Математические методы исследования
Mathematical methods of investigation

DOI: https://doi.org/10.26896/1028-6861-2025-91-1-79-88

ОЦЕНИВАНИЕ ПАРАМЕТРОВ ГАММА-РАСПРЕДЕЛЕНИЯ

© **Александр Иванович Орлов**

Московский государственный технический университет им. Н. Э. Баумана, Россия, 105005, Москва, 2-я Бауманская ул., д. 5, стр. 1; e-mail: prof-orlov@mail.ru

Статья поступила 28 февраля 2024 г. Поступила после доработки 27 марта 2024 г. Принята к публикации 26 апреля 2024 г.

Постановки задач статистического анализа данных, имеющих гамма-распределение, относятся к классической математической статистике. Однако не все они были решены в рамках параметрической статистики. Необходимо заполнить лакуны (как, например, и в случае бета-распределения), поскольку в настоящее время гамма-распределение широко используется в теоретических и прикладных работах. Примером является ГОСТ 11.011–83 «Прикладная статистика. Правила определения оценок и доверительных границ для параметров гамма-распределения». Стандартное гамма-распределение определяется параметром формы. При переходе к масштабно-сдвиговому семейству добавляются параметры масштаба и сдвига. Рассмотрены семь постановок задач оценивания параметров, поскольку каждый из трех параметров может быть как неизвестным, так и известным. Для каждой из постановок найдены оценки метода моментов и их асимптотические дисперсии. При известном параметре сдвига получены оценки максимального правдоподобия. Одношаговые оценки, асимптотически эквивалентные оценкам максимального правдоподобия, используем при неизвестном параметре сдвига. Наличие погрешностей измерения отражается на точности оценок параметров при применении тех или иных алгоритмов расчетов. В ГОСТ 11.011–83 на основе модели интервальных данных даны правила выбора метода оценивания при неизвестных параметрах формы и масштаба и известном параметре сдвига. При разработке ГОСТ 11.011–83 выявлены проблемы, для решения которых предложены новые с научной точки зрения методы. Новые научные результаты, полученные в ходе решения практической задачи (разработки ГОСТ 11.011–83), привели к созданию новых научных направлений. Речь идет о статистике интервальных данных, а также об одношаговых оценках. К настоящему времени статистика интервальных данных как раздел математической статистики достаточно развита и охватывает все основные области статистических методов. Она является важной составной частью системной нечеткой интервальной математики.

Ключевые слова: статистические методы; гамма-распределение; оценивание параметров; метод моментов; метод максимального правдоподобия; одношаговые оценки; статистика интервальных данных; асимптотические распределения; доверительные интервалы.

ESTIMATION OF GAMMA DISTRIBUTION PARAMETERS

© **Alexander I. Orlov**

Bauman Moscow State Technical University, 5, 2-ya Baumanskaya ul., Moscow, 105005, Russia; e-mail: prof-orlov@mail.ru

Received February 28, 2024. Revised March 27, 2024. Accepted April 26, 2024.

Statements of problems of statistical analysis of data with a gamma distribution are related to classical mathematical statistics. Oddly enough, not all alone were solved within the framework of parametric statistics, which was at the forefront of the development of statistical science in the first third of the 20th century. As with the beta distribution, gaps need to be filled. This is necessary because the gamma distribution is currently widely used in theoretical and applied work. An example is GOST 11.011–83 “Applied statistics. Rules for determining estimates and confidence limits for gamma distribution parameters”. The standard gamma distribution is determined by the shape parameter. When switching to a scale-shift family, scale and translation parameters are added. Seven formulations of parameter estimation problems are considered, since each of the three parameters can be either unknown or known. For each of the formulations, the estimates of the method of moments and their asymptotic variances are found. For a known shift parameter, maximum likelihood estimates are obtained. One-step estimates, asymptotically equivalent to maximum likelihood estimates, are used for an unknown shift parameter. The presence of measurement errors affects the accuracy of parameter estimates when applying certain calculation algorithms. In GOST 11.011–83, based on the interval data model, rules are given for choosing an estimation method for unknown shape and scale parameters and a known shift parameter. During the development of GOST 11.011–83, problems were identified, for the solution of which new methods from a scientific point of view were proposed. Further development of new scientific results obtained in the course of solving a practical problem (development of GOST 11.011–83) led to the creation of new scientific directions. We are talking about the statistics of interval data, as well as one-step estimates. To date, the statistics of interval data as a branch of mathematical statistics is quite developed and covers all the main areas of statistical methods. It is an important part of systemic fuzzy interval mathematics.

Keywords: statistical methods, gamma distribution, estimation of parameters, method of moments, maximum likelihood method, one-step estimates, statistics of interval data, asymptotic distributions, confidence intervals.

Введение

Математические методы исследования опираются на научную дисциплину «Теория вероятностей и математическая статистика». При рассмотрении непрерывных распределений вероятностей обычно упоминают семейство гамма-распределений [1 – 3]. Известные методы оценивания параметров вероятностных распределений могут быть применены к этому семейству. Так, в серии государственных стандартов «Прикладная статистика» нами был разработан ГОСТ 11.011–83, посвященный алгоритмам получения точечных оценок и доверительных границ для параметров семейства и подсемейств гамма-распределений. При подготовке этого нормативно-технического документа проведен ряд научно-исследовательских работ, позволивших получить достаточно продвинутые алгоритмы расчетов в рассматриваемой области. Однако указанный стандарт был отменен в 1987 г. вместе со всей серией «Прикладная статистика» [4]). После этого момента текст данного стандарта можно было бы рассматривать лишь как научную публикацию, однако этому мешал первоначальный статус официального нормативно-технического документа. Из библиотек стандартов он был исключен (утилизирован) и в научный обиход не попал. По нашему мнению, заслуживают внимания научные результаты, на основе которых он был разработан. В дальнейшем эти результаты были обобщены и получили широкое развитие. Им посвящена настоящая статья, в которой впервые систематически рассмотрена проблема оценивания параметров гамма-распределений.

