On the consistency of the maximum likelihood estimator for the three parameter lognormal distribution

HaiYing Wang

Department of Mathematics & Statistics, University of New Hampshire, Durham, New Hampshire 03824, USA

Nancy Flournoy

Department of Statistics, University of Missouri, Columbia, Missouri 65211, USA

Abstract

The three parameter log-normal distribution is a popular non-regular model, but surprisingly, whether the local maximum likelihood estimator (MLE) for parameter estimation is consistent or not has been speculated about since the 1960s. This note gives a rigorous proof for the existence of a consistent MLE for the three parameter log-normal distribution, which solves a problem that has been recognized and unsolved for 50 years. Our results also imply a uniform local asymptotic normality condition for the three parameter log-normal distribution. In addition, we give results on the asymptotic normality and the uniqueness of the local MLE.

Keywords: Consistency, Local Maximum, Maximum Likelihood, Non-regular model, Uniform Local Asymptotic Normality

1. Introduction

A random variable Y has the three parameter log-normal distribution if

$$Z = \log(Y - A) \sim N(\mu, \sigma^2), \tag{1}$$

in which μ, σ and A are unknown parameters. This model has been widely used in applications and its estimation approaches have been studied by many, including Cohen (1951), Hill (1963), Harter and Moore (1966), Munro and Wixley (1970), Giesbrecht and Kempthorne (1976), Cohen and Whitten (1980), Crow and Shimizu (1998) and Basak *et al.* (2009), among others. But the theoretical properties of the proposed methods were not fully addressed rigorously in these papers.

Cohen (1951) first considered using the method of maximum likelihood to estimate parameters in Model (1). He derived the maximum likelihood equations and the Fisher information matrix without noting

Email address: `HaiYing.Wang@unh.edu` (HaiYing Wang)

that the likelihood function was unbounded and that the solutions to the likelihood equations were not the global maximizer of the likelihood function. Aitchison and Brown (1957) gave a comprehensive summary of estimation methods developed up to that time, including the method of maximum likelihood. Hill (1963) pointed out that the likelihood function for Model (1) is unbounded and derived a sequence of parameter values along which the likelihood went to infinity. He then suggested using a Bayesian approach for parameter estimation. A strong restriction on his proposed method is that the priors used must "assign negligible prior probability in the vicinity of the singularity". Harter and Moore (1966) then suggested using the solution to the likelihood equation instead of using the global maximizer to estimate the unknown parameters and termed this *local MLE*. They also considered fitting the three parameter log-normal model from censored data. The likelihood equations were derived and a modified iterative procedure was proposed to find the estimates numerically. However, whether the local MLE is consistent or not was not discussed and this problem has been remained unsolved since then. Cohen and Whitten (1980) proposed modifying the local MLE by using the extreme order statistics to estimate the boundary parameter A. Giesbrecht and Kempthorne (1976) studied the three parameter log-normal distribution by assuming that data observed were subject to a grouping error. They defined the likelihood function in this setting for which an explicit expression was not available.

Cheng and Amin (1979) proposed the *maximum product spacings estimation* method and proved that it produced consistent estimators for the three parameter log-normal model. The maximum product spacings estimation method uses the cumulative distribution function for construction, so the target functions are always bounded. But there is no closed form for the cumulative distribution function for the three parameter log-normal distribution so the computation burden is heavy when finding the numerical solution. Cheng and Amin (1983) derived the asymptotic normality of the maximum product spacings estimator and the local MLE for the log-normal distribution. However, the normality of the local MLE was obtained based on the assumption of the existence of a consistent MLE and a rigorous proof for the consistency of the local MLE was not provided. For a class of non-regular models, Smith (1985) derived the asymptotic properties of the local MLE's. However, as shown in Section 2, a key requirement in Smith (1985)'s proof is not met for Model (1).

In this note, we conquer the obstacle in proving the consistency of the local MLE for Model (1). We also study the uniqueness of the local maximizer of the likelihood function and a theorem similar to Theorem 2 in Smith (1985) is provided to help choose a consistent solution when multiple solutions exist. The asymptotic normality of the local MLE is proved under contiguous alternatives. We briefly review the literature about the MLE in non-regular models in Section 2 and present our main results in Section 3. Proofs are given in Section 4.

