DOI 10.1007/s10598-019-09456-9
Computational Mathematics and Modeling, Vol. 30, No. 3, July, 2019

PARAMETER ESTIMATION IN A THREE-PARAMETER LOGNORMAL DISTRIBUTION

V. D. Kozlov[1] and **A. I. Maysuradze**[2]

UDC 519.233.22, 519.24+.25

Point estimation methods for the three-parameter lognormal distribution are investigated and compared. The lognormal distribution is required in many topical areas, but so far there have been no comparative studies of the various estimation methods. We show that despite the large number of traditional estimation methods, the lognormal distribution requires special methods. We accordingly consider specializations of the main parameter estimation approaches, including the actively developing distance minimization methods. Their accuracy and speed are compared on simulated data. We show that specialized parameter-estimation methods may outperform the highly popular maximum likelihood method.

Keywords: point estimation, three-parameter lognormal distribution, distance minimization methods, maximum likelihood method, moment methods, problem-oriented methods.

Introduction

Reconstruction of a distribution from a finite sample of observations is one of the core problems in statistics and data analysis in general. Many standard methods are available for its solution. In practice, however, the standard methods are inapplicable to problems arising in some subject areas. In such cases, new problem-oriented methods have to be developed for the reconstruction of distributions.

The need for a problem-oriented density-estimation itechnique arose [1] in the construction of a data transmission delays between nodes in a computer cluster, in particular a supercomputer. Previously, similar studies were conducted in local area networks and the Internet [2–4]. All these studies show that the delays are adequately described by a three-parameter gamma-distribution or a three-parameter lognormal distribution. The specific features of the problem are attributable to the following properties of the communication environment in various supercomputers [1]:

1. The distribution is multimodal.
2. The distribution has a fine multimodal structure, i.e., the large density peak actually consists of many smaller peaks.
3. The data contain many repetitions and few unique values.
4. A large number of maximum delays are observed, which is indicative of a distribution with a heavy tail.

Due to these properties, the modeling of data transmission delays in a supercomputer communication environment is far from trivial: it requires careful formulation and search for new specialized methods. This problem can be formalized as the separation of a finite mixture of three-parameter lognormal distributions [1]. However, difficulties arise even in parametric estimation of a single component of this mixture. Our study accordingly analyzes

Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, Moscow, Russia.
[1] E-mail: kozlov.vld@mail.ru.
[2] E-mail: maysuradze@cs.msu.ru.

Translated from Prikladnaya Matematika i Informatika, No. 60, 2019, pp. 79–89.

 1046-283X/19/3003–0302

Table 1

Main Numerical Characteristic of 3LN

with the Parameters $\gamma,\ \mu,\ \sigma$ [9]; $\beta = \exp(\mu),\ \omega = \exp(\sigma^2) \geqslant 1$

Moment	Formula
Expectation $\mathbb{E}$	$\gamma + \beta\sqrt{\omega}$
Variance $\mathbb{V}$	$\beta^2\omega(\omega - 1)$
Skewness α_3	$\sqrt{\omega - 1}(\omega + 2)$
Kurtosis α_4	$\omega^4 + 2\omega^3 + 3\omega^2 - 6$

and compares different parameter estimation methods for a three-parameter lognormal distribution using a finite sample.

The study is organized as follows. We start with basic notation and definitions. The sections that follow present the maximum likelihood method, the moment method, and the minimum distance method in application to the parameter estimation in a three-parameter lognormal distribution. Then we describe the data used in our tests and compare the optimization methods for distance functionals between distributions. Different parameter estimation methods are then compared. Conclusion conclude.

