УДК 519.248, 519.87
DOI: 10.33184/bulletin-bsu-2020.4.3

СРАВНЕНИЕ МЕДИАНЫ И СРЕДНЕГО ОТНОШЕНИЙ СЛУЧАЙНЫХ ВЕЛИЧИН, ОПРЕДЕЛЯЕМЫХ НОРМАЛЬНЫМИ И РАВНОМЕРНЫМИ РАСПРЕДЕЛЕНИЯМИ

© Цициашвили Г. Ш.[1,2], Осипова М. А.[1,2] *

[1]*Дальневосточный федеральный университет*
Россия, 690091 г. Владивосток, ул. Суханова, 8.

[2]*Институт прикладной математики ДВО РАН*
Россия, 690041 г. Владивосток, ул. Радио, 7.

Тел.: +7 (423) 231 23 75.
Email: mao1975@list.ru

В биоценологии нередко возникает задача обработки измерений продольных и поперечных размеров зерен различных растений и установления пропорции между ними. Эта задача в чисто статистическом плане приводит к тому, что эмпирическое среднее отношений продольных и поперечных размеров становится достаточно неудобной статистикой. Поэтому требуется определить условия, когда следует, и когда не следует пользоваться эмпирическим средним случайных отношений.

Для этой цели оказалось удобным рассмотреть модели отношения случайных величин с логарифмически нормальными распределениями и с логарифмически равномерными распределениями. Функции распределения этих случайных отношений характеризуются математическим ожиданиями и медианами.

В настоящей работе установлено, что эти характеристики близки или к единице, или к бесконечности в зависимости от некоторого коэффициента, определяющего разброс распределений числителей и знаменателей случайной дроби.

Ключевые слова: *случайная дробь, математическое ожидание, медиана, логнормальное распределение, логарифмически равномерное распределение.*

Введение

В биоценологии [1] возникает задача исследования отношения двух случайных величин, имеющих логарифмически нормальное или логарифмически равномерное распределение. В содержательном плане эта задача связана с обработкой измерений продольных и поперечных размеров зерен различных растений и установления пропорции между ними.

Для решения данной задачи автор работы [1] применил к достаточно обширному эмпирическому материалу отношения средних значений продольных размеров к средним значениям поперечных размеров зерен и к другим ботаническим наблюдениям [2]. Это потребовало формального сравнения отношения средних и среднего отношений, поскольку применение той или иной статистики является важным для приложений. Поэтому автор работы [1] обратился к нам с предложением методами теории вероятностей провести сравнение данных статистик и выяснить причины их расхождений.

В настоящей работе аналитически и численно определяются условия, когда математические ожидания и медианы распределений случайных отношений существенно различаются, и когда они близки друг к другу. Это позволяет определить условия, при которых можно или нельзя пользоваться эмпирическими средними случайных отношений.

Отношение двух случайных величин с логнормальными распределениями

Рассмотрим случайную дробь $\nu = \eta_1/\eta_2$, полагая, что $\eta_1 = \exp(c_1+\xi_1)$, $\eta_2 = \exp(c_2+\xi_2)$, где случайные величины ξ_1, ξ_2 удовлетворяют соотношениям

$$\xi_1 = a_{11}\tau_1 + a_{12}\tau_2,\ \xi_2 = a_{21}\tau_1 + a_{22}\tau_2. \qquad (1)$$

Здесь τ_1, τ_2 – независимые случайные величины со стандартным нормальным распределением, a_{ij} – вещественные числа, $i, j = 1,2$. Тогда отношение случайных величин

$$\nu = \exp(c + a_1\tau_1 + a_2\tau_2), \qquad (2)$$

$$a_1 = a_{11} - a_{21},\ a_2 = a_{12} - a_{22},\ c = c_1 - c_2,$$

причем случайная величина $\psi = c + a_1\tau_1 + a_2\tau_2$ имеет нормальное распределение со средним $E\psi = c$ и дисперсией $D\psi = d = a_1^2 + a_2^2$. Следовательно, случайная величина $\nu = \exp(\psi)$ имеет логнормальное распределение [3, гл. 6, § 6.1.6] со средним и дисперсией

$$E\nu = \exp\left(c + \frac{d}{2}\right),\ D\nu = (e^d - 1)\exp(2c + d). \qquad (3)$$

