CSSE-16 June 7, 2001

Easy Estimation by a New Parameterization in the Three-Parameter Lognormal Distribution

Yoshio Komori and Hideo Hirose

Department of Control Engineering and Science
Kyushu Institute of Technology

680-4 Kawazu
Iizuka, 820-8502, Japan
{*komori* & *hirose*}*@ces.kyutech.ac.jp*

Abstract

A new computing method is proposed for the primary relative maximum of the likelihood function in the three-parameter lognormal distribution. In the method the distribution is transformed to the extended lognormal distribution, and the three-parameter estimation problems for the extended distribution are changed to two-parameter estimation problems, which demand to maximize an object function of the two parameters. Since the function goes to $+\infty$ only if both parameters simultaneously go to $+\infty$ or $-\infty$, the two-parameter estimation problems can be expected to avoid computational difficulties caused by the non-regularity of the likelihood function in the lognormal distribution. In addition, since the employment of graphical tools makes it possible to easily find proper initial guesses given to iterative methods in the two-parameter problem, the combination of the reparameterization and graphical tools is a simple but highly effective method to cope with cases where the selection of the initial guess is difficult. In the present article, furthermore, the analysis of the object function is given, and the properties of the estimator are investigated in simulations. Some examples are given for illustration.

1 Introduction

The three-parameter lognormal distribution is one of the most important distributions in biological and sociological fields. With a variable x and three parameters α, β and γ, the probability density function is expressed by

$$f(x;\alpha,\beta,\gamma) \stackrel{\text{def}}{=} \frac{1}{\sqrt{2\pi}(x-\alpha)\beta}\exp\left[-\frac{\{\ln((x-\alpha)/\gamma)\}^2}{2\beta^2}\right], \quad x>\alpha,\ \beta>0,\ \gamma>0 \quad (1.1)$$

and the likelihood function is expressed by $L(\alpha,\beta,\gamma) \stackrel{\text{def}}{=} \prod_{i=1}^{n} f(x_i;\alpha,\beta,\gamma)$. Here, x_i $(1 \le i \le n)$ stand for independent observations. Without loss of generality, we assume $x_1 > x_2 \ge \cdots \ge x_{n-1} > x_n$.

Since $\ln(X-\alpha)$ obeys a normal distribution if a random variable X obeys a lognormal distribution, $L(\alpha,\beta,\gamma)$ achieves its maximum at a point $(\alpha_0, \hat{\beta}(\alpha_0), \hat{\gamma}(\alpha_0))$ provided that α is fixed to α_0, where

$$\hat{\beta}(\alpha) \stackrel{\text{def}}{=} \sqrt{\frac{1}{n}\sum_{i=1}^{n}\{\ln(x_i-\alpha)-\ln\hat{\gamma}(\alpha)\}^2} \quad \text{and} \quad \hat{\gamma}(\alpha) \stackrel{\text{def}}{=} \exp\left[\frac{1}{n}\sum_{i=1}^{n}\ln(x_i-\alpha)\right].$$

Consequently, if we want to obtain the maximum likelihood estimate, it suffices to find an α such that $\hat{L}(\alpha) \stackrel{\text{def}}{=} L(\alpha,\hat{\beta}(\alpha),\hat{\gamma}(\alpha))$ achieves its maximum. However, because that $\hat{L}(\alpha) \to \infty$ as $\alpha \to x_n - 0$, $L(\alpha,\beta,\gamma)$ becomes unbounded. Furthermore, the other parameters then lead to inadmissible values.

Hill (1), using the Bayes theorem, has given a statistical implication of $(\hat{\alpha}, \hat{\beta}(\hat{\alpha}), \hat{\gamma}(\hat{\alpha}))$ where $L(\alpha,\beta,\gamma)$ has its maximum in the region except the singular region; for a small $\delta > 0$, at $\alpha = \hat{\alpha}$ has $\hat{L}(\alpha)$ its relative and absolute maximum under the condition $x_n - \alpha > \delta$. The point $(\hat{\alpha}, \hat{\beta}(\hat{\alpha}), \hat{\gamma}(\hat{\alpha}))$ is used instead of the maximum likelihood estimate, and it is called the primary relative maximum (PRM) or the local maximum likelihood estimate of the likelihood function. Displaying $\hat{L}(\alpha)$ is an effective way for finding $\hat{\alpha}$, but the search might be difficult because that the shape of $\hat{L}(\alpha)$ is complicated in some cases depending on data sets (1, 2).

For problems to seek $\hat{\alpha}$, that is, the one-parameter estimation problems for the lognormal distribution, Wingo (3–5) has proposed a computing method to avoid the singular range $x_n - \alpha \le \delta$ by adopting a penalty function. On the other hand, for problems to seek simultaneously $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\gamma}$, where $(\hat{\alpha}, \hat{\beta}, \hat{\gamma})$ is the PRM, that is, the three-parameter estimation problems (6) for the distribution, Munro and Wixley (7) has proposed a parameterization to improve the convergency of an iterative method (8). As seen now, there are two ways for dealing with the parameter estimation of the distribution for complete data. Besides these, Giesbrecht (9) has proposed replacing given data with grouped data to avoid the singularity stated above.

The use of Munro's parameterization: the substitutions of $\alpha = \mu - \sigma/\lambda$, $\beta = \lambda$ and $\gamma = \sigma/\lambda$ into (1.1) yield

$$f(x;\mu-\sigma/\lambda,\lambda,\sigma/\lambda) = \frac{1}{\sqrt{2\pi}\{\sigma+\lambda(x-\mu)\}}\exp\left[-\frac{\{\ln(\sigma+\lambda(x-\mu))-\ln\sigma\}^2}{2\lambda^2}\right].$$

This can be extended by allowing $x < \mu - \sigma/\lambda$, then the generalization permits λ to be negative. In this way, we obtain the density function for the extended lognormal distribution permitting that $\lambda \neq 0$ and $\sigma > 0$. Let $\tilde{f}(x; \lambda, \mu, \sigma)$ be this density function and $\tilde{L}(\lambda, \mu, \sigma)$ the likelihood function. Cheng and Iles (10) has showed that as $\lambda \to 0$, $\tilde{f}(x; \lambda, \mu, \sigma)$ leads to the normal distribution with mean μ and variance σ^2, which is called the embedded distribution. And they have investigated tests of statistical hypothesis to see whether the embedded model should be used.