Main Definitions and Notation

The three-parameter lognormal distribution (3LN distribution, or 3LND) is an absolutely continuous one-dimensional distribution with the density function

$$p(x;\gamma,\mu,\sigma) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\sigma(x-\gamma)} \exp\left(-\dfrac{(\ln(x-\gamma)-\mu)^2}{2\sigma^2}\right), & x \geqslant \gamma, \\ 0, & x < \gamma. \end{cases} \tag{1}$$

A random variable X is 3LN with the parameters $\gamma,\ \mu$, and σ if the random variable $\ln(X-\gamma)$ is normal with the parameters μ and σ [7]. Therefore, its probability distribution function can be written in the form

$$F(x;\gamma,\mu,\sigma) = \Phi\left(\frac{\ln(x-\gamma)-\mu}{\sigma}\right) \qquad \text{for} \quad x > \gamma,$$

where $\Phi(x)$ is the standard normal distribution. The main numerical characteristics of 3LN are given in Table 1. They will be needed in what follows.

We use the following notation throughout this article:

- $F(\cdot;\theta)$ is a probability distribution with the set of parameters $\theta = (\theta_1, \ldots, \theta_m)$; for 3LN, $\theta = (\gamma,\mu,\sigma)$, $\theta_1 = \gamma,\ \theta_2 = \mu,\ \theta_3 = \sigma$;
- $X \sim F(\cdot)$ is the random variable X with the probability distribution $F(\cdot)$;
- $X^n = \{X_1, \ldots, X_n\}$ is a sample of n independent identically distributed random variables; $X_{(k,n)}$ is the k-th order statistics X^n;

– $x^n = \{x_1, \dots, x_n\}$ is the observed sample, $x_{(k)}$ is the k-th element in the ordered sample;

– $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$ is the standard normal distribution, $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ is the error function;

– $\overline{1, k}$ is the set of integers from 1 to k, inclusive.

The Maximum Likelihood Method

The most popular approach to parameter estimation is the maximum likelihood method. The goodness of fit of the distribution $F(\cdot; \theta)$ to the observations x^n is measured by the likelihood function $L(\theta)$ equal to $p(x_1, \dots, x_n; \theta)$ — the joint probability density of the sample observations. It is assumed that the higher the likelihood, the better the fit of the model to the observations [5].

For 3LN, the log likelihood is written as

$$\ln L(\theta) = -n \ln \sigma \sqrt{2\pi} - \sum_{i=1}^{n} \ln(x_i - \gamma) - \sum_{i=1}^{n} \left(\frac{(\ln(x_i - \gamma) - \mu)^2}{2\sigma^2} \right), \tag{2}$$

this expression is meaningful only for $\gamma < x_{(1)}$. The necessary extremum conditions for (2) are

$$\begin{cases} \dfrac{\partial \ln L}{\partial \gamma} = \displaystyle\sum_{i=1}^{n} \frac{1}{x_i - \gamma} \left(1 + \frac{\ln(x_i - \gamma) - \mu}{\sigma^2} \right) = 0, \\ \dfrac{\partial \ln L}{\partial \mu} = \displaystyle\sum_{i=1}^{n} \frac{\ln(x_i - \gamma) - \mu}{\sigma^2} = 0, \\ \dfrac{\partial \ln L}{\partial \sigma} = \displaystyle\sum_{i=1}^{n} \frac{1}{\sigma} \left(-1 + \frac{(\ln(x_i - \gamma) - \mu)^2}{\sigma^2} \right) = 0. \end{cases} \tag{3}$$

The maximum likelihood method is successfully applied to many parameter estimation problems. However, its application to 3LN parameter estimation is problematic. It has been shown [10] that for every sample x^n the maximum likelihood function $L(\theta)$ is unbounded, namely there exist trajectories in the three-dimensional parameter space (γ, μ, σ) that converge to $(x_{(1)}, -\infty, +\infty)$ and along which $L(\theta)$ converges to $+\infty$, while at the point $(x_{(1)}, -\infty, +\infty)$ the likelihood function is zero. It is nevertheless asserted [10] that despite the overall unboundedness of the likelihood function $L(\theta)$, it often has a local maximum if the sample observations take sufficiently many different values "near" the true values of the parameters θ. This suggest the idea of using local maximum likelihood estimates that correspond to a local maximum of the likelihood function. Local estimates can be obtained as the numerical solution of system (3) [11]:

$$\begin{aligned} &\mu(\gamma) = \frac{1}{n} \sum_{i=1}^{n} \ln(x_i - \gamma), \\ &\sigma^2(\gamma) = \frac{1}{n} \sum_{i=1}^{n} \left(\ln(x_i - \gamma) - \mu(\gamma) \right)^2, \\ &\gamma\colon \ \sum_{i=1}^{n} \frac{1}{x_i - \gamma} \left(1 + \frac{\ln(x_i - \gamma) - \mu(\gamma)}{\sigma^2(\gamma)} \right) = 0. \end{aligned} \tag{4}$$