Определим теперь медиану распределения случайной величины ν:

$$\frac{1}{2} = P(\nu < T) = \int_0^T \frac{p_\psi(\ln t)dt}{t} = (u = \ln t) = \int_{-\infty}^{\ln T} p_\psi(u)du =$$

$$= \int_{-\infty}^{\ln T} \frac{\exp(-(u-c)^2/2d)}{\sqrt{2\pi d}} du = \left(v = \frac{u-c}{\sqrt{d}} \right) = \int_{-\infty}^{(\ln T - c)/\sqrt{d}} \frac{\exp(-v^2/2)}{\sqrt{2\pi}} dv.$$

Из этой формулы следует, что $\frac{\ln T - c}{\sqrt{d}} = 0$ и, значит, медиана $M\nu$ случайной величины ν определяется равенством

$$M\nu = e^c. \qquad (4)$$

Тогда справедливы следующие предельные соотношения

$$\frac{E\nu}{M\nu} \to 1,\ d \to 0,\ \frac{E\nu}{M\nu} \to \infty,\ d \to \infty. \qquad (5)$$

Отношение двух случайных величин, определяемых равномерными распределениями

Рассмотрим вероятностную модель (*1*), (*2*) в предположении, что независимые случайные величины τ_1, τ_2 имеют равномерное распределение на отрезке $[-1,1]$. Как и в предыдущем пункте остановимся на вычислении среднего и медианы случайной величины ν. Для этого сначала вычислим плотность $f(t)$ случайной величины $a_1\tau_1 + a_2\tau_2$. Без ограничения общности рассмотрим случай, когда $a_1 > 0,\ a_2 > 0,$ поскольку случайные величины τ_1, τ_2 имеют симметрические распределения.

Случайные величины $a_1\tau_1, a_2\tau_2$ независимы и равномерно распределены на отрезках $[-a_1, a_1], [-a_2, a_2]$. Обозначим плотности их распределений

$$p_i(u) = \frac{1}{2a_i},\ u \in [-a_i, a_i],\ \ p_i(u) = 0,\ u \notin [-a_i, a_i],\ i = 1.2.$$

Построим в плоскости u_1Ou_2 параллелограмм (см., *рис.* 1), ограниченный прямыми $u_1 = \pm a_1,\ u_1 + u_2 = \pm a_2.$ Обозначим $a(t)$ длину отрезка прямой $u_2 = t,$ содержащегося в этом параллелограмме. Тогда плотность $f(t)$ случайной величины $a_1\tau_1 + a_2\tau_2$ удовлетворяет равенству

$$f(t) = \frac{a(t)}{4a_1a_2},\ -a_1 - a_2 \le t \le a_1 + a_2. \qquad (6)$$

Построенный параллелограмм удовлетворяет свойством симметрии относительно точки (0,0), поэтому выполняется равенство $a(t) = a(-t),\ 0 \le t \le a_1 + a_2,$ и, значит, достаточно вычислить $a(t)$ при $0 \le t \le a_1 + a_2$.

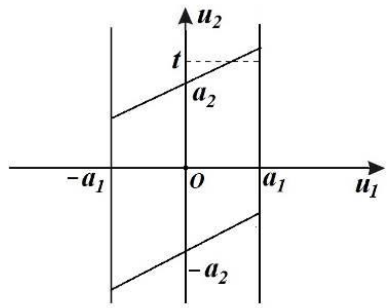

Рис. 1. Параллелограмм в плоскости u_1Ou_2.

Если $a_2 \ge a_1$, то

$a(t) = a_2 + a_1 - t,\ a_2 - a_1 \le t \le a_2 + a_1;$

$a(t) = 2a_1,\ 0 \le t \le a_2 - a_1.$

Если $a_2 \le a_1$, то

$a(t) = a_1 + a_2 - t,\ a_1 - a_2 \le t \le a_1 + a_2;$

$a(t) = 2a_2,\ 0 \le t \le a_1 - a_2.$

Объединяя последние формулы, получаем

$$a(t) = a_1 + a_2 - t, |a_1 - a_2| \le t \le a_1 + a_2;$$

$$a(t) = 2\min(a_1, a_2),\ 0 \le t \le |a_1 - a_2|. \qquad (7)$$

Используя формулы (*6*), (*7*), вычислим плотность распределения случайной величины $\ln\nu = \psi$: $p_\psi(t) = \frac{a(t-c)}{4a_1a_2}$. Следовательно, плотность $p_\nu(t)$ распределения случайной величины ν вычисляется по формуле [4, гл. II, § 8]

$$p_\nu(t) = p_\psi(\ln t)(\ln t)' = \frac{a(\ln t - c)}{4a_1a_2t},\ t \ge 0. \qquad (8)$$