Постановки задач статистического анализа данных, имеющих гамма-распределение, относятся к классической математической статистике. Однако не все они решены в рамках параметрической статистики, находившейся на переднем крае развития статистической науки в первой трети XX в. Как и в случае бета-распределения [5], необходимо заполнить лакуны. Это необходимо, поскольку гамма-распределение часто используется в теоретических и прикладных работах. Приведем примеры.

Гамма-распределения широко применяются в различных областях науки и практики, в частности, при оценке надежности (например, в модели «нагрузка – прочность» [6]) и в теории испытаний, различных областях техники и технологии (в том числе при моделировании точности технологических процессов [7]), метеорологии и т.д. [8]. В частности, установлено, что с помощью гамма-распределений могут быть смоделированы распределения общего срока службы изделия, длины цепочек токопроводящих пылинок, время достижения изделием предельного состояния при коррозии [9], время наработки до k-го отказа [10]. В ряде случаев продолжительность жизни больных хроническими заболеваниями, время достижения определения эффекта при лечении и другие используемые в медико-биологических исследованиях случайные величины имеют гамма-распределения. В экономико-математических моделях управления запасами [11] гамма-распределение может быть наиболее адекватным и для моделирования длин путей следования пассажиров маршрутным транспортом [12]. В настоящее время гамма-распределение широко используется в теоретических и прикладных работах (см., например, [13 – 16]).

Семейство гамма-распределений

Плотность стандартного гамма-распределения такова:

$$f(x;a) = \begin{cases} \dfrac{1}{\Gamma(a)} x^{a-1} \exp(-x), \ x \geq 0, \\ 0, \ x < 0, \end{cases} \quad (1)$$

где

$$\Gamma(a) = \int_0^{+\infty} x^{a-1} \exp(-x) \mathrm{d}x \ — \quad (2)$$

гамма-функция.

Стандартное гамма-распределение определяется одним параметром $a > 0$, который называется параметром формы.

Для случайной величины X, имеющей стандартное гамма-распределение, математическое ожидание и дисперсия равны a. Она имеет третий центральный момент $M(X - M(X))^3 = 2a$, асимметрию $2/\sqrt{a}$ и эксцесс $6/a$ [1 – 3].

Стандартное гамма-распределение порождает масштабно-сдвиговое семейство распределений случайных величин

$$Y = bX + c, \qquad (3)$$

где b — параметр масштаба; c — параметр сдвига. Как следует из формулы (1), заданная формулой (3) случайная величина Y имеет плотность

$$f(x; a, b, c) = \begin{cases} \dfrac{1}{\Gamma(a)}(x-c)^{a-1}b^{-a}\exp\left(-\dfrac{x-c}{b}\right), \\ \qquad x \ge c, \\ 0, \ \ x < c. \end{cases} \qquad (4)$$

Согласно (1) и (4) $f(x; a) = f(x; a, 1, 0)$. Тогда

$$M(Y) = ab + c, \ D(Y) = ab^2,$$

$$M(Y - M(Y))^3 = 2ab^3. \qquad (5)$$

Обоснованием возможности применения гамма-распределения в ряде прикладных областей может быть справедливое для него свойство воспроизводимости [1 – 3]. А именно, сумма k независимых экспоненциально распределенных случайных величин с одним и тем же параметром λ (без сдвига) имеет гамма-распределение с параметрами формы $a = k$, масштаба $b = 1/\lambda$ и сдвига $c = 0$. Поэтому гамма-распределение часто применяют в тех же прикладных областях, в которых полезно экспоненциальное распределение.

Пусть $Y_1, Y_2, \ldots, Y_n$ — выборка независимых одинаково распределенных случайных величин, имеющих гамма-распределение (4). В данной статье рассмотрим задачу оценивания параметров этого распределения.

Гамма-распределение (4) имеет три параметра — формы, масштаба и сдвига. Каждый из них может быть как известным, так и неизвестным, всего $2^3 = 8$ вариантов. Один из них не имеет отношения к рассматриваемым задачам, поскольку все параметры известны. Для обеспечения полноты исследования необходимо рассмотреть $2^3 - 1 = 7$ постановок задач оценивания (табл. 1).

Оценки метода моментов

Для оценивания параметров распределения наиболее простым является метод моментов [17]. Он состоит в том, что теоретические моменты выражаются через параметры распределения (см., например, (5)), потом решается обратная задача — находится функциональная зависимость параметров распределения от теоретических моментов, затем для получения статистических оценок параметров в эту функциональную зависимость подставляют выборочные моменты вместо теоретических. При росте объема выборок статистические оценки являются асимптотически нормальными и асимптотически несмещенными, а для получения их асимптотических дисперсий применяют метод линеаризации [17, разд. 4.4]. Для выборок из бета-распределений эта исследовательская программа реализована в [5].

Неоднозначность метода моментов определяется тем, что для оценивания определенного параметра могут быть использованы различные выборочные моменты. Например, в варианте 1 (см. табл. 1) можно исходить из любого из трех соотношений, приведенных в (5). Получаем три оценки метода моментов:

$$a_1 = a_1(\omega) = \frac{\bar{Y} - c}{b},$$

$$a_2 = a_2(\omega) = \frac{s^2}{b^2}, \ a_2 = a_2(\omega) = \frac{m_3}{2b^3}, \qquad (6)$$

Таблица 1. Варианты постановок задач оценивания

Table 1. Options for setting estimation tasks

Номер варианта	Параметр формы	Параметр масштаба	Параметр сдвига
1	Оценивается	Известен	Известен
2	Известен	Оценивается	Известен
3	Известен	Известен	Оценивается
4	Известен	Оценивается	Оценивается
5	Оценивается	Известен	Оценивается
6	Оценивается	Оценивается	Известен
7	Оценивается	Оценивается	Оценивается

где

$$\bar{Y} = \frac{Y_1 + Y_2 + \ldots + Y_n}{n},$$

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2, \; m_3 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^3. \qquad (7)$$

Выбор между тремя оценками (6) проводят в пользу первой из них как наиболее устойчивой к выбросам [17, разд. 7.2]. К аналогичным результатам приводит сравнение по асимптотическим дисперсиям оценок (6), найденным с помощью линеаризации, как пояснено выше.