The following is pointed out in (11):

> although the use of the extended lognormal distribution may reduce the difficulty of the parameter estimation accompanying the lognormal distribution, selecting a proper initial guess remains as a troublesome task because that we have to finally use an iterative method to complete the estimation.

And it has been shown in simulations that the convergency of the sequence of approximate solutions by the continuation method is better than those by other competitors. Possible missing PRMs, however, remains even if the continuation method is used.

In the present article we propose a new computing method for PRMs. It consists of four factors; that is, an extension of the lognormal distribution, a reparameterization, the selection of initial values in graphical ways, and the employment of iterative methods. One merit of the reparameterization is the avoidance of computational difficulties caused by the non-regularity of the likelihood function in the lognormal distribution. Another merit is that through graphic tools of computer we can finger the object function to be maximized. This makes it easier to select proper initial guesses.

In Section 2 we describe the reparameterization and consider the search region for estimates. In Section 3 we introduce a searching example with the proposed method, and investigate how the presented reparameterization influences the bias and root mean squared error (RMSE) in the original parameterization by means of Monte Carlo simulation experiments. In the last of the section we show an example, in which it is difficult to select a proper initial guess even if the continuation method is chosen as a solver. The summary and remarks are given in Section 4. Finally, The proof of a lemma stated in the body of this article are given in Appendix.

2 Two-parameter estimation

In this section we analyze a function maximized to find a PRM. First of all we introduce the function.

Set $\tau = \sigma - \lambda\mu$ and $s = \ln \sigma$, and define $\bar{f}(x; \lambda, \tau, s) \overset{\text{def}}{=} \tilde{f}(x; \lambda, (\mathrm{e}^s - \tau)/\lambda, \mathrm{e}^s)$:

$$\bar{f}(x; \lambda, \tau, s) \overset{\text{def}}{=} \frac{1}{\sqrt{2\pi}(\lambda x + \tau)} \exp\left[-\frac{\{\ln(\lambda x + \tau) - s\}^2}{2\lambda^2}\right], \qquad \lambda \neq 0.$$

By arranging $\ln \bar{L}(\lambda, \tau, s) \overset{\text{def}}{=} \sum_{i=1}^n \ln \bar{f}(x_i; \lambda, \tau, s)$, we obtain

$$\begin{aligned}\ln \bar{L}(\lambda, \tau, s) \;=\; & -\frac{n}{2\lambda^2}\left\{s - \frac{1}{n}\sum_{i=1}^n \ln(\lambda x_i + \tau)\right\}^2 - n \ln\sqrt{2\pi} \\ & + \frac{1}{2n\lambda^2}\left\{\sum_{i=1}^n \ln(\lambda x_i + \tau)\right\}^2 - \frac{1}{2\lambda^2}\sum_{i=1}^n \{\ln(\lambda x_i + \tau)\}^2 - \sum_{i=1}^n \ln(\lambda x_i + \tau).\end{aligned}$$

Only the first term depends on s in the right-hand side of the above equation. And this term has the maximum value 0 when $s = (1/n)\sum_{i=1}^n \ln(\lambda x_i + \tau)$. Hence it suffices to maximize the sum of the third, the fourth and the fifth terms in the equation. Expressing the sum by $F(\lambda,\tau)$, let us deal with it:

$$F(\lambda,\tau) \stackrel{\text{def}}{=} \frac{1}{2n\lambda^2}\left\{\sum_{i=1}^{n}\ln(\lambda x_i+\tau)\right\}^2 - \frac{1}{2\lambda^2}\sum_{i=1}^{n}\{\ln(\lambda x_i+\tau)\}^2 - \sum_{i=1}^{n}\ln(\lambda x_i+\tau). \tag{2.1}$$

2.1 Case of $\lambda > 0$

First, we observe F around the boundary of the domain of definition.

The anti-logarithm condition implies $\lambda x_n+\tau>0$. When λ is fixed, we have $F(\lambda,\tau)\to -\infty$ as $\tau\to-\lambda x_n+0$. On the other hand, when λ diverges along with some paths, different situations appear. We can see it in the lemma below whose proof is in Appendix.

Lemma 2.1 We set $\tau = -\lambda x_n + 1/g(\lambda)$, where $g(\lambda)>0$ and $g(\lambda)\to\infty$ $(\lambda\to\infty)$. As $\lambda\to\infty$,

i) $F(\lambda,\tau)\to-\infty$ if $(g(\lambda))^{-1} = o(\mathrm{e}^{-\lambda^\varepsilon})$ for $\varepsilon>2$,

ii) $F(\lambda,\tau)\to\infty$ if $g(\lambda) = C\mathrm{e}^{\lambda^\varepsilon+o(\lambda^\varepsilon)}$ for $C>0$ and $0<\varepsilon\le 2$,

iii) $F(\lambda,\tau)\to\infty$ if $g(\lambda)\sim C\lambda^\varepsilon$ for $C>0$ and $\varepsilon>n-1$,

iv) $F(\lambda,\tau)\to-\infty$ if $g(\lambda)=o(\lambda^{n-1})$.

■

On the other hand, as $\tau\to\infty$, $F(\lambda,\tau)\to-\infty$ if λ is finite.