Определим теперь медиану распределения случайной величины ν:

$$\frac{1}{2} = \int_0^T \frac{a(\ln t - c)}{4a_1a_2t} dt = \int_{-\infty}^{\ln T - c} \frac{a(v)}{4a_1a_2} dv. \qquad (9)$$

Из формулы (*9*) вытекает равенство $\ln T - c = 0$ и, значит, $M\nu = e^c$. Далее вычислим среднее $E\nu$, используя формулу (*8*):

$$E\nu = \int_0^\infty t\frac{p_\nu(t)}{t}dt = \int_0^\infty t\frac{a(\ln t - c)}{4a_1a_2t}dt = (w = \ln t - c) =$$

$$= \frac{e^c}{4a_1a_2}\int_{-\infty}^{\infty} e^w a(w)dw = \frac{e^c}{4a_1a_2}\int_0^{a_1+a_2}(e^w + e^{-w})a(w)dw =$$

$$= \frac{e^c\left(e^{a_1+a_2} + e^{-(a_1+a_2)} - e^{|a_1-a_2|} - e^{-|a_1-a_2|}\right)}{4a_1a_2}.$$

Следовательно, справедливо соотношение

$$\frac{E\nu}{M\nu} = \frac{1}{4a_1a_2}\left(e^{a_1+a_2} + e^{-(a_1+a_2)} - e^{|a_1-a_2|} - e^{-|a_1-a_2|}\right). \quad (10)$$

Сделаем в формуле *(10)* замену

$$a_1 \to ka_1,\ a_2 \to ka_2,\ \frac{E\nu}{M\nu} \to \left(\frac{E\nu}{M\nu}\right)_k,$$

тогда

$$\left(\frac{E\nu}{M\nu}\right)_k = \frac{1}{4k^2 a_1 a_2}\left(e^{k(a_1+a_2)} + e^{-k(a_1+a_2)} - e^{k|a_1-a_2|} - e^{-k|a_1-a_2|}\right).$$

Нетрудно получить следующие предельные соотношения

$$\lim_{k\to 0}\left(\frac{E\nu}{M\nu}\right)_k = 1,\ \lim_{k\to\infty}\left(\frac{E\nu}{M\nu}\right)_k = \infty.$$

Заключение

Полученные в работе результаты позволяют сказать, что случайная величина ν, как в модели 1, определяемой нормальными распределениями, так и в модели 2, определяемой равномерными распределениями, обладает следующим свойством. Ее средние и медианы сильно отличаются друг от друга при большом разбросе распределения и мало отличаются при малом разбросе. Это определяет качество оценки эмпирического среднего для выборки из генеральной совокупности случайных дробей, задаваемых логарифмически нормальными и логарифмически равномерными распределениями. Данный вывод проиллюстрируем в случае $a_1 = 3,\ \ a_2 = 2,\ \ c = 0.$

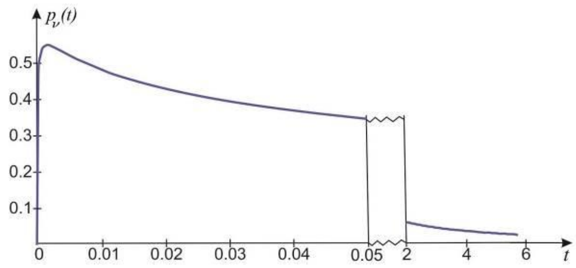

Рис. 2. Модель 1.

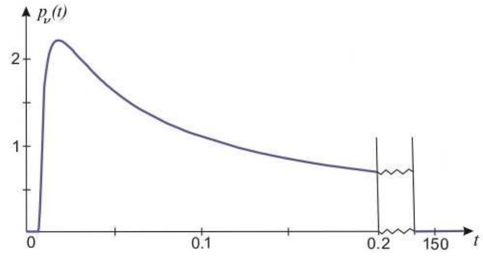

Рис. 3. Модель 2.

Расчеты по модели 1 показали, что медиана $M\nu = 1$, а среднее $E\nu \approx 665.2$. Из *рис.* 2 видно, что среднее $E\nu$ далеко от отрезка [0.2], в котором сосредоточено практически все распределение случайной величины ν. Из формулы (*5*), следует, что среднее $E\nu$ близко к медиане $M\nu$ при малой дисперсии d.