Аналогично для варианта 2 (см. табл. 1) используем оценку

$$b^* = \frac{\bar{Y} - c}{a}, \; D(b^*) = \frac{1}{n}\frac{b^2}{a},$$

а для варианта 3 — оценку

$$c^* = \bar{Y} - ab, \; D(c^*) = \frac{1}{n}ab^2.$$

Таблица 2. Оценки метода моментов и их асимптотические дисперсии

Table 2. Estimates of the method of moments and their asymptotic variances

Варианты постановок задач	Оцениваемый параметр	Вид оценки	Асимптотическая дисперсия оценки
1	a	$\frac{\bar{Y}-c}{b}$	$\frac{a}{n}$
2	b	$\frac{\bar{Y}-c}{a}$	$\frac{1}{n}\frac{b^2}{a}$
3	c	$\bar{Y}-ab$	$\frac{1}{n}ab^2$
4	b	$\frac{s}{\sqrt{a}}$	$\frac{b}{2n}(a+3)$
4	c	$\bar{Y}-s\sqrt{a}$	$\frac{ab^2}{2n}(a+1)$
5	a	$\frac{s^2}{b^2}$	$\frac{2a}{n}(a+3)$
5	c	$\bar{Y}-\frac{s^2}{b}$	$\frac{ab^2}{n}(2a+3)$
6	a	$\frac{(\bar{Y}-c)^2}{s^2}$	$\frac{2a(a+1)}{n}$
6	b	$\frac{s^2}{\bar{Y}-c}$	$\frac{b^2}{n}\left(2+\frac{3}{a}\right)$
7	a	$\frac{4s^6}{m_3^2}$	$\frac{6a}{n}(a^2+6a+5)$
7	b	$\frac{m_3}{2s^2}$	$\frac{b^2}{2an}(6a^2+25a+24)$
7	c	$\bar{Y}-\frac{2s^4}{m_3}$	$\frac{ab^2}{n}(3a^2+13a+10)$

В вариантах 4 – 6 неизвестны два параметра, поэтому необходимо использовать два выборочных момента. Будем применять выборочное среднее арифметическое и выборочную дисперсию. Исходим из равенств

$$\bar{Y} = ab + c, \; s^2 = ab^2. \qquad (8)$$

В варианте 4 на основе выборочных характеристик и параметра формы оцениваются параметры масштаба и сдвига:

$$b^* = \frac{s}{\sqrt{a}}, \; c^* = \bar{Y} - ab^* = \bar{Y} - s\sqrt{a}.$$

В варианте 5 на основе выборочных характеристик и параметра масштаба определяются оценки параметров формы и сдвига:

$$a^* = s^2/b^2, \; c^* = \bar{Y} - a^*b = \bar{Y} - s^2/b.$$

В варианте 6 на основе выборочных характеристик и параметра сдвига находятся оценки параметров формы и масштаба:

$$\bar{Y} - c = ab, \; s^2 = ab^2;$$

$$b^* = \frac{s^2}{\bar{Y} - c}, \; a^* = \frac{\bar{Y} - c}{b^*} = \frac{(\bar{Y} - c)^2}{s^2}.$$

В варианте 7 неизвестны все три параметра, поэтому необходимо использовать три выборочных момента. Будем применять выборочные среднее арифметическое, дисперсию и третий центральный момент:

$$\bar{Y} = ab + c, \; s^2 = ab^2, \; m_3 = 2ab^3,$$

откуда

$$\frac{m_3}{2s^2} = b^*, \; a^* = \frac{s^2}{(b^*)^2} = \frac{4s^6}{m_3^2}, \; c^* = \bar{Y} - a^* b^* = \bar{Y} - \frac{2s^4}{m_3}.$$

Итоги оценивания параметров гамма-распределения методом моментов для всех семи вариантов (см. табл. 1) приведены в табл. 2. Асимптотические дисперсии этих оценок получены нами методом линеаризации, который подробно рассмотрен в [5 и 17, разд. 4.4].

Асимптотические доверительные границы для параметров гамма-распределения строятся на основе асимптотической нормальности оценок этих параметров. Они имеют вид

$$(\theta_n - C(\gamma)\sqrt{D^*(\theta_n)}; \theta_n + C(\gamma)\sqrt{D^*(\theta_n)}),$$

где θ — рассматриваемый параметр; θ_n — его оценка (см. табл. 2); $D^*(\theta_n)$ — оценка дисперсии θ_n, полученная путем замены в формуле для асимптотической дисперсии (крайний правый столбец табл. 2) неизвестных значений параметров их оценками; $C(\gamma)$ — коэффициент, соответствующий доверительной вероятности γ (для $\gamma = 0{,}95$ имеем $C(\gamma) = C(0{,}95) = 1{,}96$).

Здесь и далее доверительные интервалы строятся на основе асимптотической нормальности оценок параметров. Для некоторых выборок $Y_1, Y_2, ..., Y_n$ рассчитанные по асимптотическим формулам доверительные границы могут выйти за пределы областей определения соответствующих параметров. Тогда необходимо внести коррективы. Если нижняя доверительная граница для параметра формы или параметра масштаба оказывается отрицательной, то ее следует заменить на 0. Если верхняя доверительная граница для параметра сдвига оказывается больше какого-либо из элементов выборки, то ее следует заменить на $\min(Y_1, Y_2, ..., Y_n)$. Вероятности таких событий стремятся к нулю при росте объема выборки.

