяются аналитические зависимости и графические решения.

Аналитическое определение дисперсий выборочных распределений статистик коэффициентов асимметрии и эксцесса

В соответствии с теоремой об асимтотическом поведении выборочных распределений функций от моментов исходного распределения [1, 2], которыми являются и оценки коэффициентов асимметрии и эксцесса, в [1, 2] делается вывод о том, что выборочные коэффициенты асимметрии и эксцесса при большом объеме выборки распределены приближенно нормально около соответствующих характеристик совокупности, из которой была сделана выборка, с дисперсиями по следующим формулам:

$$D_A = \frac{4\mu_2^2\mu_6 - 12\mu_2\mu_3\mu_5 - 24\mu_2^3\mu_4 + 9\mu_3^2\mu_4 + 35\mu_2^2\mu_3^2 + 36\mu_2^5}{4\mu_2^5 N} + O\left(\frac{1}{N^{3/2}}\right), \qquad (1)$$

где N – объем выборочных данных;

для оценки коэффициента эксцесса ($\tilde{E}$):

$$D_E = \frac{\mu_2^2\mu_8 - 4\mu_2\mu_4\mu_6 - 8\mu_2^2\mu_3\mu_5 + 4\mu_4^3 - \mu_2^2\mu_4^2 + 16\mu_2\mu_3^2\mu_4 + 16\mu_2^3\mu_3^2}{\mu_2^6 N} + O\left[\frac{1}{N^{3/2}}\right]. \qquad (2)$$

Поскольку предлагаемый алгоритм ориентирован на работу с большими объемами выборочных данных (более 1000 элементов), то в соответствии с выше сказанным, принимаем допущение о нормальном распределении выборочных коэффициентов асимметрии и эксцесса. Математическое ожидание для оценок этих характеристик может быть вычислено с использованием таблиц, приведенных в [1, 3] где представлены формулы для вычисления теоретических значений коэффициентов асимметрии и эксцесса типовых распределений, входящих в систему кривых К.Пирсона, а дисперсии по формулам (1), (2).

Дисперсии оценок рассматриваемых характеристик могут быть представлены в виде $\frac{C}{N}$ (2'), где C – коэффициент пропорциональности, который определяется свойствами исходного распределения. Поэтому для одного и того же объема выборочных данных дисперсия оценок $\tilde{A}$ и $\tilde{E}$ для различных типовых распределений будет различной. Вследствие этого для решения задачи построения разделяющих границ и других задач, связанных с дальнейшей разработкой алгоритма аппроксимации, необходимо исследовать величину дисперсий выборочных коэффициентов асимметрии и эксцесса типовых распределений.

В [1, 2] приводятся формулы для дисперсии $\tilde{A}$ и $\tilde{E}$ нормального распределения. Дисперсии оценок $\tilde{A}$ и $\tilde{E}$ в этом случае зависят только от объема выборочных данных:

$$D_A = \frac{6(N-2)}{(N+1)(N+3)} + O\left[\frac{1}{N^3}\right], \tag{3}$$

$$D_E = \frac{24N(N-2)(N-3)}{(N+1)^2(N+3)(N+5)} + O\left[\frac{1}{N^3}\right]. \tag{4}$$

Поскольку в систему типовых распределений входят сложные распределения с различными значениями параметров формы, которые соответствуют определенным эталонным классамсистемы типовых распределений, то необходимо исследовать зависимость дисперсий оценок коэффициентов асимметрии и эксцесса от параметров формы сложных распределений. Для этого, в общем случае, по известной схеме получения центральных моментов, дифференцируют характеристическую функцию распределения в соответствии с формулой [2, 5]:

$$\alpha'_n = \frac{1}{i^n} \cdot \frac{d^n W(U)}{d^n U}\bigg| U = 0, \tag{5}$$

получая начальные моменты распределения α'_n, и переходят от них к соответствующим центральным моментам с использованием формулы [1, 2]:

$$\mu_n = \sum_{K=0}^{n} (-1)^K \binom{n}{K} \alpha'_{n-K} (\alpha'_1)^K. \tag{6}$$

Затем, подставляя полученные значения моментов в (1), (2), можно получить выражения для дисперсии выборочных характеристик асимметрии и эксцесса в зависимости от параметров формы распределений.