The analysis of F around $\lambda=0$ is as follows. Rewriting (2.1) into

$$F(\lambda,\tau) = \frac{-1}{2n\lambda^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\{\ln(\lambda x_i+\tau)-\ln(\lambda x_j+\tau)\}^2 - \sum_{i=1}^{n}\ln(\lambda x_i+\tau), \tag{2.2}$$

we obtain

$$\lim_{\lambda\to+0}F(\lambda,\tau) = \frac{-1}{2n}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left(\frac{x_i-x_j}{\tau}\right)^2 - \sum_{i=1}^{n}\ln\tau, \tag{2.3}$$

which achieves the relative maximum $-n/2 - n\ln\tau^*$ when

$$\tau=\tau^* \stackrel{\text{def}}{=} \frac{1}{n}\sqrt{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(x_i-x_j)^2}.$$

In order to narrow the search region for PRMs, we next look into the search interval in the τ direction corresponding to some special cases of values of λ.

From (2.2)

$$\frac{\partial F}{\partial\tau}(\lambda,\tau) \le -\frac{n-1}{2\lambda^2(\lambda x_1+\tau)}\left\{\left(1-\frac{\lambda x_1+\tau}{\lambda x_n+\tau}\right)\ln\frac{\lambda x_1+\tau}{\lambda x_n+\tau} + \frac{2n\lambda^2}{n-1}\right\}.$$

Since the term outside brace is negative, we devote our attention to the variation of the sign of

$$c^+(\lambda,\tau) \stackrel{\text{def}}{=} \left(1-\frac{\lambda x_1+\tau}{\lambda x_n+\tau}\right)\ln\frac{\lambda x_1+\tau}{\lambda x_n+\tau} + \frac{2n\lambda^2}{n-1}. \tag{2.4}$$

This increases monotonously as τ increases. Hence, if there exists a τ_{c0}^{+} such that $c^{+}(\lambda,\tau_{c0}^{+}) = 0$, it suffices to search in the interval $(-\lambda x_n, \tau_{c0}^{+})$. For instance, setting $c^{+}(\lambda,0)=0$ in the case that $x_n > 0$, we obtain

$$\lambda = \lambda^{*} \stackrel{\mathrm{def}}{=} \sqrt{-\frac{n-1}{2n}\left(1-\frac{x_1}{x_n}\right)\ln\frac{x_1}{x_n}}.$$

Then, the interval is included in $(-\lambda x_n, 0)$ for λ larger than λ^{*}. To investigate the search interval $(-\lambda x_n, \tau_{c0}^{+})$, let us substitute

$$X \stackrel{\mathrm{def}}{=} \frac{\lambda x_1 + \tau}{\lambda x_n + \tau}$$

into the right-hand side of (2.4) and set

$$C^{+}(\lambda, X) \stackrel{\mathrm{def}}{=} (1-X)\ln X + \frac{2n\lambda^2}{n-1} \qquad (X > 1). \tag{2.5}$$

In addition, denote by X_{C0}^{+} the solution of $C^{+}(\lambda,X)=0$ for given n and λ. Since the first term in the right-hand side of (2.5) is a monotonously decreasing function of X, X_{C0}^{+} is necessarily large if λ is large. This implies that τ_{c0}^{+} should be close to $-\lambda x_n$ because

$$X_{C0}^{+} = 1 + \frac{\lambda(x_1 - x_n)}{\lambda x_n + \tau_{c0}^{+}}. \tag{2.6}$$

Therefore, as λ becomes much larger, the search interval $(-\lambda x_n, \tau_{c0}^{+})$ becomes narrower, and it will be included in the interval where $F(\lambda,\tau)$ is of singularity as stated in Lemma 2.1.

Incidentally, τ_{c0}^{+} approaches $-\lambda x_n$ not quickly but slowly. From (2.6) and the definition of X_{C0}^{+}, we can see that the distance between τ_c^{+} and $-\lambda x_n$ is given as

$$d^{+} = (x_1 - x_n)\sqrt{\frac{(n-1)\ln X_{C0}^{+}}{2n(X_{C0}^{+}-1)}}.$$

For instance, when $X_{C0}^{+} = 290$,

$$d^{+} \approx \sqrt{\frac{\ln X_{C0}^{+}}{2(X_{C0}^{+}-1)}} \approx 0.1, \qquad \lambda \approx 29.$$

Next, let us consider the opposite case, where λ is small. Then, τ_{c0}^{+} might be very large from (2.4). Suppose that λ is so small that the following relations hold:

$$\lambda \max(|x_1|,|x_n|) \ll \tau^{*} \text{ and } \frac{\lambda(x_1-x_n)}{\lambda x_n + \tau^{*}} \ll 1. \tag{2.7}$$

Then, because that

$$\begin{aligned}
\ln\frac{\lambda x_i + \tau}{\lambda x_j + \tau} &= \ln\left\{1 + \frac{\lambda(x_i - x_j)}{\lambda x_j + \tau}\right\} \\
&= \frac{\lambda(x_i - x_j)}{\lambda x_j + \tau} - \frac{1}{2\left\{1+\theta\dfrac{\lambda(x_i-x_j)}{\lambda x_j+\tau}\right\}^2}\left\{\frac{\lambda(x_i - x_j)}{\lambda x_j + \tau}\right\}^2 \quad (\text{here, } 0<\theta<1) \\
&\approx \frac{\lambda(x_i - x_j)}{\tau} \quad (\text{from (2.7)})
\end{aligned}$$

for $\tau \geq \tau^*$ and $j > i$, $F(\lambda, \tau)$ is approximately equal to the right-hand side of (2.3). Thus, since the point τ_0 satisfying $\frac{\partial F}{\partial \tau}(\lambda, \tau) = 0$ is around $\tau = \tau^*$, it would be sufficient to search the τ_0 just beyond τ^* in the τ-direction.

2.2 Case of $\lambda < 0$

First, we observe F around the boundary of the domain of definition.

The anti-logarithm condition implies $\lambda x_1 + \tau > 0$. When λ is fixed, we have $F(\lambda, \tau) \to -\infty$ as $\tau \to -\lambda x_1 + 0$. On the other hand, when λ goes to the infinity, the following lemma holds.