Метод моментов может быть использован для проверки согласия с семейством гамма-распределений. Так, для проверки согласия с подсемейством гамма-распределений $f(x; a, b, 0)$ с параметром сдвига $c = 0$ можно использовать критерий со статистикой

$$Z = \sqrt{n}\left(\frac{\bar{Y}m_3}{2s^4} - 1\right).$$

С помощью описанного выше метода нахождения асимптотического распределения функции от выборочных моментов получаем, что при росте объема выборки распределение Z сходится к нормальному, причем при справедливости гипотезы о том, что параметр сдвига равен нулю, предельное распределение Z имеет нулевое математическое ожидание и дисперсию

$$\frac{1}{2a}(3a^2 + 13a + 10).$$

Поскольку параметр формы a неизвестен, в выражение для дисперсии следует подставить его состоятельную оценку a^*. Статистический критерий для проверки нулевой гипотезы $c = 0$ при альтернативе $c > 0$ имеет критическую область

$$\left\{Z{:}\, Z > u(1-\alpha)\sqrt{\frac{3(a^*)^2 + 13a^* + 10}{2a^*}}\right\},$$

где $u(1-\alpha)$ — квантиль порядка $1-\alpha$ стандартного нормального распределения с математическим ожиданием 0 и дисперсией 1. При $n \to \infty$ уровень значимости рассматриваемого критерия стремится к α. Подробнее использование метода моментов для проверки гипотезы согласия с параметрическим семейством распределений рассмотрено в [17, разд. 7.1].

Оценки максимального правдоподобия (при известном параметре сдвига)

Оценки метода моментов — наиболее простой тип оценок параметров. Однако во многих случаях они не являются наилучшими асимптотически нормальными оценками (НАН-оценками), поскольку их дисперсия больше, чем у некоторых других. Среди НАН-оценок наиболее известны оценки максимального правдоподобия. Их получают путем максимизации по параметрам произведения плотностей, в которые вместо аргументов подставлены элементы выборки. В случае гамма-распределения с плотностью (4) речь идет о максимизации функции правдоподобия

$$F = \prod_{1\le i\le n} f(Y_i; a, b, c) =$$
$$= \prod_{1\le i\le n} \frac{1}{\Gamma(a)}(Y_i - c)^{a-1} b^{-a} \exp\left(-\frac{Y_i - c}{b}\right) \qquad (9)$$

по параметрам $a > 0$, $b > 0$, $c < \min(Y_1, Y_1, ..., Y_n)$.

Проще иметь дело с логарифмической функцией правдоподобия

$$L(a, b, c) = \ln F = -\mathrm{n}\ln\Gamma(a) +$$
$$+ (a-1)\sum_{i=1}^{n}\ln(Y_i - c) - an\ln b - \frac{1}{b}\sum_{i=1}^{n} Y_i + \frac{nc}{b}. \qquad (10)$$

Для варианта постановки 1 оценка параметра a находится из уравнения

$$\frac{\partial L}{\partial a} = -n\frac{\mathrm{d}\ln\Gamma(a)}{\mathrm{d}a} + \sum_{i=1}^{n}\ln(Y_i - c) - n\ln b = 0. \qquad (11)$$

Введем в рассмотрение логарифмическую производную гамма-функции

$$\Psi(a) = \frac{\mathrm{d}\ln\Gamma(a)}{\mathrm{d}a} = \frac{1}{\Gamma(a)}\frac{\mathrm{d}\Gamma(a)}{\mathrm{d}a}$$

(эту специальную функцию называют пси-функцией [18]). Из (11) следует, что оценку параметра формы при известных параметрах масштаба и сдвига находят как решение уравнения

$$\Psi(a) = \frac{1}{n}\sum_{i=1}^{n}\ln\left(\frac{Y_i - c}{b}\right),$$

т.е. асимптотически нормальная оценка параметра формы имеет вид

$$a^* = G\left(\frac{1}{n}\sum_{i=1}^{n}\ln\left(\frac{Y_i - c}{b}\right)\right),$$

где G — функция, обратная к пси-функции. В разделе 4 ГОСТ 11.011–83 приведены правила расчета значений функции G, дисперсии оценки a^* и доверительных границ, основанные на использовании таблиц и приближенных формул. В настоящее время целесообразно использовать соответствующие компьютерные программы.

Для варианта постановки 2 оценка b^* параметра b — решение уравнения

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial b}\left(-an\ln b - \frac{1}{b}\sum_{i=1}^{n}(Y_i - c)\right) =$$

$$= -\frac{an}{b} + \frac{1}{b^2}\sum_{i=1}^{n}(Y_i - c) = 0, \tag{12}$$

т.е.

$$b^* = \frac{1}{na}\sum_{i=1}^{n}(Y_i - c). \tag{13}$$

Таблицы и приближенные (асимптотические) формулы для расчета доверительных границ для параметра b приведены в разделе 3 ГОСТ 11.011–83. На современном этапе развития прикладной статистики в этих целях применяют компьютерные программы.

Оценим параметры формы и масштаба при известном параметре сдвига (вариант постановки 6). Оценки максимального правдоподобия находят при решении системы уравнений (11) – (12). Оценку параметра масштаба дает формула (13). Подставив ее в (11), получим уравнение для параметра формы:

$$-n\frac{\mathrm{d}\ln\Gamma(a)}{\mathrm{d}a} + \sum_{i=1}^{n}\ln(Y_i - c) -$$

$$-n\ln\left(\frac{1}{n}\sum_{i=1}^{n}(Y_i - c)\right) + n\ln a = 0.$$

Преобразуем это уравнение:

$$\ln a - \Psi(a) = \ln\left(\frac{1}{n}\sum_{i=1}^{n}(Y_i - c)\right) - \frac{1}{n}\sum_{i=1}^{n}\ln(Y_i - c).$$

Следовательно, оценки параметров формы и масштаба при известном параметре сдвига имеют следующий вид:

$$a^* = H\left[\ln\left(\frac{1}{n}\sum_{i=1}^{n}(Y_i - c)\right) - \frac{1}{n}\sum_{i=1}^{n}\ln(Y_i - c)\right],$$

$$b^* = \frac{1}{na^*}\sum_{i=1}^{n}(Y_i - c), \tag{14}$$

где H — функция, обратная к функции

$$Q(a) = \ln a - \Psi(a).$$

В ГОСТ 11.011–83 (раздел 7) приведены алгоритмы расчетов оценок (14), их дисперсий и доверительных границ, основанные на соответствующих таблицах и приближенных (асимптотических) формулах, разработанных при подготовке этого стандарта. Как и в ранее разобранных вариантах постановок, полученные результаты прикладной математики в настоящее время могут быть использованы при подготовке соответствующего программного обеспечения.