В случае гамма-распределения, распределения Вейбулла, логарифмически-нормального распределения известны формулы для вычисления центральных или начальных α'_n моментов этих распределений в зависимости от параметров формы. Они имеют следующий вид:

– для гамма-распределения [77]:

$$\mu_n = (n-1)(\mu_{n-1} + K\mu_{n-2}); \tag{7}$$

– для распределения Вейбулла [73]:

$$\alpha'_n = \Gamma\left(1 + \frac{\text{т}}{K}\right); \tag{8}$$

– для логарифмически-нормального распределения [2, 6]:

$$\alpha_n = \omega^{n^2}\rho^n, \tag{9}$$

$$\omega = exp\left(\frac{1}{2}K^2\right), \tag{9′}$$

$$\rho = ex\,p(-MK). \tag{9″}$$

Подставляя выражения для центральных моментов, вычисленных с использованием формул (7), (9), в общие формулы для вычислений дисперсий оценок коэффициентов асимметрии и эксцесса (1), (2), получаем окончательные выражения для дисперсий этих выборочных характеристик в зависимости от параметров формы:

– для гамма-распределения:

$$D_A = \frac{6K^2 + 36K + 30}{K^2 N}, \tag{10}$$

$$D_E = \frac{24K^3 + 1008K^2 + 4008K + 3024}{K^3 N}; \tag{11}$$

– для логарифмически-нормального распределения:

$$\begin{aligned} D_A = \Bigg\{&(\omega^{30} - 6\omega^{20} + 15\omega^{12} - 20\omega^6 + 15\omega^2 - 5)\cdot(\omega^2 - 1)^{-3} \\ &- \quad -3(\omega^2 + 2)(\omega^{20} - 5\omega^{12} + 10\omega^6 - 10\omega^2 + 4)(\omega^2 - 1)^{-2} \\ &- (\omega^8 + 2\omega^6 + 3\omega^4 - 3)\left[6 - \frac{9(\omega^2 + 2)^2(\omega^2 - 1)}{4}\right] \\ &+ \frac{35(\omega^2 - 1)(\omega^2 + 2)^2}{4} + 9\Bigg\}/N\,; \end{aligned} \tag{12}$$

$$\begin{aligned} D_E = \{[(&\omega^{56} - 8\omega^{42} + 28\omega^{30} - 56\omega^{20} + 70\omega^{12} - 56\omega^6 + 28\omega^2 - 7) \\ &- 4(\omega^8 + 2\omega^6 + 3\omega^4 - 3)(\omega^{30} - 6\omega^{20} + 15\omega^{12} - 20\omega^6 + 15\omega^2 \\ &- 5)](\omega^2 - 1)^{-4} \\ &- 8(\omega^2 + 2)(\omega^{20} - 5\omega^{12} + 10\omega^6 - 10\omega^2 + 4)(\omega^2 - 1)^{-2} \\ &+ (\omega^8 + 2\omega^6 + 3\omega^4 - 3)[4(\omega^8 + 2\omega^6 + 3\omega^4 - 3)^2 \\ &- (\omega^8 + 2\omega^6 + 3\omega^4 - 3) \\ &+ 16(\omega^2 - 1)(\omega^2 + 2)^2(\omega^8 + 2\omega^6 + 3\omega^4 - 3)] + 16(\omega^2 \\ &- 1)(\omega^2 + 2)^2\}/N. \end{aligned} \tag{13}$$

Для бета-распределения известна его характеристическая функция [4]:

$$W_x(U) = 1 + \frac{K1}{(K1 + K2)}iU + \frac{1}{2!}\frac{K1(K1 + 1)}{(K1 + K2)(K1 + K2 + 1)}iU^2 + \cdots \tag{14}$$

В соответствии с ней начальные моменты могут быть определены по следующей формуле:

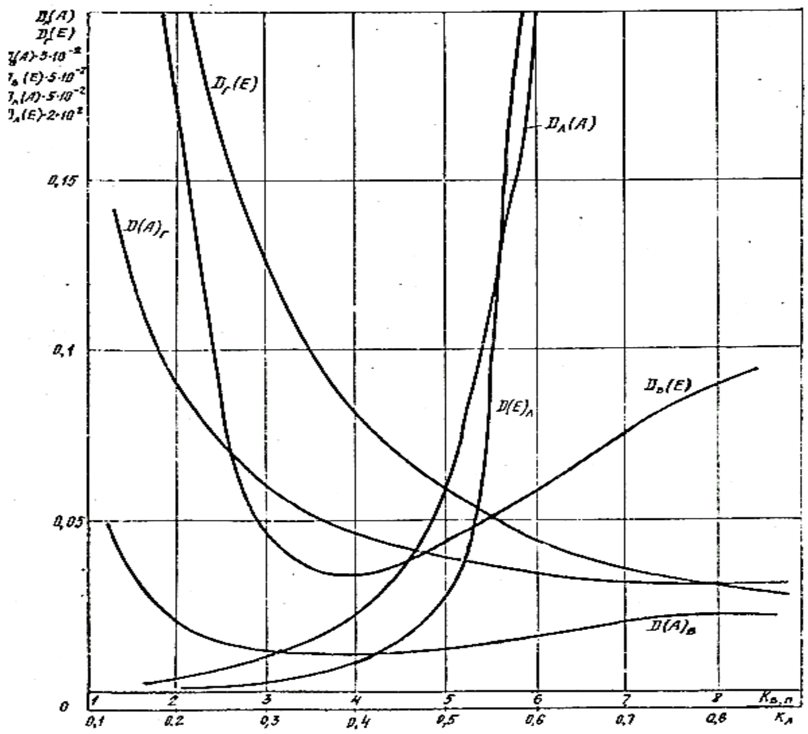
$$\alpha_n' = \frac{(K1 + n - 1)!\,(K1 + K2 - 1)!}{(K1 - 1)!\,(K1 + K2 + n - 1)!}. \tag{15}$$

Центральные моменты могут быть определены в зависимости от параметров формы распределения:

$$\mu_n = \frac{(n-1)K2}{K1} \sum_{K=0}^{n-2} \frac{(-1)^K \alpha'_{n-K-1} (\alpha'_1)^K C_{n-2}^K}{K1 + n - K - 1}. \quad (16)$$

Для определения дисперсий оценок $\tilde{A}$ и $\tilde{E}$ сложных типовых распределений составлена программа.

По результатам решения с использованием этой программы при объеме выборки $N = 7000$ и диапазонах изменения параметров формы: для гамма-распределения от 1,5 до 9; для распределения Вейбулла от 1,5 до 6; для логарифмически-нормального распределения от 0,1 до 0,6; для бета-распределения от $K1 = 1{,}5 = K2$ до $K1 = K2 = 9$ построены графики рис 1, 2, 3.



Г – гамма распределение;
В – распределение Вейбулла;
Л – логарифмически нормальное распределение

Рис. 1. Зависимость дисперсии оценок коэффициентов асимметрии *D (A)* и эксцесса *D (E)* сложных распределений от параметров формы

В связи с тем, что простые распределения являются частными случаями сложных с определенными значениями параметров формы, то с учетом формулы (2), и с использование указанной программы можно определить дисперсии оценок $\tilde{A}$ и $\tilde{E}$ этих простых распределений. Для простых распределений они равны:

– для равномерного распределения:

$$D_A = \frac{2{,}24}{N}, \quad (17)$$

$$D_E = \frac{2{,}8}{N}; \quad (18)$$

– для распределения Симпсона:

$$D_A = \frac{2,75}{N}, \tag{19}$$

$$D_E = \frac{4,41}{N}\ ; \tag{20}$$

– для распределения Рэлея:

$$D_A = \frac{7,2}{N}, \tag{21}$$

$$D_E = \frac{84}{N}\ ; \tag{22}$$

– для экспоненциального распределения:

$$D_A = \frac{72}{N}, \tag{23}$$

$$D_E = \frac{8048}{N}\,. \tag{24}$$

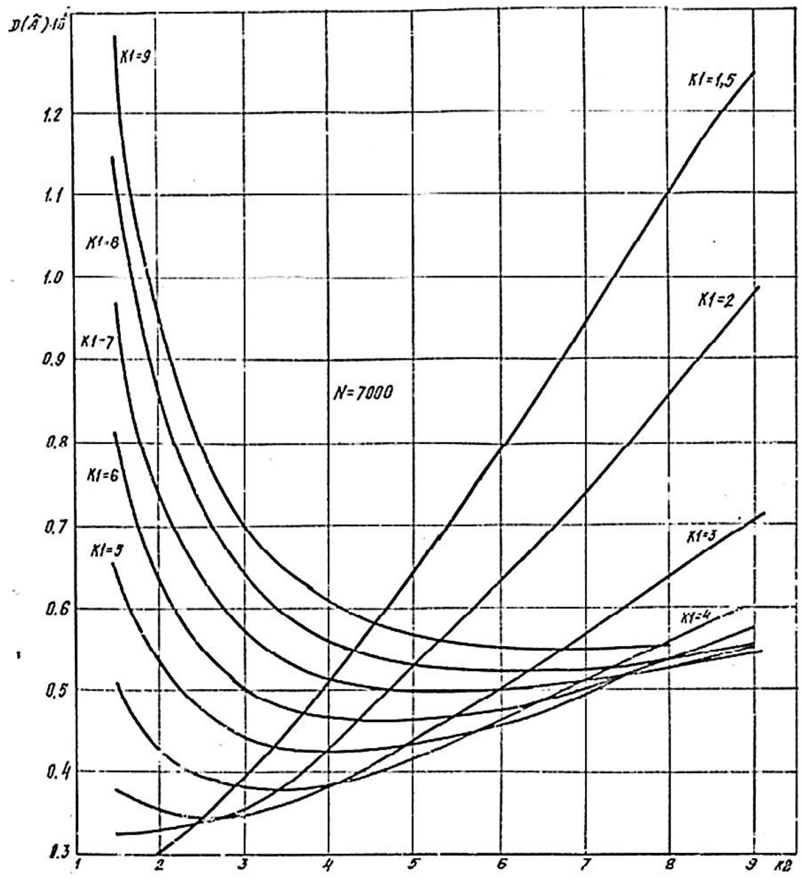


Рис. 2. Зависимость дисперсии оценки коэффициента асимметрии бета-распределения от параметров К1, К2

Из анализа выражений (17)– (24) видно, что при постоянном объеме выборочных данных N, максимальными по величине среди всех простых распределений будут дисперсии оценок $\tilde{A}$ и $\tilde{E}$ экспоненциального распределения.