Lemma 2.2 We set $\tau = -\lambda x_1 + 1/g(\lambda)$, where $g(\lambda) > 0$ and $g(\lambda) \to \infty$ $(\lambda \to -\infty)$. As $\lambda \to -\infty$,

i) $F(\lambda, \tau) \to -\infty$ if $(g(\lambda))^{-1} = o(\mathrm{e}^{-|\lambda|^\varepsilon})$ for $\varepsilon > 2$,

ii) $F(\lambda, \tau) \to \infty$ if $g(\lambda) = C\mathrm{e}^{|\lambda|^\varepsilon + o(|\lambda|^\varepsilon)}$ for $C > 0$ and $0 < \varepsilon \leq 2$,

iii) $F(\lambda, \tau) \to \infty$ if $g(\lambda) \sim C|\lambda|^\varepsilon$ for $C > 0$ and $\varepsilon > n - 1$,

iv) $F(\lambda, \tau) \to -\infty$ if $g(\lambda) = o(|\lambda|^{n-1})$.

■

Proof. The proof is similar to that of Lemma 2.1. □

As $\tau \to \infty$, $F(\lambda, \tau) \to -\infty$ if λ is finite.

Since $\lim_{\lambda \to -0} F(\lambda, \tau)$ has the same expression as the right-hand side of (2.3), thus it achieves the relative maximum $-n/2 - n \ln \tau^*$ at $\tau = \tau^*$.

As in the case of $\lambda > 0$, we next look into the search interval in the τ direction.

From (2.2)

$$\frac{\partial F}{\partial \tau}(\lambda, \tau) \leq -\frac{n-1}{2\lambda^2(\lambda x_n + \tau)} \left\{ \left(1 - \frac{\lambda x_n + \tau}{\lambda x_1 + \tau}\right) \ln \frac{\lambda x_n + \tau}{\lambda x_1 + \tau} + \frac{2n\lambda^2}{n-1} \right\}.$$

We set

$$c^-(\lambda, \tau) \stackrel{\text{def}}{=} \left(1 - \frac{\lambda x_n + \tau}{\lambda x_1 + \tau}\right) \ln \frac{\lambda x_n + \tau}{\lambda x_1 + \tau} + \frac{2n\lambda^2}{n-1}. \tag{2.8}$$

If there exists a τ_{c0}^- such that $c^-(\lambda, \tau_{c0}^-) = 0$, it suffices to search in the interval $(-\lambda x_1, \tau_{c0}^-)$.

To investigate the search interval $(-\lambda x_1, \tau_{c0}^-)$, let us substitute

$$Y \stackrel{\text{def}}{=} \frac{\lambda x_n + \tau}{\lambda x_1 + \tau}$$

into the right-hand side of (2.8) and set

$$C^-(\lambda, Y) \stackrel{\text{def}}{=} (1 - Y) \ln Y + \frac{2n\lambda^2}{n-1} \qquad (Y > 1).$$

In addition, denote by Y_{C0}^- the solution of $C^-(\lambda, Y) = 0$ for given n and λ. The equation above has the same form as that of (2.5). Consequently, as $|\lambda|$ becomes much larger, the search interval $(-\lambda x_1, \tau_{c0}^-)$ becomes narrower, and it will be included in the interval where $F(\lambda, \tau)$ is of singularity as stated in Lemma 2.2. Incidently, τ_{c0}^- approaches $-\lambda x_1$ not quickly but slowly.

On the other hand, suppose that $|\lambda|$ is so small that the following relations hold:

$$|\lambda| \max(|x_1|, |x_n|) \ll \tau^* \text{ and } \left| \frac{\lambda(x_1 - x_n)}{\lambda x_1 + \tau^*} \right| \ll 1.$$

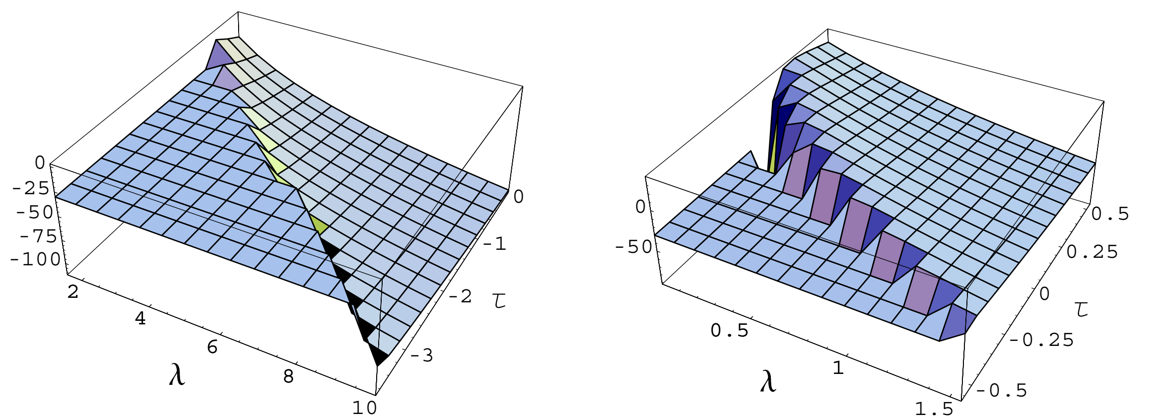
Then, since the point τ_0 satisfying $\frac{\partial F}{\partial \tau}(\lambda, \tau) = 0$ is around $\tau = \tau^*$, it would be sufficient to search the τ_0 just beyond τ^* in the τ-direction.