С помощью метода линеаризации (см. также [19, с. 83, 98]) можно показать, что при больших n с точностью до бесконечно малых более высокого порядка

$$M(a^* - a)^2 = \frac{a}{n(aI(a) - 1)},$$

$$M(b^* - b)^2 = \frac{b^2 I(a)}{n(aI(a) - 1)},\ I(a) = \frac{\mathrm{d}\Psi(a)}{\mathrm{d}a}.$$

На основе свойств гамма-функции [19] в ГОСТ 11.011–83 установлено, что при больших a

$$M(a^* - a)^2 = \frac{a(2a - 1)}{n},\ M(b^* - b)^2 = \frac{2b^2}{n}$$

с точностью до бесконечно малых более высокого порядка. Используя табл. 2, заключаем, что средние квадраты ошибок для оценок метода моментов параметров формы a^{**} и масштаба b^{**} (при известном параметре сдвига)

$$M(a^{**} - a)^2 = \frac{2a(a + 1)}{n},$$

$$M(b^{**} - b)^2 = \frac{b^2}{n}\left(2 + \frac{3}{a}\right)$$

больше соответствующих средних квадратов ошибок для оценок максимального правдоподобия. Таким образом, для варианта постановки 6 оценки максимального правдоподобия имеют преимущество по сравнению с оценками метода моментов. Как установлено в теории классической математической статистики, именно так обстоит дело в большинстве задач оценивания параметров. Оценки максимального правдоподобия всегда являются наилучшими асимптотически нормальными оценками, а оценки метода моментов — лишь в отдельных случаях.

Одношаговые оценки (при неизвестном параметре сдвига)

В вариантах постановок 3, 4, 5, 7 параметр сдвига c неизвестен и подлежит оцениванию.

Оценку максимального правдоподобия c^* параметра c в варианте 3 можно определить из условия

$$\frac{\partial L}{\partial c} = \frac{\partial L}{\partial c}\left[(a-1)\sum_{i=1}^{n}\ln(Y_i - c) + \frac{nc}{b}\right] =$$

$$= -(a-1)\sum_{i=1}^{n}\frac{1}{Y_i - c} + \frac{n}{b} = 0. \qquad (15)$$

В варианте 4 оценки параметров масштаба и сдвига — это решения системы уравнений максимального правдоподобия (12) и (15). Подставляя выражение для оценки параметра масштаба (13) в (15), получаем оценку параметра сдвига c^* из уравнения

$$-\frac{a-1}{n}\sum_{i=1}^{n}\frac{1}{Y_i - c} + \frac{a}{\frac{1}{n}\sum_{i=1}^{n}(Y_i - c)} = 0. \qquad (16)$$

В варианте 5 оценки параметров формы и сдвига находят при решении системы уравнений максимального правдоподобия (11) и (15). Подставив решение уравнения (11) в (15), получим уравнение для оценки параметра сдвига

$$\left\{G\left(\frac{1}{n}\sum_{i=1}^{n}\ln\frac{Y_i - c}{b}\right) - 1\right\} - \frac{n}{b} = 0, \qquad (17)$$

где G — функция, обратная к пси-функции. Правила расчетов приведены в ГОСТ 11.011–83.

При трех неизвестных параметрах (вариант 7) оценки максимального правдоподобия следует находить при решении системы трех уравнений — (11), (12) и (15).

Таким образом, в ряде постановок для получения оценок максимального правдоподобия надо решать соответствующие уравнения или системы уравнений. Казалось бы, достаточно применить распространенные численные методы. Однако при этом возникает ряд проблем.

Во многих задачах оценивания параметров для нахождения оценок максимального правдоподобия отсутствуют алгоритмы на основе явных формул и приходится применять численные методы. Так обстоит дело не только для гамма-распределения, но и, например. распределения Вейбулла – Гнеденко. Тогда тем или иным итерационным методом решают систему уравнений максимального правдоподобия (см., например, [20]) или непосредственно максимизируют функцию правдоподобия (см., например, [21]).

Однако применение численных методов в отличие от алгоритмов на основе явных формул (как для нормального распределения) сталкивается с рядом проблем. Так, сходимость итерационных методов требует обоснования. В ряде примеров функция правдоподобия имеет много локальных максимумов, а потому естественные итерационные процедуры не сходятся [22]. Например, для данных по усталостным испытаниям стали уравнение максимального правдоподобия имеет 11 корней [23]. Какой из одиннадцати использовать в качестве оценки параметра? Для решения проблемы сходимости итерационных методов проводят исследования по доказательству сходимости итерационных алгоритмов нахождения оценок максимального правдоподобия для конкретных вероятностных моделей и конкретных алгоритмов (см., например, [24]).

Следует подчеркнуть, что математическое доказательство сходимости итерационного алгоритма — это еще не все. Важен вопрос об обоснованном выборе момента прекращения вычислений в связи с достижением требуемой точности. В большинстве случаев он не решен. Кроме того, точность вычислений необходимо увязывать с объемом выборки — чем он больше, тем точнее надо находить оценки параметров, в противном случае нельзя говорить о состоятельности метода оценивания. Существенно, что при увеличении объема выборки необходимо усложнять алгоритмы, в частности, увеличивать количество используемых в компьютере разрядов, переходить от одинарной точности расчетов к двойной и далее — также ради достижения состоятельности оценок.

Итак, если нет явных формул для оценок максимального правдоподобия, их нахождение натыкается на ряд проблем вычислительного характера. Специалисты по математической статистике обычно игнорируют все эти проблемы, рассуждая об оценках максимального правдоподобия в теоретическом плане. Однако прикладная статистика не может этого допустить во избежание сомнений в возможности практического использования оценок максимального правдоподобия.