The Generalized Moment Method

In parameter estimation by the generalized moment method, the probability distribution $F(\cdot;\theta)$ is usually constrained by a system of equalities $g_i(\theta) = h_i(x^n),\ i = \overline{1,k}$, where the functions $g_i(\theta)$ characterize the distribution and $h_i(x^n)$ are their sample estimates, unbiased or at least asymptotically unbiased [5].

The general moment method has been previously applied to estimate 3LN parameters because of the difficulties with the maximum likelihood method [8, 9]. If $g_{1,2,3}(\theta)$ are the expectation, the variance, and the skewness (see Table 1), and $h_{1,2,3}(X^n)$ are their sample estimates [6], we obtain the system of equations

$$\begin{cases} \gamma + \beta\sqrt{\omega} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i = \overline{x}, \\ \beta^2\omega(\omega-1) = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(x_i-\overline{x})^2, \\ \sqrt{\omega-1}(\omega+2) = \dfrac{\dfrac{n}{(n-1)(n-2)}\sum_{i=1}^{n}(x_i-\overline{x})^3}{\left(\dfrac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2\right)^{\frac{3}{2}}} = a. \end{cases} \tag{5}$$

The third equation, for any values of the right-hand side a, has the unique solution

$$\omega = 1 + \left(\sqrt[3]{\frac{\sqrt{a^2+4}+a}{2}} - \sqrt[3]{\frac{\sqrt{a^2+4}-a}{2}}\right)^2$$

(see [9]) which analytically produces the estimates for $\sigma = \sqrt{\ln\omega},\ \mu = \ln\beta$ and γ.

The 3LN parameters also have been estimated by the L-moment method [12]. The L-moment of order r for the distribution $F(\cdot)$ is defined as

$$\lambda_r = \frac{1}{r}\sum\nolimits_{k=0}^{r-1}(-1)^k\binom{r-1}{k}\mathbb{E}X_{(r-k,r)}.$$

The statistic

$$l_r = \binom{n}{r}^{-1}\sum_{1\leqslant i_1\leqslant\cdots\leqslant i_r\leqslant n}\frac{1}{r}\sum_{k=0}^{r-1}(-1)^k\binom{r-1}{k}x_{(i_{r-k})}$$

is an unbiased estimate of λ_r [13]. For 3LN we write the following system of constraints [12, 13]:

$$\begin{cases} \gamma + \exp\left(\mu + \dfrac{\sigma^2}{2}\right) = l_1, \\ \exp\left(\mu + \dfrac{\sigma^2}{2}\right)\operatorname{erf}\left(\dfrac{\sigma}{2}\right) = l_2, \\ \dfrac{6}{\sqrt{\pi}\operatorname{erf}\left(\frac{\sigma}{2}\right)}\displaystyle\int_0^{\frac{\sigma}{2}}\operatorname{erf}\left(\dfrac{x}{\sqrt{3}}\right)\exp\left(-x^2\right)dx = \dfrac{l_3}{l_2}. \end{cases} \tag{6}$$

Table 2

Distances Between the Distributions $F(\cdot)$ and $G(\cdot)$ [6]

Distance	$d[F,G]$
Kolmogorov–Smirnov	$\sup_x \lvert F(x)-G(x)\rvert$
Cramer–von Mises	$\int\limits_{-\infty}^{+\infty} (F(x)-G(x))^2 dF(x)$
Anderson–Darling	$\int\limits_{-\infty}^{+\infty} \frac{(F(x)-G(x))^2}{F(x)\,(1-F(x))} dF(x)$