3 Computational experience

3.1 Searching examples with graphs

In this subsection we show an example, where we select an initial guess by employing a graphic tool. The data from Smith and Naylor (12) are used here. The functions $F(\lambda, \tau)$ for $\lambda > 0$ and $\lambda < 0$ are displayed on Fig. 1 and Fig. 2, respectively. On both figures, the right graphs are the magnifications of the left ones around $\lambda = 0$. The left graph on Fig. 2 shows $F(\lambda, \tau)$ up to 9.98 in the τ direction, which is the value of τ_{c0}^- for $\lambda = -6$. In addition, we have set $F(\lambda, \tau) \stackrel{\text{def}}{=} -30$ at (λ, τ) not satisfying the anti-logarithm condition.

These graphs imply that the PRM is near $(0, \tau^*)$. So, carrying out Newton iteration from the initial guess $(-0.001, 0.27)$ since $\tau^* \simeq 0.27$, we get $(\lambda, \tau) = (-0.295268, 0.598082)$. On the basis of this, $(\lambda, \mu, \sigma) = (-0.295268, 1.16810, 0.253179)$ is obtained.

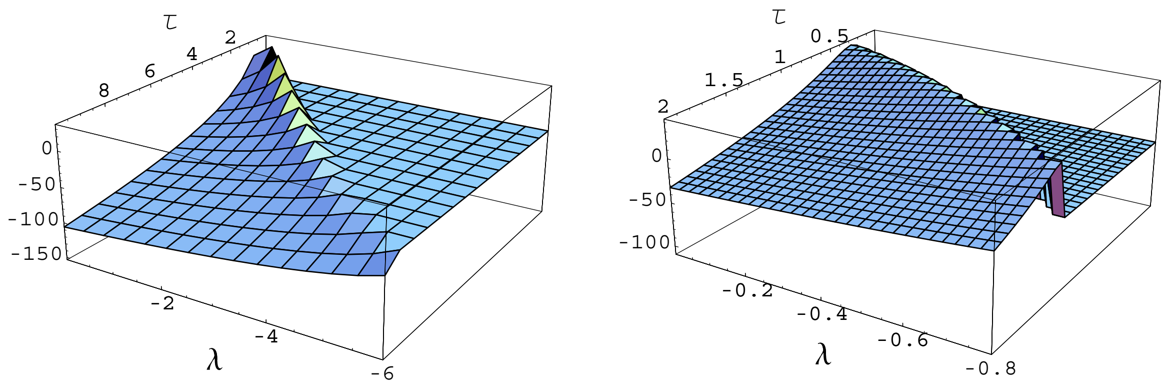


$(\lambda, \tau) \in [\lambda^*, 10] \times [-10x_n + 0.001, 0]$ $\quad$ $(\lambda, \tau) \in [0.001, \lambda^*] \times [-\lambda^* x_n + 0.001, 2\tau^*]$

Figure 1: $F(\lambda, \tau)$ for $\lambda > 0$

3.2 Monte Carlo studies

By means of Monte Carlo simulation we confirm that, in the same way as Subsection 3.1, we can select initial guesses such that the sequence of the approximations generated by an iterative method converges in more general data. In addition, we numerically evaluate the biases and RMSEs of the estimators for the extended lognormal distribution in the case that the estimation is carried out through the proposed reparameterization.



$(\lambda, \tau) \in [-6, 0.001] \times [0.001x_1 + 0.001, 9.98]$ $(\lambda, \tau) \in [-0.8, -0.001] \times [0.001x_1 + 0.001, 2]$

Figure 2: $F(\lambda, \tau)$ for $\lambda < 0$

3.2.1 Simulation conditions

The sample number n is set at 100, 50, 20 or 10. The parameter λ is set at 0.01, 0.2, 0.4, 0.6 or 0.8, whereas the other parameters σ and μ are fixed at 1 and 0, respectively. For each combination of values of n and λ, 1000 independent pseudo-random samples are considered.

To imitate the process that we display graphs and select an initial guess, we adopt the following strategy: 1) scatter many mesh points on the λ-τ plane; 2) choose as an initial guess the point at which $F(\lambda, \tau)$ has the largest value among the values that $F(\lambda, \tau)$ have on mesh points. We put mesh points as follows:

i) In the λ direction.
The interval $(0, 10]$ and the interval $[-6, 0)$ are divided into 100 equal parts by the mesh points.

ii) In the τ direction.
Two intervals are set corresponding to the values (10 or -6) of the end points in the λ coordinate. And both intervals are divided into 200 equal parts by the mesh points.

3.2.2 Simulation results

We obtained the result in Tab. 1 by using Newton's method as the iterative method. It indicates the numbers of cases, say m, when PRMs have been gotten and the numbers (showed in parentheses) of cases when the λ is positively estimated. From this we can see that the iterative method is not always given the initial guesses such that the sequence of the approximations converges. Although the result for m is worse than that for "n_{exist} in Tab. 4 in (11)", which is the number of successful cases for the three-parameter estimation with the continuation method, the difference is small. Thus, at worst the estimation as in Section 3.1 can be expected to work better than the three-parameter estimation with other competitors in the table. In the next section we will once again deal with the data that the convergent sequence was not obtained here.

Next, we investigate the influence of the two-stage process on the RMSEs and biases of the estimator of (λ, μ, σ), in which we estimate (λ, τ) and then evaluate (λ, μ, σ) on the base of the estimate. The evaluated RMSEs and biases are showed in Tab. 2 and Tab. 3. The comparisons between these tables and the results (Tab. 5, 6) in the article indicates that our reparameterization does not influence the RMSEs and biases of (λ, μ, σ).