По графикам рис. 3 можно сделать общий вывод, что с увеличением степени асимметричности и островершинности кривой распределения дисперсия оценок коэффициентов асимметрии и эксцесса увеличивается. Причем, дисперсия оценок коэффициента эксцесса при тех же значениях параметров формы K каждого сложного распределения и при том же объеме выборки примерно на порядок больше, чем дисперсия коэффициента асимметрии. С увеличением параметра формы гамма-распределения дисперсия оценки коэффициента асимметрии изменяется от 0,0102 при $K = 1$, что соответствует экспоненциальному распределению (до 0,00148) при $K = 9$; дисперсия оценки коэффициента эксцесса от 1,152 (при $K = 1$) до 0,0270 (при $K = 9$). Сравнивая значения дисперсий оценок $\tilde{A}$ и $\tilde{E}$ при $K = 9$ для гамма-распределения с соответствующими значениями для нормального распределения, полученными с использованием формул (3), (4) при том же объеме выборки $N = 7000$, и учитывая, что при этом значении параметра формы гамма-распределения близко к нормальному, видим, что дисперсии оценок, соответствующих характеристик близки ($D_{A\text{норм.}} = 0{,}0008$; $D_{E\text{норм.}} = 0{,}00342$). Однако, в силу большей симметрии нормального распределения и его меньшему E, D_A и D_E этого распределения остаются меньшими, чем для гамма-распределения с $K = 9$. С уменьшением параметра формы (с приближением кривой распределения к нормальной) дисперсия оценок $\tilde{A}$ и $\tilde{E}$ логарифмически-нормального распределения уменьшается и стремится к значениям D_A и D_E нормального распределения. Так при $K = 0{,}1$ $D_A = 0{,}00142, D_E = 0{,}0075$. При увеличении параметра формы от $K = 1$ до $K = 4$ для распределения Вейбулла дисперсия оценок $\widetilde{A}$ и $\tilde{E}$ уменьшается, при дальнейшем увеличении параметра формы, в соответствии с удалением формы кривой от нормального вида, дисперсия оценок $\widetilde{A}$ и $\tilde{E}$ растет. Для бета-распределения в частном случае равномерного распределения с $K1 = K2 = 1$ дисперсия оценок $\widetilde{A}$ и $\tilde{E}$ меньше, чем дисперсия этих оценок для нормального распределения, что соответствует тому факту, что при равной степени симметричности кривых этих распределений коэффициент эксцесса меньше для равномерного распределения ($D_{A\text{р.}} = 0{,}00020, D_{E\text{р.}} = 0{,}00049$). В целом для бета-распределения сохраняется общая тенденция в увеличении дисперсии оценок $\widetilde{A}$ и $\tilde{E}$ с увеличением степени асимметричности и островершинности кривой по сравнению с нормальным распределением. Экспериментально величины дисперсий выборочных коэффициентов асимметрии и эксцесса при конечных объемах выборки были исследованы методом статистического моделирования В качестве исходного распределения было взято экспоненциальное распределение, так как оно обладает одним из самых больших значений A и E. Для моделирования использовался датчик псевдослучайных чисел, работающий по алгоритму, предложенному в [6]:

$$x_i = \frac{ln(R_i)}{2}, \tag{25}$$

где x_i – значение СВ, имеющий экспоненциальный закон распределения;

R_i – значение равномерно распределенной СВ.

Графические решения иллюстрируют достаточно хорошую близость теоретических и экспериментальных значений дисперсий оценок коэффициентов асимметрии и эксцесса уже при $N \geqslant 750$. Различие между теоретическими и экспериментальными значениями можно объяснить конечностью объема выборки M.