Table 3

Sets of Parameters for Simulation

	θ^1	θ^2	θ^3	θ^4	θ^5	θ^6	θ^7
γ	3	10	16	10	10	10	10
μ	3	3	3	2	4	3	3
σ	0.23	0.23	0.23	0.23	0.23	0.1	0.35

For an ordered sample x^n the computation of $l_1,\ l_2,$ and l_3 is of complexity $O(n)$. For system (6) we obtain the approximate solution [12]

$$z = \sqrt{\frac{8}{3}}\Phi^{-1}\left(\frac{1+\frac{l_3}{l_2}}{2}\right),$$

$$\sigma \approx 0.999281z - 0.006118z^3 + 0.000127z^5,$$

$$\mu = \ln\left(\frac{l_2}{\operatorname{erf}\left(\frac{\sigma}{2}\right)}\right) - \frac{\sigma^2}{2}, \qquad (7)$$

$$\gamma = l_1 - \exp\left(\mu + \frac{\sigma^2}{2}\right).$$

The Minimum Distance Method

In the class of distance minimization methods, the goodness of fit measure for the model and the observations is some distance $d[\cdot,\cdot]$ between the theoretical and the empirical data distributions. It is assumed that the smaller the distance, the better is the model fit. For continuous distributions, the distance is usually taken between the model distribution function $F(x;\theta)$ and the empirical distribution

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}[x > x_{(i)}].$$

The term "distance" in this case is strictly conventional: the functional d may even prove to be non-symmetric, the only requirements are that it be nonnegative and vanish only if the distributions are equal [14].

The application of the minimum distance method to 3LN parameter estimation is rarely covered in the literature. In the present article, we consider minimization of three distance measures: Kolmogorov–Smirnov, Cramer–von Mises, and Anderson–Darling (Table 2). Note that for an ordered sample x^n the distance $d[F, F_n]$ is computed in time $O(n)$.

Computer Experiments

In this section, we compare the various 3LN parameter estimation methods: the local maximum likelihood method (LML in what follows), the generalized moment method (GM), the L-moment method (L-mom), Kolmogorov-Smirnov distance minimization (KS), Cramer–von Mises distance minimization (CvM), and Anderson-Darling distance minimization (AD). The methods were programmed in Python using the numerical optimization libraries SciPy and NLopt.

The methods were tested on simulated data — samples from 3NL with known parameter values $\theta = (\gamma, \mu, \sigma)$. We took seven sets of parameters (see Table 3) obtained from the analysis of message transmission delays in a local-area network and on the Internet [3]. Then for each set of parameters we generated from 3NL $n = 100$ samples of $N = 10{,}000$ each. In this way, we could compare the estimates obtained by various methods with the true parameter value and investigate their statistical properties.

The performance of the various estimation methods was compared by the following measures:

– the bias of parameter z: $b_z = (\overline{z} - z)^2$;

– the efficiency of the estimate of parameter z: $p_z = \frac{1}{n}\sum_{i=1}^{n}(\hat{z}_i - \overline{z})^2$;

– mean estimation time.

Here z stands for one of the parameters γ, μ, or σ, $\hat{z}_i$ is the estimate of parameter z from sample i,

$$\overline{z} = \frac{1}{n}\sum_{i=1}^{n}\hat{z}_i.$$

The overall accuracy of the estimate of the parameter z, $a_z = \frac{1}{n}\sum_{i=1}^{n}(\hat{z}_i - z)^2$, is expressed as $a_z = b_z + p_z$.

The results are presented in Tables 4 and 5.

Discussion

We should first note that for all estimation methods the estimate bias is close to 0, which suggests that all the methods are asymptotically unbiased. The overall estimate accuracy az is predominantly determined by its efficiency: we see from Table 4 that the efficiency is usually two orders of magnitude greater than the bias. It is the estimate efficiency that plays the main role in these cases.