Table 1: Successful numbers in finding PRMs

		n			
		100	50	20	10
	0.01	1000 (527)	1000 (529)	1000 (507)	972 (478)
	0.2	1000 (986)	1000 (939)	1000 (795)	968 (656)
λ	0.4	1000 (1000)	1000 (997)	999 (947)	958 (781)
	0.6	1000 (1000)	1000 (1000)	998 (991)	924 (823)
	0.8	1000 (1000)	1000 (1000)	996 (993)	884 (839)

Table 2: Evaluated RMSEs for (λ, μ, σ)

		n			
		100	50	20	10
	0.01	(.090, .111, .069)	(.137, .155, .103)	(.258, .248, .176)	(.529, .373, .262)
	0.2	(.090, .111, .071)	(.139, .155, .107)	(.261, .247, .181)	(.530, .370, .266)
λ	0.4	(.092, .111, .077)	(.146, .154, .117)	(.271, .246, .195)	(.544, .369, .283)
	0.6	(.096, .111, .088)	(.155, .153, .132)	(.289, .245, .218)	(.539, .366, .308)
	0.8	(.102, .110, .102)	(.167, .151, .150)	(.313, .243, .247)	(.547, .359, .346)

Table 3: Evaluated biases for (λ, μ, σ)

		n			
		100	50	20	10
	0.01	$(-.04, .15, -.04)$	$(-.02, .14, -.19)$	$(-.07, .18, -.63)$	$(-.31, .26, -1.40)$
	0.2	$(.00, .14, -.02)$	$(.06, .12, -.17)$	$(.10, .12, -.60)$	$(-.01, .19, -1.36)$
λ	0.4	$(.05, .13, .01)$	$(.14, .09, -.13)$	$(.28, .05, -.55)$	$(.23, .12, -1.31)$
	0.6	$(.09, .12, .05)$	$(.23, .07, -.09)$	$(.46, -.00, -.48)$	$(.30, .15, -1.20)$
	0.8	$(.14, .11, .08)$	$(.31, .05, -.04)$	$(.67, -.05, -.40)$	$(.27, .20, -1.07)$

Here, evaluated biases $\times$ 10 are showed for saving space.

3.3 Difficult examples

Concerned with finding PRMs, the estimation results in Subsection 3.2 were slightly inferior to that by the continuation method in the article. In some of the data sets judged as "not converged" in the subsection, however, we might have obtained the PRMs successfully if we had actually displayed F's and had given initial guesses sufficiently close to the PRMs. Now, we introduce such a data example. And we show that the selection of an initial guess is difficult in case that we try to solve the three-parameter estimation problem with regard to the data by the continuation method.

First, we introduce the continuation method. Assume that λ, μ and σ are the functions of t, and set that $\boldsymbol{\theta}(t) \stackrel{\text{def}}{=} (\lambda(t),\ \mu(t),\ \sigma(t))^T$,

$$\boldsymbol{M}(\boldsymbol{\theta}(t)) \stackrel{\text{def}}{=} \left(\ \frac{\partial \ln \tilde{L}}{\partial \lambda}(\lambda(t),\mu(t),\sigma(t)), \quad \frac{\partial \ln \tilde{L}}{\partial \mu}(\lambda(t),\mu(t),\sigma(t)), \quad \frac{\partial \ln \tilde{L}}{\partial \sigma}(\lambda(t),\mu(t),\sigma(t)) \ \right)^T$$

and

$$\boldsymbol{h}(t,\boldsymbol{\theta}(t)) \stackrel{\text{def}}{=} t\boldsymbol{M}(\boldsymbol{\theta}(t)) + (1-t)\left\{\boldsymbol{M}(\boldsymbol{\theta}(t)) - \boldsymbol{M}(\boldsymbol{\theta}(0))\right\}.$$

If the curve $\{\boldsymbol{\theta}(t) : 0 \le t \le 1 \,|\boldsymbol{h}(t,\boldsymbol{\theta}(t)) = \boldsymbol{0}\}$ is continuous, its tracing starting from $\boldsymbol{\theta}(0)$ can lead to finding a stationary point $\boldsymbol{\theta}(1)$ of $\ln \tilde{L}(\lambda,\mu,\sigma)$. Assuming that $\boldsymbol{h}(t,\boldsymbol{\theta}(t))$ is continuously differentiable, we numerically solve

$$\frac{\mathrm{d}\boldsymbol{\theta}(t)}{\mathrm{d}t} = -\boldsymbol{J}^{-1}(\boldsymbol{\theta}(t))\boldsymbol{M}(\boldsymbol{\theta}(0)) \qquad (0 \le t \le 1), \tag{3.1}$$

which is obtained by differentiating $\boldsymbol{h}(t,\boldsymbol{\theta}(t)) = \boldsymbol{0}$ with respect to t. Here,

$$\boldsymbol{J}(\boldsymbol{\theta}(t)) \stackrel{\text{def}}{=} \begin{bmatrix} \frac{\partial^2 \ln \tilde{L}}{\partial \lambda^2}(\lambda(t),\mu(t),\sigma(t)) & \frac{\partial^2 \ln \tilde{L}}{\partial \mu \partial \lambda}(\lambda(t),\mu(t),\sigma(t)) & \frac{\partial^2 \ln \tilde{L}}{\partial \sigma \partial \lambda}(\lambda(t),\mu(t),\sigma(t)) \\ \frac{\partial^2 \ln \tilde{L}}{\partial \lambda \partial \mu}(\lambda(t),\mu(t),\sigma(t)) & \frac{\partial^2 \ln \tilde{L}}{\partial \mu^2}(\lambda(t),\mu(t),\sigma(t)) & \frac{\partial^2 \ln \tilde{L}}{\partial \sigma \partial \mu}(\lambda(t),\mu(t),\sigma(t)) \\ \frac{\partial^2 \ln \tilde{L}}{\partial \lambda \partial \sigma}(\lambda(t),\mu(t),\sigma(t)) & \frac{\partial^2 \ln \tilde{L}}{\partial \mu \partial \sigma}(\lambda(t),\mu(t),\sigma(t)) & \frac{\partial^2 \ln \tilde{L}}{\partial \sigma^2}(\lambda(t),\mu(t),\sigma(t)) \end{bmatrix}$$

and it is supposed that $|\boldsymbol{J}(\boldsymbol{\theta}(t))| \ne 0$ $(0 \le t \le 1)$. Set that $\boldsymbol{\theta}_0 \stackrel{\text{def}}{=} \boldsymbol{\theta}(0)$ and $\Delta t \stackrel{\text{def}}{=} 1/N$ (N is a natural number). Then, the application of Euler method to (3.1) yields

$$\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i - \Delta t \boldsymbol{J}^{-1}(\boldsymbol{\theta}_i)\boldsymbol{M}(\boldsymbol{\theta}_0) \qquad (i = 0, \ldots, N-1).$$

$\boldsymbol{\theta}_i$ stands for the approximation of $\boldsymbol{\theta}(i\Delta t)$. $\boldsymbol{\theta}_N$ may be obtained by this recurrence formula.