В прикладной статистике, кроме оценок метода моментов и максимального правдоподобия, разработано много других видов оценок. Например, квантильные оценки основаны на идее, аналогичной методу моментов, но вместо выборочных и теоретических моментов приравниваются выборочные и теоретические квантили. Другая группа оценок базируется на идее минимизации (по параметрам распределения) того или иного расстояния (показателя различия) между эмпирическими данными и элементом параметрического семейства. Например, минимизируется евклидово или блочное расстояние между эмпирическими и теоретическими гистограммами, а точнее — векторами, составленными из высот столбиков гистограмм.

Нет необходимости абсолютизировать оценки максимального правдоподобия. Кроме них, существуют и другие виды оценок, обладающих столь же хорошими статистическими свойствами. Например, одношаговые оценки являются результатом первой итерации метода Ньютона – Рафсона при решении уравнения или системы уравнений максимального правдоподобия. Эти оценки — наилучшие асимптотически нормальные, т.е. имеют столь же хорошие асимптотические свойства, как и оценки максимального правдоподобия, причем при тех же математических условиях регулярности. Одношаговые оценки получены в виде явных формул, т.е. для них снимаются рассмотренные выше проблемы вычислительного характера. При их расчете в качестве начального приближения можно взять оценку метода моментов, а затем найти поправки на основе первой итерации метода Ньютона – Рафсона.

Для гамма-распределения в варианте постановки 3 вместо решения уравнения (15) рекомендуют использовать одношаговую оценку. В качестве начального приближения берут оценку метода моментов $\bar{Y} - ab$. Расчетные формулы приведены в ГОСТ 11.011–83, пп. 8.15 – 8.18. Для вариантов 4 и 7 правила определения одношаговых оценок даны там же, в разделе 8.

Разработка ГОСТ 11.011–83 дала стимул к дальнейшему развитию теории одношаговых оценок. В настоящее время эта теория подробно изложена в учебниках [17, разд. 6.2; 25, разд. 3.25], где рассмотрено ее применение и для оценок параметров гамма-распределения. Поэтому указанные научные результаты в настоящей работе не рассматриваем.

Статистика интервальных данных

При разработке ГОСТ 11.011–83 были заложены основы уже достаточно развитого к настоящему времени направления современной математической статистики — статистики интервальных данных. Толчком к этой теме послужил анализ данных о наработке резцов до предельного состояния (в часах). Данное направление математической статистики отражено и в дальнейших публикациях (см., например, [17, разд. 6.1]). Совершенно очевидно, что все исходные данные заданы либо натуральными числами, либо суммами натуральных чисел и 0,5 (например, 127,5 ч). Это значит, что наработка резцов до предельного состояния измерена с точностью до 0,5 часа. Очевидно, наличие погрешностей необходимо учитывать при статистической обработке данных. При разработке ГОСТ 11.011–83 было принято, что

$$Z_i = Y_i + \varepsilon_i,\ i = 1, 2, ..., n,\ |\varepsilon_i| \le \Delta, \qquad (18)$$

где n — объем выборки; Z_i — предъявленные для анализа статистические данные; Y_i — описывающие реальность независимые случайные величины, имеющие гамма-распределение; ε_i — независимые ошибки (погрешности) измерения; Δ — максимально возможная ошибка.

Погрешности измерения отражаются на точности оценок параметров при применении тех или иных алгоритмов расчетов. В ГОСТ 11.011–83 (см. раздел 5) на основе модели интервальных данных (18) даны правила выбора метода оценивания при неизвестных параметрах формы и масштаба и известном параметре сдвига.

Дальнейшее развитие статистики интервальных данных отражено во многих работах (см., например, обобщающую статью [26]). К настоящему времени это направление математической статистики достаточно развито и охватывает все основные области статистических методов. Оно является важной составной частью системной нечеткой интервальной математики [27, гл. 7] и подробно отражено в ряде наших учебников (см., например, [17, гл. 12]).

Заключение

Представлена информация об основных результатах в области разработки методов оценивания параметров гамма-распределения, основанных на предельных теоремах теории вероятностей и математической статистики. Следующий этап — построение и изучение алгоритмов оценивания для конечных объемов выборок.

Отметим, что одношаговые оценки, как и оценки максимального правдоподобия, являются асимптотически наилучшими, но при конкретных объемах выборок в отдельных случаях более точными по сравнению с ними могут оказаться, например, несмещенные оценки [28].

Интересно посмотреть на развитие исследований в рассматриваемой области с точки зрения науковедения. При решении практической задачи (т.е. при разработке ГОСТ 11.011–83) выявились проблемы, для решения которых были предложены новые с научной точки зрения методы. Новые научные результаты, полученные в ходе решения практической задачи, привели к созданию новых научных направлений. Речь идет о статистике интервальных данных, а также об одношаговых оценках. Естественно, за развитием теории последовали ее практические применения.

Подобное развитие исследований — не единичный случай. Другой пример — разработка АСППАП (автоматизированной системы прогнозирования и предотвращения авиационных происшествий). Как отмечено в [29], ее создание оказалось возможным лишь в результате выполнения большого числа конкретных научных работ. Перечислим некоторые из них. Моделирование развития авиационных событий на различных этапах полета реализовано для 12 «деревьев» (ориентированных графов) с передаточными коэффициентами при движении от входных данных к вершинам «деревьев». Созданы и осуществлены новые процедуры экспертных оценок. Разработаны методы краткосрочного и долгосрочного прогнозирования, оценки и управления рисками в рассматриваемой области. С удовлетворением можно констатировать, что АСППАП позволяет значительно уменьшить вероятности авиационных происшествий и соответствующие риски. В [29] предложены 12 направлений дальнейшего развития АСППАП. Научные работы в этих направлениях продолжаются.