2. Maximum likelihood estimation in non-regular models

When some of the classical regularity conditions required in Cramér (1946) and Wald (1949) are not true, examples can be found in which desirable results of MLE's fail (e.g., Le Cam, 1990). Such situations are often termed *non-regular*. With different violations of regularity conditions, there are different types of non-regular problems. Model (1) has parameter-dependent support and unbounded likelihood so we only include results that are related to this type of non-regularity. Other types of non-regularity and related literature can be found in Smith (1989), Cheng and Traylor (1995) and the references therein. There is a large body of literature rigorously developing likelihood based inference methods for models with parameter-dependent support, including Woodroofe (1972), Weiss and Wolfowitz (1973), Woodroofe (1974), Hall (1982), Cheng and Amin (1981), Smith (1985), Cheng and Iles (1987), Smith (1994) and Hall and Wang (2005). But the consistency of the MLE for Model (1) is not covered in the literature.

LeCam (1970) pointed out that the classical conditions used for the asymptotic normality of the MLE's are too strong, especially the requirement of two or three derivatives of the likelihood function. Interestingly, this requirement can often be weakened to involve only the first derivatives, which is called *differentiable in quadratic mean* (Le Cam, 1960). Model (1) belongs to the quadratic mean differentiable family, but this does not assure the existence of a consistent MLE (page 506 of Lehmann and Romano, 2005). Woodroofe (1972) studied the properties of the MLE's for the following density.

$$f_A(y) = f(y - A), \quad A, y \in \mathbb{R}, \tag{2}$$

where $f(y)$ is uniformly continuous and equals 0 for $y \in (-\infty, 0]$. It is also required that $\lim_{y\downarrow 0} f'(y) \in (0, \infty)$. Note that if $\log f(y)$ is integrable, then the classical conditions for consistency of the MLE's in Wald (1949) still hold, but the classical conditions for asymptotic normality fail since the support of f_A depends on A. Woodroofe (1972) showed that, with some extra assumptions on f, the MLE is still asymptotically normal with a convergence rate of $(n \log n)^{-1/2}$. Weiss and Wolfowitz (1973) worked further on Model (2) and proved that the MLE is asymptotically efficient for the single parameter A. Woodroofe (1974) worked on Model (2) with an alternative restriction on f that

$$f(y) \sim \alpha y^{\alpha-1} L(y) \qquad \text{as } y \downarrow 0,$$

where $\alpha \in (1, 2)$ and $L(y)$ varies slowly as $y \downarrow 0$ in the sense that $\lim_{y\downarrow 0} L(ay)/L(y) \in (0, \infty)$ for any $a > 0$. This class of functions f includes some commonly used densities such as the Gamma density and the Pareto density. For this class of models, the convergence rate of the MLE, say γ_n, is determined by

$$n\gamma_n^\alpha L(\gamma_n) \to 1.$$

Note that the convergence rate is $n^{-1/\alpha}$ if $\lim_{y\downarrow 0} L(y)$ is a positive constant. The asymptotic distributions of the MLE's are rather complex and were given in the form of complicated characteristic functions in

Woodroofe (1974). Hall and Wang (2005) derived a simple representation for the asymptotic distribution under the special case that $\lim_{y\downarrow 0} L(y)$ is a positive constant. The efficiency of estimators for this class of models is still an unsolved question.

Smith (1985) generalized the framework of Woodroofe (1974) to multivariate parameter models and extended the theory of local maximum likelihood estimation to a broad class of non-regular models without covariates by an elegant mathematical derivation. The models of interest in Smith (1985) take densities of the form:

$$f(y; A, \phi) = (y - A)^{\alpha-1} g(y - A; \phi), \quad A < y < \infty, \tag{3}$$