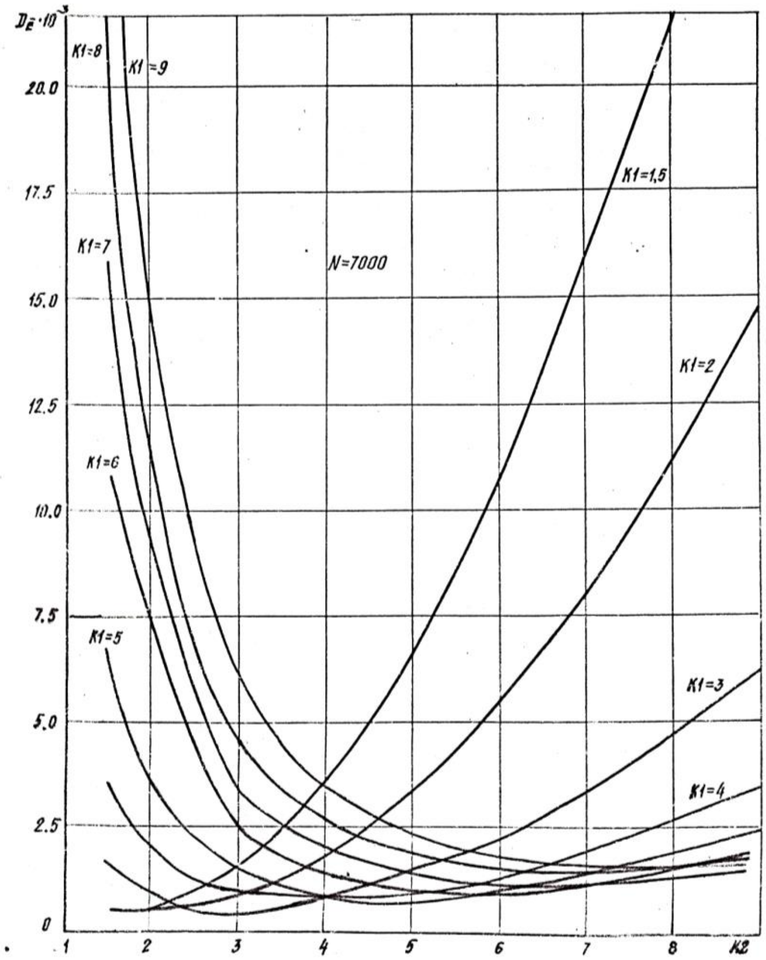


Рис. 3. Зависимость дисперсии оценки коэффициента эксцесса бета-распределения от параметров формы $K1$, $K2$

Учитывая (2'), для определения дисперсий оценок $\widetilde{A}$ и $\tilde{E}$ типовых распределений можно использовать графики рис. 1–3 при любом объеме выборочных данных $N \geqslant 750$, пользуясь пересчетными формулами:

– для дисперсии оценки коэффициента асимметрии:

$$D_A = \frac{D_{A\Gamma} \cdot 7000}{N}; \tag{26}$$

– для дисперсии оценки коэффициента эксцесса:

$$D_E = \frac{D_{E\Gamma} \cdot 7000}{N}, \tag{27}$$

где $D_{A\Gamma}, D_{E\Gamma}$ – значения дисперсий $\widetilde{A}$ и $\tilde{E}$ типового распределения снятые с графиков, построенных для $N\ = 7000$.

Заключение

Результаты исследования величин дисперсий оценок коэффициентов асимметрии и эксцесса аппроксимирующих распределений, дают возможность определения точных выборочных распределений коэффициентов асимметрии и эксцесса при решении многих статистических задач.

Список источников

[7] *Кендалл М.Дж., Стьюарт А.* Теория распределений. М.: Наука, 1966. 587 с.
[8] *Кендалл М.Дж., Стьюарт А.* Статистические выводы и связи. М.: Наука, 1974. 899 с.
[9] *Прокопчина С.В.* Байесовские интеллектуальные технологии в задачах моделирования законов распределения в условиях неопределенности. М.: Издательский дом "НАУЧНАЯ БИБЛИОТЕКА", 2020. 292 с.
[10] *Корн Г.А.* Справочник по математике для научных работников и инженеров. М.: Наука, 1974. 831 с.
[11] *Крамер Г.* Математические методы статистики. М.: Мир. 1975. 647 с.
[12] *Бендат Дж., Пирсон А.* Измерение и анализ случайных процессов. М.: Мир, 1974. 463 с.

Referances

[7] *Kendall M.J., Stewart A.* Theory of distributions. M.: Nauka, 1966. 587 p.
[8] *Kendall M.J., Stewart A.* Statistical conclusions and connections. M.: Nauka, 1974. 899 p.
[9] *Prokopchina S.V.* Bayesian intelligent technologies in problems of modeling distribution laws under uncertainty. M.: Publishing House "SCIENTIFIC LIBRARY", 2020. 292 p.
[10] *Korn G.A.* Handbook of Mathematics for scientists and engineers. M.: Nauka, 1974. 831 p.
[11] *Kramer G.* Mathematical methods of statistics. M.: Mir, 1975. 647 p.
[12] *Bendat J., Pearson A.* Measurement and analysis of random processes. M.: Mir, 1974, 463 p.

Статья поступила в редакцию 14.07.2023; одобрена после рецензирования 21.07.2023; принята к публикации 24.07.2023.

The article was submitted 14.07.2023; approved after reviewing 21.07.2023; accepted for publication 24.07.2023.