If $\ln \tilde{L}(\lambda,\mu,\sigma)$ has a maximal value at $(\lambda,\mu,\sigma) = \boldsymbol{\theta}(1)$, it is necessary that $|\boldsymbol{J}(\boldsymbol{\theta}(t))| < 0$ $(0 \le t \le 1)$ since $|\boldsymbol{J}(\boldsymbol{\theta}(1))| < 0$. Thus, we have to select an initial guess $\boldsymbol{\theta}_0$ satisfying $|\boldsymbol{J}(\boldsymbol{\theta}_0)| < 0$.

We give two examples, which show the regions where $|\boldsymbol{J}((\lambda,\mu,\sigma)^T)| < 0$ is unsatisfied. Fig. 3 and Fig. 4, respectively, represent the case from Smith and Naylor and the case of Tab. 4. On both figures, the black points in the centers stand for the PRMs, whose coordinates are $(-0.295268, 1.16810, 0.253179)$ and $(2.21152, -0.505806, 0.903114)$, respectively. The colored parts stand for the region where the inequality is unsatisfied. Incidentally, the data in Tab. 4 are those given from one of the cases that were judged as "not converged" in the simulation when $n = 20$ and $\lambda = 0.4$ in Subsection 3.2. And we obtained the PRM with respect to the data by drawing the graphs of $F(\lambda,\tau)$ and selecting a more proper initial guess.

Table 4: failed data from Subsection 3.2

2.325620	−0.836045	5.663170	−0.905886	4.333967
−0.912527	−0.382619	0.326860	0.030242	−0.319501
−0.116709	0.583490	−0.745602	0.591303	−0.564003
0.214998	−0.665881	0.696181	−0.909085	0.630385

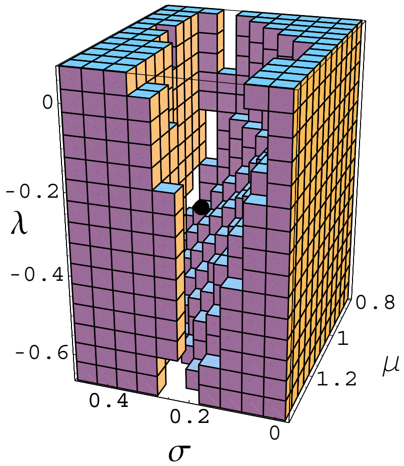


Figure 3: The region where $|\boldsymbol{J}((\lambda,\mu,\sigma)^T)| < 0$ is unsatisfied (Smith & Naylor data)

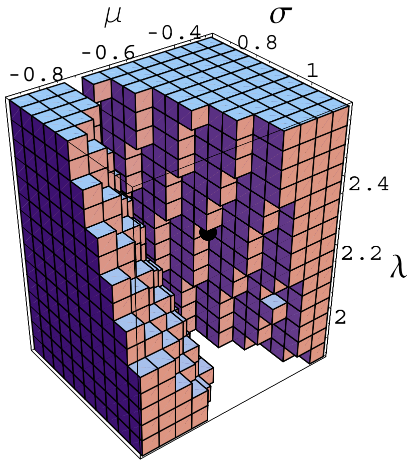


Figure 4: The region where $|\boldsymbol{J}((\lambda,\mu,\sigma)^T)| < 0$ is unsatisfied (the data in Tab. 4)

Noting that the embedded distribution is the normal distribution, let us consider the selection of the initial guess by $(\lambda,\mu,\sigma) = (0.005, \bar{\mu}, \{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{\mu})^2\}^{1/2})$, where $\bar{\mu} \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} x_i$. Then, the initial guesses with regard to data on Fig. 3 and Fig. 4 are $(0.005, 1.13, 0.268)$ and $(0.005, 0.452, 1.71)$, respectively. In cases like as in Fig. 3, the homotopy curves are probably continuous from the starting points (initial guesses) to the end points (PRMs) since the colorless regions are sufficiently wide for the curves to pass. Thus, the combination of such a selection of the initial guess and the continuation method would work well in the cases. However, it is not the case like as Fig. 4, for there the PRM is far from the initial guess but is located nearby as well as surrounded by the boundary of the region that may cut off the homotopy curve. In fact, the initial guess on Fig. 4 does not satisfy even $|\boldsymbol{J}((\lambda,\mu,\sigma)^T)| < 0$. Besides, it should be noted that the continuation method may not successfully give PRMs even if a point is selected as the initial guess from the colorless region. The data on Fig. 4 is also such a case, and the colorless region includes many points such that the continuation method does not give the PRM when those points are selected as the initial guesses.

4 Summary and remarks

We proposed a computing method for PRMs on the base of the extended lognormal distribution. In the method we change the three-parameter estimation problems to the two-parameter estimation problems, select initial guesses with the help of graphic tools, and find PRMs by iterative methods. The features are as follows:

i) The singular region associated with the non-regularity of the likelihood function in the lognormal distribution is in the region where both new parameters simultaneously go to ∞ or $-\infty$. Hence, the difficulty of the estimation caused by the non-regularity is removed.

ii) By displaying the function defined on the two-parameter plane, the selection of the initial guess becomes easier. This makes it possible to cope with data cases such that the estimation by the other computing methods is difficult.

iii) The indirect estimation through the proposed reparameterization does not influence the RMSEs and biases of the estimators in the extended lognormal distribution.