ЛИТЕРАТУРА

1. **Balakrishnan N., Nevzorov V. B.** A primer of statistical distributions. — New Jersey: Wiley-Interscience, 2003. — 328 p.
2. **Хастингс Н., Пикок Дж.** Справочник по статистическим распределениям. — М.: Статистика, 1980. — 95 с.
3. Вероятность и математическая статистика: Энциклопедия / Гл. ред. Ю. В. Прохоров. — М.: Большая Рос. Энцикл., 1999. — 910 с.
4. **Орлов А. И.** Сертификация и статистические методы (обобщающая статья) / Заводская лаборатория. Диагностика материалов. 1997. Т. 63. № 3. С. 55 – 62.
5. **Орлов А. И.** Статистический анализ выборок из бета-распределения / Научный журнал КубГАУ. 2023. № 03(187). С. 184 – 206.
6. **Ившин В. В.** Статистические задачи оценивания надежности в модели «нагрузка – прочность» в случаях гамма-распределения и распределения Вейбулла / Вестник Пермского государственного технического университета. Математическое моделирование систем и процессов. 1994. № 2. С. 43 – 49.
7. **Рязанский В. П., Юдин С. В.** Гамма-функция как основа трехпараметрического распределения параметров точности и надежности изделий оборонной промышленности / Известия Тульского государственного университета. Технические науки. 2022. № 2. С. 603 – 612.
8. **Johnson N. L., Kotz S., Balakrishnan N.** Continuous univariate distributions. Vol. 1. Second edition. — New York – Chichester – Brisbane – Toronto – Singapore: Jonh Wiley and Sons, 1994. — 777 p.
9. **Бендерский А. М., Невельсон М. Б.** Оценивание параметра формы гамма-распределения / Надежность и контроль качества. 1981. № 10. С. 14 – 21.
10. **Шор Я. Б.** Статистические методы анализа и контроля качества и надежности. — М.: Советское радио, 1962. — 553 с.
11. **Burgin T. A.** The gamma distribution and inventory control / Operational Research Quarterly. 1975. Vol. 26. N 4. P. 377 – 519.
12. **Свичинский С. В., Макаричев А. В.** Обоснование гамма-распределения длин путей следования пассажиров маршрутным транспортом / Модернизация и научные исследования в транспортном комплексе. 2013. Т. 2. С. 341 – 349.
13. **Волк А. М.** Анализ свойств статистических оценок параметров обобщенного гамма-распределения / Труды БГТУ. Серия 3: Физико-математические науки и информатика. 2023. № 1(266). С. 10 – 14.
14. **Кудрявцев А. А., Шестаков О. В., Шоргин В. С.** Метод оценивания параметров изгиба, формы и масштаба гамма-экспоненциального распределения / Информатика и ее применения. 2021. Т. 15. № 3. С. 57 – 62.
15. **Карпов И. Г., Зырянов Ю. Т., Мельник О. В.** Модель закона распределения непрерывных случайных величин на основе гамма-распределения / Фундаментальные и прикладные проблемы техники и технологии. 2014. № 3(305). С. 26 – 30.
16. **Долгов А. Ю., Терещенко Е. В., Сорочан А. А.** Исследование статистической гипотезы в условиях гамма-распределения / Вестник Приднестровского университета. Серия: Физико-математические и технические науки. Экономика и управление. 2018. № 3(60). С. 112 – 118.
17. **Орлов А. И.** Прикладной статистический анализ: учебник. — М.: Ай Пи Ар Медиа, 2022. — 812 с.
18. **Янке Е., Эмде Ф., Леш Ф.** Специальные функции. — М.: Наука, 1964. — 344 с.
19. **Кендалл М. Дж., Стьюарт А.** Статистические выводы и связи. — М.: Наука, 1973. — 896 с.
20. **Сатаров Г. А., Шмерлинг Д. С.** Новая статистическая модель парных сравнений / Экспертные оценки в задачах управления: Сб. тр. — М.: Изд-во Института проблем управления АН СССР, 1982. С. 67 – 79.
21. **Лапига А. Г.** Многокритериальные задачи управления качеством: построение прогноза качества в балльной шкале / Заводская лаборатория. 1983. Т. 49. № 7. С. 55 – 59.
22. **Закс Ш.** Теория статистических выводов. — М.: Мир, 1975. — 776 с.
23. **Бахмутов В. О., Косарев Л. Н.** Использование метода максимального правдоподобия для оценки однородности результатов усталостных испытаний / Заводская лаборатория. 1986. Т. 52. № 5. С. 52 – 57.
24. **Резникова А. Я., Шмерлинг Д. С.** Оценивание параметров вероятностных моделей парных и множественных сравнений / Статистические методы оценивания и проверки гипотез: Межвузовский сб. науч. тр. — Пермь: Изд-во Пермского госуниверситета, 1984. С. 110 – 120.
25. **Орлов А. И.** Искусственный интеллект: статистические методы анализа данных: учебник. — М.: Ай Пи Ар Медиа, 2022. — 843 с.
26. **Орлов А. И.** Статистика интервальных данных (обобщающая статья) / Заводская лаборатория. Диагностика материалов. 2015. Т. 81. № 3. С. 61 – 69.
27. **Орлов А. И., Луценко Е. В.** Системная нечеткая интервальная математика. Монография (научное издание). — Краснодар, КубГАУ, 2014. — 600 с.
28. **Лумельский Я. П.** К вопросу сравнения несмещенных и других оценок / Прикладная статистика: Сб. тр. — М.: Наука, 1983. С. 316 – 319.
29. **Бутов А. А., Волков М. А., Макаров В. П., Орлов А. И., Шаров В. Д.** Автоматизированная система прогнозирования и предотвращения авиационных происшествий при организации и производстве воздушных перевозок / Известия Самарского научного центра Российской академии наук. 2012. Т. 14. № 4(2). С. 380 – 385.