where ϕ is a vector parameter and $g(x; \phi)$ tends to a positive constant as $x \downarrow 0$. Because of the additional unknown parameter ϕ, the likelihood function of (3) may be unbounded. So the MLE considered by Smith (1985) is a local maximum that satisfies the likelihood equations. Smith (1985) evaluated the asymptotic behavior of the MLE's for different true values of α. When $\alpha > 2$, the convergence rate of the MLE's of A and ϕ are both $n^{-1/2}$, and they converge jointly to a normal distribution with the asymptotic variance being the inverse of the Fisher information matrix; when $\alpha = 2$, the MLE's of A and ϕ are asymptotically independent and are also asymptotically normal, but their convergence rates are $(n \log n)^{-1/2}$ and $n^{-1/2}$, respectively; when $\alpha \in (1, 2)$, the MLE's of A and ϕ are still asymptotically independent but the former is not asymptotically normal anymore, and the convergence rates are $n^{-1/\alpha}$ and $n^{-1/2}$, respectively; when $\alpha < 1$, the MLE's do not exist or are inconsistent.

The requirement that $g(x; \phi)$ tends to a positive constant as $x \downarrow 0$ in Model (3) is very crucial. It guarantees that there exist a constant sequence, say c_n, such that $c_n\{Y_{(1)} - A\}$ converges weakly to a non-degenerated distribution, where $Y_{(1)}$ is the smallest order statistic. Smith 1985's derivation relies heavily on this fact. But for Model (1), from the result in Lemma 1 of Section 4, $\exp(-\mu_0 + \sigma_0 r_n)\{Y_{(1)} - A_0\} \xrightarrow{d} 1$, and there does not exist a constant sequence $c_n^* \to \infty$ such that $c_n^*\{Y_{(1)} - A_0\}$ converges to a non-degenerate distribution. So Smith (1985)'s technique cannot be applied directly for the three parameter log-normal distribution to prove the consistency of the MLE.

3. Main results

Now we address the consistency of the MLE for the three parameter log-normal Model (1). Interestingly, although the likelihood function is unbounded, the Fisher information matrix of Model (1) exists and is finite, as given in Theorem 1 below. Denote $\boldsymbol{\theta} = (\mu, \sigma, A)^{\mathrm{T}}$. For a independent and identically distributed random sample $\{Y_1, Y_2, \ldots, Y_n\}$ of size n from Model (1), the log-likelihood function is

$$\ell_n(\boldsymbol{\theta}) = I(A < Y_{(1)}) \left[-\frac{n}{2} \log(2\pi) - n \log(\sigma) - \sum_{i=1}^{n} \log(Y_i - A) - \frac{\sum_{i=1}^{n} \{\log(Y_i - A) - \mu\}^2}{2\sigma^2} \right]. \tag{4}$$

The MLE are defined as local maximizers that satisfy the following likelihood equations.

$$
\begin{aligned}
\frac{\partial \ell_n(\boldsymbol{\theta})}{\partial \mu} &= \sum_{i=1}^{n} \frac{\log(Y_i - A) - \mu}{\sigma^2} = 0, \\
\frac{\partial \ell_n(\boldsymbol{\theta})}{\partial \sigma} &= -\frac{n}{\sigma} + \sum_{i=1}^{n} \frac{\{\log(Y_i - A) - \mu\}^2}{\sigma^3} = 0, \\
\frac{\partial \ell_n(\boldsymbol{\theta})}{\partial A} &= \sum_{i=1}^{n} \frac{1}{Y_i - A} + \sum_{i=1}^{n} \frac{\log(Y_i - A) - \mu}{\sigma^2 (Y_i - A)} = 0.
\end{aligned}
\tag{5}
$$

We obtain the following theorem describing the consistency and asymptotic normality of the MLE.

Theorem 1. *Denote the true vale of $\boldsymbol{\theta}$ by $\boldsymbol{\theta}_0 = (\mu_0, \sigma_0, A_0)^{\mathrm{T}}$ for the three parameter log-normal distribution. With probability approaching 1 under $\boldsymbol{\theta}_0$, there exists a sequence of local maximizers, $\hat{\boldsymbol{\theta}}_n$, of the likelihood function in* (4) *that is $\sqrt{n}$-consistent and satisfies $\partial \ell_n(\hat{\boldsymbol{\theta}}_n)/\partial \boldsymbol{\theta} = 0$. Furthermore, under $\boldsymbol{\theta}_0$,*

$$
\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} N\left\{0,\ \mathcal{I}^{-1}(\boldsymbol{\theta}_0)\right\},
$$

where the Fisher information matrix

$$
\mathcal{I}(\boldsymbol{\theta}) = \frac{1}{\sigma^2} \begin{pmatrix} 1 & 0 & -\exp(-\mu + \frac{1}{2}\sigma^2) \\ 0 & 2 & -2\sigma \exp(-\mu + \frac{1}{2}\sigma^2) \\ -\exp(-\mu + \frac{1}{2}\sigma^2) & -2\sigma \exp(-\mu + \frac{1}{2}\sigma^2) & (1 + \sigma^2) \exp(-2\mu + 2\sigma^2) \end{pmatrix}.
$$