Расчеты по модели 2 показали, что среднее $E\nu \approx 6.05$, а медиана $M\nu = 1$. Аналогично из формулы (*11*) следует, что среднее $E\nu$ близко к медиане $M\nu$ при малых a_1, a_2.

ЛИТЕРАТУРА

1. Semkin B. I. The elementary theory of similarity and its application in biology and geography. Multi-site measures of similarity and dissimilarity // Pattern Recognition and Image Analysis. 2015. №25(1). P. 1–9.
2. Семкин Б. И., Гусарова И. С., Горшков М. В. Об инвариантности средних отношений величин (на примере некоторых морфологических признаков слоевищ ламинарии японской (Laminaria japonica Aresch.) из сублиторали северного Приморья) // Изв. ТИНРО. 2012. Т. 171. С. 313–320.
3. Айвазян С. А., Енюков И. С., Мешалкин Л. Д. Основы моделирования и первичная обработка данных. М.: Финансы и статистика, 1983. С. 471.
4. Ширяев А. Н. Вероятность. Учеб. пособ. для вузов. М.: Наука, 1989. С. 580.

Поступила в редакцию 23.03.2020 г.
После доработки – 06.12.2020 г.

DOI: 10.33184/bulletin-bsu-2020.4.3

COMPARISON OF MEDIAN AND MEAN RATIOS OF VALUES OF RANDOM VARIABLES DETERMINED BY NORMAL AND UNIFORM DISTRIBUTIONS

© G. Sh. Tsitsiashvili[1,2], M. A. Osipova[1,2]*

[1]Far Eastern Federal University
8 Sukhanov Street, 690091 Vladivostok, Russia.

[2]Institute of Applied Mathematics, FEB RAS
7 Radio Street, 690041 Vladivostok, Russia.

Phone: +7 (423) 231 23 75.
**Email: mao1975@list.ru*

In biocenology, the problem of processing measurements of longitudinal and transverse grain sizes of various plants and establishing proportion between them arises often. This problem in purely statistical terms leads to the fact that the empirical mean of the ratio of longitudinal and transverse dimensions becomes quite inconvenient statistics. Therefore, it is necessary to determine conditions under which one should and should not use the empirical mean of random relations. For this purpose, it is convenient to consider models of the relation of random variables with logarithmically normal distributions and with logarithmically uniform distributions. The distribution functions of these random relations are characterizing by mathematical expectations and medians. It was established that either a ratio of these characteristics is close to unity or to infinity depending on some coefficient that determines the spread of distributions of numerators and denominators of a random fraction. Consider a random fraction $\nu = \exp(c_1+\xi_1)/\exp(c_2+\xi_2)$, where $\xi_1 = a_{11}\tau_1 + a_{12}\tau_2$, $\xi_2 = a_{21}\tau_1 + a_{22}\tau_2$, τ_1, τ_2 are independent random variables with a standard normal distribution, a_{ij} are real numbers, i,j = 1,2. The random variable $\nu = \exp(c + a_1\tau_1 + a_2\tau_2)$, $a_i = a_{1i} - a_{2i}$, $c = c_1 - c_2$, has a lognormal distribution. Let us denote the median as $M\nu$ and the mean of a random variable ν as $E\nu$, $d = a_1^2 + a_2^2$. Then the following limit relations are true

$$\frac{E\nu}{M\nu} \to 1,\ d \to 0,\ \frac{E\nu}{M\nu} \to \infty,\ d \to \infty.$$

Analogously, it is possible to analize the ratio of random variables with logorithmically uniform distributions.

Keywords: random fraction, mathematical expectation, median, logarithmically normal distribution, logarithmically uniform distribution.

Published in Russian. Do not hesitate to contact us at bulletin_bsu@mail.ru if you need translation of the article.

REFERENCES

1. Semkin B. I. Pattern Recognition and Image Analysis. 2015. No. 25(1). Pp. 1–9.
2. Semkin B. I., Gusarova I. S., Gorshkov M. V. Izv. TINRO. 2012. Vol. 171. Pp. 313–320.
3. Aivazyan S. A., Enyukov I. S., Meshalkin L. D. Osnovy modelirovaniya i pervichnaya obrabotka dannykh [Basics of modeling and primary data processing]. Moscow: Finansy i statistika, 1983. Pp. 471.
4. Shiryaev A. N. Veroyatnost'. Ucheb. posob. dlya vuzov [Probability. Textbook for universities]. Moscow: Nauka, 1989. Pp. 580.

Received 23.03.2020.
Revised 06.12.2020.