REFERENCES

1. **Balakrishnan N., Nevzorov V. B.** A primer of statistical distributions. — New Jersey: Wiley-Interscience, 2003. — 328 p.
2. **Hastings N., Peekock J.** Handbook of Statistical Distributions. — Moscow: Statistika, 1980. — 95 p. [Russian translation].
3. Probability and mathematical statistics: Encyclopedia / Prokhorov Yu. V., Ed. — Moscow: Bol'shaya Ros. Entsikl., 1999. — 910 p. [in Russian].

4. **Orlov A. I.** Certification and statistical methods (summarizing article) / Zavod. Lab. Diagn. Mater. 1997. Vol. 63. N 3. P. 55 – 62 [in Russian].
5. **Orlov A. I.** Statistical analysis of samples from the beta distribution / Nauch. Zh. KubGAU. 2023. N 03(187). P. 184 – 206 [in Russian].
6. **Ivshin V. V.** Statistical problems of reliability estimation in the load-strength model in the cases of gamma and Weibull distributions / Vestn. Perm. Gos. Tekhn. Univ. Matem. Model. Sist. Prots. 1994. N 2. P. 43 – 49 [in Russian].
7. **Ryazanskii V. P., Yudin S. V.** Gamma function as the basis of a three-parameter distribution of parameters of accuracy and reliability of defense industry products / Izv. Tul. Gos. Univ. Tekhn. Nauki. 2022. N 2. P. 603 – 612 [in Russian].
8. **Johnson N. L., Kotz S., Balakrishnan N.** Continuous univariate distributions. Vol. 1. Second edition. — New York – Chichester – Brisbane – Toronto – Singapore: Jonh Wiley and Sons, 1994. — 777 p.
9. **Benderskii A. M., Nevel'son M. B.** Estimating the Shape Parameter of the Gamma Distribution / Nadezhn. Kontrol' Kach. 1981. N 10. P. 14 – 21 [in Russian].
10. **Shor Ya. B.** Statistical methods of analysis and quality and reliability control. — Moscow: Sovetskoe radio, 1962. — 553 p. [in Russian].
11. **Burgin T. A.** The gamma distribution and inventory control / Operational Research Quarterly. 1975. Vol. 26. N 4. P. 377 – 519.
12. **Svichinskii S. V., Makarichev A. V.** Justification for the gamma distribution of route lengths for passengers traveling by route transport / Modern. Nauch. Issl. Transp. Kompl. 2013. Vol. 2. P. 341 – 349 [in Russian].
13. **Volk A. M.** Analysis of the properties of statistical estimates of the parameters of the generalized gamma distribution / Tr. BGTU. Ser. 3: Fiz.-Mat. Nauki Inform. 2023. N 1(266). P. 10 – 14 [in Russian].
14. **Kudryavtsev A. A., Shestakov O. V., Shorgin V. S.** Method for estimating the bending parameters, shape and scale of the gamma exponential distribution / Inform. Primen. 2021. Vol. 15. N 3. P. 57 – 62 [in Russian].
15. **Karpov I. G., Zyryanov Yu. T., Mel'nik O. V.** Model of the distribution law of continuous random variables based on the gamma distribution / Fund. Prikl. Probl. Tekhn. Tekhnol. 2014. N 3(305). P. 26 – 30 [in Russian].
16. **Dolgov A. Yu., Tereshchenko E. V., Sorochan A. A.** Study of a statistical hypothesis under the conditions of a gamma distribution / Vestn. Pridnestr. Univ. Ser. Fiz.-Mat. Tekhn. Nauki. Ékon. Upravl. 2018. N 3(60). P. 112 – 118 [in Russian].
17. **Orlov A. I.** Applied statistical analysis: textbook. — Moscow: IPR Media, 2022. — 812 p. [in Russian].
18. **Yanke E., Emde F., Lesh F.** Special functions. — Moscow: Nauka, 1964. — 344 p. [in Russian].
19. **Kendall M. J., Stewart A.** Statistical inferences and relationship. — Moscow: Nauka, 1973. — 896 p. [Russian translation].
20. **Satarov G. A., Shmerling D. S.** New statistical model for paired comparisons / Expert assessments in management problems: Collection of papers. — Moscow: Izd. IPU AN SSSR, 1982. P. 67 – 79 [in Russian].
21. **Lapiga A. G.** Multicriteria quality management problems: constructing a quality forecast on a point scale / Zavod. Lab. 1983. Vol. 49. N 7. P. 55 – 59 [in Russian].
22. **Zaks Sh.** The theory of statistical inference. — Moscow: Mir, 1975. — 776 p. [Russian translation].
23. **Bakhmutov V. O., Kosarev L. N.** Using the maximum likelihood method to evaluate the homogeneity of fatigue test results / Zavod. Lab. 1986. Vol. 52. N 5. P. 52 – 57 [in Russian].
24. **Reznikova A. Ya., Shmerling D. S.** Estimating the parameters of probabilistic models of paired and multiple comparisons / Statistical methods of estimation and testing of hypotheses: Interuniversity collection of scientific papers. — Perm': Izd. Perm. Gos. Univ., 1984. P. 110 – 120 [in Russian].
25. **Orlov A. I.** Artificial intelligence: statistical methods of data analysis: textbook. — Moscow: IPR Media, 2022. — 843 p. [in Russian].
26. **Orlov A. I.** Statistics of interval data (summarizing article) / Zavod. Lab. Diagn. Mater. 2015. Vol. 81. N 3. P. 61 – 69 [in Russian].
27. **Orlov A. I., Lutsenko E. V.** System fuzzy interval mathematics. Monograph (scientific publication). — Krasnodar, KubGAU, 2014. — 600 p. [in Russian].
28. **Lumel'skii Ya. P.** On the issue of comparison of unbiased and other estimates / Applied statistics: Collection of works. — Moscow: Nauka, 1983. P. 316 – 319 [in Russian].
29. **Butov A. A., Volkov M. A., Makarov V. P., Orlov A. I., Sharov V. D.** Automated system for forecasting and preventing aviation accidents during the organization and production of air transportation / Izv. Samar. NTs RAN. 2012. Vol. 14. N 4(2). P. 380 – 385 [in Russian].