Table 4
Bias and Efficiency of the Estimates of $\gamma,\ \mu,\ \sigma$ *

		θ^1	θ^2	θ^3	θ^4	θ^5	θ^6	θ^7
$b_\gamma \sim 10^{-2}$	LML	0.2	0.002	1.2	0.03	3.6	18.1	0.2
	GM	0.6	3.4	0.3	0.2	1.4	17.1	0.2
	L-mom	0.6	1.1	0.03	0.005	1.0	11.9	0.1
	KS	3.4	2.6	0.3	0.2	1.1	5.5	0.1
	CvM	3.3	4.4	1.6	0.1	6.3	1.9	0.1
	AD	3.6	3.6	0.01	0.04	0.1	6.0	0.03
$b_\mu \sim 10^{-5}$	LML	0.4	0.0002	3.4	1.0	1.6	31.5	0.5
	GM	1.0	7.5	1.2	5.4	1.3	27.2	0.3
	L-mom	1.1	2.4	0.2	0.0004	0.7	17.5	0.3
	KS	2.2	1.7	1.2	5.2	2.5	3.8	0.8
	CvM	2.3	3.9	2.1	0.7	5.5	0.2	0.8
	AD	2.5	2.9	0.2	0.3	1.5	4.8	0.6
$b_\sigma \sim 10^{-7}$	LML	2.1	0.7	12.7	7.5	14.7	21.4	3.0
	GM	4.1	23.5	4.7	36.6	17.3	16.1	0.4
	L-mom	4.9	5.0	0.4	0.4	10.3	9.6	1.0
	KS	0.2	2.6	7.0	38.2	44.7	0.2	41.2
	CvM	0.4	0.05	7.3	0.1	69.7	1.6	42.1
	AD	0.4	0.6	0.7	0.2	34.7	0.5	36.2
$p_\gamma \sim 10^{0}$	LML	0.47	0.48	0.36	0.06	3.84	2.81	0.17
	GM	0.83	0.83	0.69	0.13	9.09	3.41	0.76
	L-mom	0.65	0.62	0.51	0.09	5.99	3.34	0.32
	KS	4.0	3.8	0.70	0.13	18.9	4.61	1.83
	CvM	3.89	4.0	1.47	0.22	17.34	4.42	1.83
	AD	4.05	3.91	0.71	0.11	19.44	4.31	1.82
$p_\mu \sim 10^{-3}$	LML	1.21	1.25	0.95	1.19	1.33	6.80	0.47
	GM	2.15	2.16	1.81	2.57	3.20	8.17	2.06
	L-mom	1.68	1.59	1.31	1.72	2.04	7.81	0.88
	KS	9.50	8.81	1.81	1.53	6.39	11.23	4.42
	CvM	9.26	9.24	3.59	4.15	5.95	11.73	4.42
	AD	9.6	9.06	1.78	2.13	6.57	10.76	4.40
$p_\sigma \sim 10^{-5}$	LML	6.78	6.42	5.09	6.47	6.91	6.59	6.15
	GM	11.15	10.88	9.08	12.97	16.56	7.94	23.29
	L-mom	9.14	7.91	6.64	9.06	10.37	7.32	10.53
	KS	52.24	47.09	8.97	12.42	34.68	11.28	60.23
	CvM	51.07	49.44	18.27	22.47	32.94	13.06	60.20
	AD	52.71	48.13	9.42	11.64	35.94	11.12	59.94

* For readability, the significance is shown in the first column.

Table 5

Execution time of various methods in seconds

	θ^1	θ^2	θ^3	θ^4	θ^5	θ^6	θ^7
LML	0.0355	0.0333	0.0335	0.0313	0.0377	0.0399	0.0317
GM	0.0015	0.0016	0.0016	0.0016	0.0016	0.0015	0.0015
L-mom	0.0019	0.0022	0.002	0.0022	0.002	0.0018	0.0020
KS	0.3588	0.3472	0.1761	0.2079	0.3994	0.1916	0.3129
CvM	0.3087	0.3672	0.2139	0.4212	0.3259	0.2095	0.2505
AD	0.3623	0.3728	0.2659	0.2454	0.3571	0.21	0.2467

The various methods are ranked roughly in the same order by the quality of the various parameter estimates. There are no instances when one of any two methods produces a substantially better estimate for one of the parameters and a substantially worse estimate for another. It is therefore superfluous to estimate different parameters by different methods. Furthermore, the relative estimation quality is not sensitive to the specific choice of parameters. The methods can be easily ranked by performance.