Remark 1. *The asymptotic normality in Theorem 1 is established under $\boldsymbol{\theta}_0$. It can also be established under contiguous alternatives. Let $\boldsymbol{\theta}_n = \boldsymbol{\theta}_0 + n^{-1/2}\mathbf{h}_n$, where $\mathbf{h}_n \to \mathbf{h} \in \mathbb{R}^3$. Then under $\boldsymbol{\theta}_n$, $\widehat{\boldsymbol{\theta}}_n$ satisfies*

$$
\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{d} N\left\{\mathbf{h},\ \mathcal{I}^{-1}(\boldsymbol{\theta}_0)\right\}.
$$

Theorem 1 assures the existence of a consistent MLE for a large sample that is the solution to the likelihood equation (5), of which a rigorous proof has been missing from the literature for 5 decades. For practical use, one just needs to solve (5) to find the consistent estimator. From the first two equations of (5), $\mu = n^{-1}\sum_{i=1}^{n} \log(Y_i - A)$ and $\sigma^2 = n^{-1}\sum_{i=1}^{n}\{\log(Y_i - A) - n^{-1}\sum_{i=1}^{n}\log(Y_i - A)\}^2$, which, inserted into the last equation of (5), yields

$$
\lambda(A) = \sum_{i=1}^{n} \frac{1}{Y_i - A} + \sum_{i=1}^{n} \frac{n \log(Y_i - A) - \sum_{i=1}^{n} \log(Y_i - A)}{(Y_i - A)\sum_{j=1}^{n}\{\log(Y_j - A) - n^{-1}\sum_{k=1}^{n}\log(Y_k - A)\}^2} = 0. \tag{6}
$$

To solve (5), one just needs to solve the univariate non-linear equation (6). For given data, $\lambda(A) \to 0$ as $A \to -\infty$. But since the log-likelihood goes to $-\infty$ along this path, one can ignore values of A that are far from $Y_{(1)}$. The first term of $\lambda(A)$ dominates the second term when $A \uparrow Y_{(1)}$, so $\lambda(A) \uparrow \infty$ as $A \uparrow Y_{(1)}$ and hence there is no solution to (6) when A is too close to $Y_{(1)}$. From these two facts, one only needs to numerically search in a closed interval on the left hand side of $Y_{(1)}$ and Theorem 1 assures the existence of a consistent solution to (6) for large samples.

Figure 1 plots $\lambda(A)$ for a sample of $n = 15$ with $\boldsymbol{\theta}_0 = (0,1,1)^{\mathrm{T}}$. The lower panel of Figure 1 is the log-likelihood as a function of A with μ and σ satisfying the first two equations of (5). It is seen that there are two solutions to (6), one (blue round point) is a local maximizer and the other (red diamond point) is a local minimizer that is very close to $Y_{(1)}$. The log-likelihood goes to infinity as A approaches $Y_{(1)}$, but the local maximizer $\hat{A}$ gives better estimation. We also tried various values of $\boldsymbol{\theta}$ and n and found that the shapes of $\lambda(A)$ and the log-likelihood are similar to those in Figure 1 if a solution to (5) exists. In our numerical studies, a unique local maximizer always exists if n is not too small and σ is not too large. We provide the following theorem on the uniqueness of the local maximizer for the log-likelihood to justify this observation.