We introduced the continuation method and used $|\boldsymbol{J}(\boldsymbol{\theta}(t))| < 0$ $(0 \leq t \leq 1)$ to get the condition $|\boldsymbol{J}(\boldsymbol{\theta}_0)| < 0$. Incidentally, there is the more sophisticated continuation method except the method introduced in the present article, and it does not always impose $|\boldsymbol{J}(\boldsymbol{\theta}(t))| < 0$ $(0 < t < 1)$ on the paths that lead to PRMs. Even the sophisticated method, however, requires the condition $|\boldsymbol{J}(\boldsymbol{\theta}_0)| < 0$. A detailed explanation for this is given in (13, p. 48).

Acknowledgements

We would like to thank Professor T. Mitsui for improving the presentation of this manuscript.

5 Bibliography

(1) Hill, B. M. The Three-parameter Lognormal Distribution and Bayesian Analysis of a Point-source Epidemic. J. Amer. Statist. Assoc., **1963**, *58*, 72–84.

(2) Johnson, N. L.; Kotz, S.; Balakrishnan, N. *Continuous Univariate Distributions*, John Wiley & Sons, New York, **1994**.

(3) Wingo, D. R. The Use of Interior Penalty Functions to Overcome Lognormal Distribution Parameter Estimation Anomalies. J. Statist. Comput. Simulation, **1975**, *4*, 49–61.

(4) Wingo, D. R. Moving Truncations Barrier-function Methods for Estimation in Three-parameter Lognormal Models. Comm. Statist.–Simulation Comput., **1976**, *B5*, 65–80.

(5) Wingo, D. R. Fitting Three-parameter Lognormal Models by Numerical Global Optimization–an Improved Algorithm. Comput. Statist. Data Anal., **1984**, *2*, 13–25.

(6) Lambert, J. A. Estimation of Parameters in the Three Parameter Lognormal Distribution. Austral. J. Statist., **1964**, *6*, 29–32.

(7) Munro, A. H.; Wixley, R. A. J. Estimation on Order Statistics of Small Samples from a Three-parameter Lognormal Distribution. J. Amer. Statist. Assoc., **1970**, *65*, 212–225.

(8) Eastham, J. F.; LaRiccia, V. N.; Schuenemeyer, J. H. Small Sample Properties of the Maximum Likelihood Estimators for an Alternative Parameterization of the Three-parameter Lognormal Distribution. Comm. Statist.–Simulation Comput., **1987**, *16*, 871–884.

(9) Giesbrecht, F.; Kempthorne, O. Maximum Likelihood Estimation in the Three-parameter Lognormal Distribution. J. Royal Statist. Soc., Ser. B, **1976**, *38*, 257–264.

(10) Cheng, R. C. H.; Iles, T. C. Embedded Models in Three-parameter Distributions and Their Estimation. J. Royal Statist. Soc., Ser. B, **1990**, *52*, 135–149.

(11) Hirose, H. Maximum Likelihood Parameter Estimation in the Three-parameter Log-normal Distribution Using the Continuation Method. Comput. Statist. Data Anal., **1997**, *24*, 139–152.

(12) Smith, R. L.; Naylor, J. C. A Comparison of Maximum Likelihood and Bayesian Estimators for the Three-parameter Weibull Distribution. Appl. Statist., **1987**, *36*, 358–369.

(13) Iri, M. *Suuti Keisan (in Japanese)*, Asakura-Shoten, Tokyo, **1981**.

Appendix

Proof of Lemma 2.1: Substituting $\tau = -\lambda x_n + 1/g(\lambda)$ into (2.1) and rewriting the result by

$$\ln(\lambda(x_i - x_n) + 1/g(\lambda)) \sim \ln\lambda \quad (\lambda \to \infty)$$

for $1 \le i \le n-1$, we get

$$F(\lambda, -\lambda x_n + 1/g(\lambda)) \sim \frac{1-n}{2n}\left(\frac{\ln g(\lambda)}{\lambda}\right)^2 + \ln g(\lambda) - (n-1)\ln\lambda \quad (\lambda \to \infty). \qquad (A.1)$$

First of all, let us consider the case i). Since $e^{\lambda^\varepsilon} \le g(x)$ for a sufficiently large λ,

$$0 < \frac{\ln\lambda}{\ln g(\lambda)} \le \frac{\ln\lambda}{\lambda^\varepsilon} \to 0 \quad (\lambda \to \infty), \qquad \frac{\ln g(\lambda)}{\lambda^2} \ge \frac{\lambda^\varepsilon}{\lambda^2} \to \infty \quad (\lambda \to \infty).$$

From these and (A.1), we can obtain the conclusion.

In the case ii), (A.1) becomes

$$F(\lambda, -\lambda x_n + 1/g(\lambda)) \sim \frac{1-n}{2n}\lambda^{2\varepsilon-2} + \lambda^\varepsilon \quad (\lambda \to \infty)$$

since $\ln g(\lambda) \sim \lambda^\varepsilon$ $(\lambda \to \infty)$. From the above, we can get the conclusion.

In the case iii), (A.1) becomes

$$F(\lambda, -\lambda x_n + 1/g(\lambda)) \sim \varepsilon\ln\lambda - (n-1)\ln\lambda \quad (\lambda \to \infty)$$

since $\ln g(\lambda) \sim \varepsilon\ln\lambda$ $(\lambda \to \infty)$. Thus we get the conclusion.

In the case iv), (A.1) becomes

$$F(\lambda, -\lambda x_n + 1/g(\lambda)) \sim \ln g(\lambda) - (n-1)\ln\lambda \quad (\lambda \to \infty)$$

since $\ln g(\lambda) = O(\ln\lambda)$ $(\lambda \to \infty)$. Here,

$$\ln g(\lambda) - (n-1)\ln\lambda = \ln\frac{g(\lambda)}{\lambda^{n-1}} \to -\infty \quad (\lambda \to \infty).$$

Consequently, our proof has completed.