We see from Table 4 that the local maximum likelihood method produces the best estimates, close results are obtained with the L-moment method, and these two are followed by the Kolmogorov-Smirnov and Anderson-Darling methods; the last in the ranking is the Cramer–von Mises method. We emphasize that the L-moment method and the generalized moment method give results that are comparable with the local maximum likelihood method, but they are an order of magnitude faster (see Table 5). Moreover, the generalized moment method, unlike the local maximum likelihood method, guarantees the existence of an estimate for every sample.

Conclusion

We have examined various parameter estimation method for the three-parameter lognormal distribution. This is the first detailed investigation of the application of the class of minimum distance methods to parameter estimation of 3LN distributions. This is also the first testing and comparison of the methods on simulation data based on efficiency and accuracy. It has been shown that the methods can be ranked by estimate quality. We have identified methods that achieve high computational efficiency with a relatively small loss of accuracy compared with the leaders. The results of this study are used in the next study focusing on the separation of a mixture of three-parameter lognormal distribution, which in turn is necessary for the simulation of information transmission delays in communication environments of various computer clusters.

Supported by Moscow State University (grant NIR 5.1.16) and the Russian Foundation for Basic Research (grant 17-01-00917).

REFERENCES

1. A. A. Gorelov, A. I. Maysuradze, and A. N. Salnikov, "Delay structure mining in computing cluster," *CEUR Workshop Proceedings*, **1482**, 546–551, M. Jeusfeld c/o Redaktion Sun SITE, Informatik V, RWTH, Aachen, Germany (2015).
2. A. Corlett, D. I. Pullin, and S. Sargood, "Statistics of one-way internet packet delays," *Proceedings of the 53rd Internet Engineering Task Force* (2002).

3. M. Karakas, *Determination of Network Delay Distribution over the Internet*, Master Thesis, The Middle East Technical University (2003).
4. A. Mukherjee, "On the dynamics and significance of low frequency components of internet load," *Internetworking: Research and Experience*, **5**, 163–205 (1994).
5. S. A. Aivazyan and V. S. Mkhitaryan, *Applied Statistics. Foundations of Econometrics, Vol. 1: Probability Theory and Applied Statistics* [in Russian], Yuniti-Dana, Moscow (2001).
6. A. I. Kobzar', *Applies Mathematical Statistics* [in Russian], Fizmatlit, Moscow (2006).
7. N. L. Johnson, S. B. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, Wiley, New York (1994).
8. A. C. Cohen and B. J. Whitten, "Estimation in the three-parameter lognormal distribution," *J. Amer. Stat. Association*, **75(370)**, 399–404 (1980).
9. A. C. Cohen and B. J. Whitten, *Parameter Estimation in Reliability and Life Span Models*, Marcel Dekker (1988).
10. B. M. Hill, "The three-parameter lognormal distribution and bayesian analysis of a point-source epidemic," *J. Amer. Stat. Association*, **58(301)**, 72–84 (1963).
11. F. Calitz, "maximum likelihood estimation of the parameters of the three-parameter lognormal distribution — a reconsideration," *Australian J. Statistics*, **15(3)**, 185–190 (1973).
12. D. Bilkova, "Three-parametric lognormal distribution and estimating its parameters using the method of L-moments," *RELIK — Reprodukce Lidskeho Kapitalu* (2011).
13. J. R. M. Hosking, "L-moments: Analysis and estimation of distributions using linear combinations of order statistics," *J. Royal Stat. Soc. Series B (Methodological)*, **52(1)**, 105–124.
14. A. Basu, H. Shioya, and C. Park, *Statistical Inference: The Minimum Distance Approach*, Taylor & Francis (2011).