Theorem 2. *Let δ be some fixed value and $\delta_n = n^{-\alpha}$ for some $\alpha > 0$. Denote by $S_\delta = \{\boldsymbol{\theta} : A \leq A_0 - \delta\}$ and $T_{\delta,n} = \{\boldsymbol{\theta} : A_0 - \delta \leq A \leq A_0 + \delta_n$ and $|\mu - \mu_0| + |\sigma - \sigma_0| > \delta\}$. Then, for each compact set $K \subset \mathbb{R} \times (0,\infty) \times \mathbb{R}$,*

$$\lim_{n\to\infty} P_{\boldsymbol{\theta}_0}\left\{\sup_{S_\delta \cap K} \ell_n(\boldsymbol{\theta}) < \ell_n(\boldsymbol{\theta}_0)\right\} = 1, \text{ and} \tag{7}$$

$$\lim_{n\to\infty} P_{\boldsymbol{\theta}_0}\left\{\sup_{T_{\delta,n} \cap K} \ell_n(\boldsymbol{\theta}) < \ell_n(\boldsymbol{\theta}_0)\right\} = 1, \tag{8}$$

where $P_{\boldsymbol{\theta}_0}$ is the probability measure under $\boldsymbol{\theta}_0$.

This result is useful when there are multiple solutions to the likelihood equations. By this theorem, one can insert each solution into the log-likelihood (4) and choose the solution that yields the largest value.

4. Proof

Proofs of the theorems in Section 3 rely on some lemmas which we now establish. In the following, we use $\dot{\ell}_n(\cdot)$ and $\ddot{\ell}_n(\cdot)$ to denote the gradient and Hessian matrix of the log-likelihood (4), respectively.

Lemma 1. *As $n \to \infty$ under $\boldsymbol{\theta}_0$,*

$$\frac{e^{-\mu_0+\sigma_0 r_n}}{\sigma_0 s_n}\left\{\frac{1}{e^{-\mu_0+\sigma_0 r_n}} - (Y_{(1)} - A_0)\right\} \xrightarrow{d} G, \tag{9}$$

where $r_n = \sqrt{2\log n} - \{\log\log n + \log(4\pi)\}/\sqrt{8\log n}$, $s_n = 1/\sqrt{2\log n}$, and G is a random variable with distribution function $F(t) = e^{-e^{-t}}$.

Proof. First, following the idea in Chapter 2.3 of Galambos (1978), for any $t \neq 0$,

$$\begin{aligned}
&\lim_{n\to\infty} \mathrm{P}_{\boldsymbol{\theta}_0}\left[\left\{-\frac{\log(Y - A_0) - \mu_0}{\sigma_0}\right\}_{(n)} < r_n + s_n t\right] = e^{-e^{-t}} \\
&= \lim_{n\to\infty} \mathrm{P}_{\boldsymbol{\theta}_0}\left[-\left\{\frac{\log(Y - A_0) - \mu_0}{\sigma_0}\right\}_{(1)} < r_n + s_n t\right] \\
&= \lim_{n\to\infty} \mathrm{P}_{\boldsymbol{\theta}_0}\left\{-(Y_{(1)} - A_0) < -e^{\mu_0 - \sigma_0 r_n} + \sigma_0 s_n e^{\mu_0 - \sigma_0 r_n} t \frac{1 - e^{-\sigma_0 s_n t}}{\sigma_0 s_n t}\right\}
\end{aligned}$$

From Lemma 2.2.2 of Galambos (1978), the result in (9) follows from the fact that $(1-e^{-\sigma_0 s_n t})/(\sigma_0 s_n t) \to 1$ for any $t \neq 0$. When $t = 0$, the result can be verified by using the properties of the extreme order statistics of normal distribution directly. $\square$

Lemma 2. *Let $\delta_n = n^{-\alpha}$ for some $\alpha > 0$ and let A satisfy $|A - A_0| < \delta_n$. Then for any positive constants k_1 and k_2, there exists a constant M such that*

$$\lim_{n\to\infty} P_{\boldsymbol{\theta}_0}\left\{\sup_{|A-A_0|<\delta_n} \frac{1}{n}\sum_{i=1}^{n} \frac{|\log(Y_i - A)|^{k_1}}{(Y_i - A)^{k_2}} < M\right\} = 1. \tag{10}$$

Proof. Let $Z_i = \log(Y_i - A_0)$. Note that on $\{Y_i \geq A_0 + 2\delta_n\}$, or equivalently on $\{e^{Z_i} \geq 2\delta_n\}$, $0.5e^{Z_i} < e^{Z_i} + A_0 - A < 1.5e^{Z_i}$. So

$$\begin{aligned}
\frac{|\log(Y_i - A)|^{k_1}}{(Y_i - A)^{k_2}} &= \frac{|\log(e^{Z_i} + A_0 - A)|^{k_1} I(e^{Z_i} \geq 2\delta_n)}{(e^{Z_i} + A_0 - A)^{k_2}} + \frac{|\log(Y_i - A)|^{k_1} I(Y_i < A_0 + 2\delta_n)}{(Y_i - A)^{k_2}} \\
&\leq \frac{[\max\{|\log(0.5e^{Z_i})|, |\log(1.5e^{Z_i})|\}]^{k_1} I(e^{Z_i} \geq 2\delta_n)}{(0.5e^{Z_i})^{k_2}} + \frac{|\log(Y_i - A)|^{k_1} I(Y_{(1)} < A_0 + 2\delta_n)}{(Y_i - A)^{k_2}} \\
&\leq 2^{k_2 - 1}\Big[\{\log(1.5) + |Z_i|\}^{2k_1} + e^{-2k_2 Z_i}\Big] + \frac{|\log(Y_i - A)|^{k_1}}{(Y_i - A)^{k_2}} I\{\delta_n^{-1}(Y_{(1)} - A_0) < 2\}
\end{aligned} \tag{11}$$

Since a normal distribution has finite moments of positive orders and a finite moment generating function, the first term in the right hand side of (11) has an average that converges to a finite constant in probability. The indicator function in the second term does not depend on i, and from the result in (9) of Lemma 1 the probability that the indicator function is nonzero is $o(1)$. This shows that the average of the second term is $o_P(1)$. So any $M > 2^{k_2-1}E\Big[\{\log(1.5) + |Z_i|\}^{2k_1} + e^{-2k_2 Z_i}\Big]$ satisfies (10). $\square$

Lemma 3. *Let $\delta_n = n^{-\alpha}$ for some $\alpha > 0$. For Model* (1),

$$\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|<\delta_n} \left\|\frac{1}{n}\ddot{\ell}_n(\boldsymbol{\theta}) + \mathcal{I}(\boldsymbol{\theta}_0)\right\| = O_{P_{\boldsymbol{\theta}_0}}\left\{\max\left(\frac{1}{\sqrt{n}},\ \delta_n\right)\right\}.$$

Proof. The second order partial derivatives of the log-likelihood are

$$\begin{aligned}
\frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \mu^2} &= -\frac{n}{\sigma^2}, & \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \sigma \partial \mu} &= -\sum_{i=1}^{n} \frac{2\{\log(Y_i - A) - \mu\}}{\sigma^3}, \\
\frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial A \partial \mu} &= -\sum_{i=1}^{n} \frac{1}{\sigma^2 (Y_i - A)}, & \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \sigma^2} &= \frac{n}{\sigma^2} - \sum_{i=1}^{n} \frac{3\{\log(Y_i - A) - \mu\}^2}{\sigma^4}, \\
\frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial \sigma \partial A} &= -\sum_{i=1}^{n} \frac{2\{\log(Y_i - A) - \mu\}}{\sigma^3 (Y_i - A)}, & \frac{\partial^2 \ell_n(\boldsymbol{\theta})}{\partial A^2} &= \sum_{i=1}^{n} \frac{1}{(Y_i - A)^2}\left(1 - \frac{1}{\sigma^2}\right) + \sum_{i=1}^{n} \frac{\log(Y_i - A) - \mu}{\sigma^2 (Y_i - A)^2}.
\end{aligned}$$

So each third order partial derivative of the log-likelihood can be represented as a linear combination of the form $\sum_{i=1}^{n}\left[\{\log(Y_i - A)\}^{k_1}/(Y_i - A)^{k_2}\right]$, with k_1 being 0 or 1, k_2 being 0, 1, 2 or 3, and coefficients being continuous functions of μ and σ. Thus Lemma 2 and a first order Taylor's expansion of each element of $n^{-1}\ddot{\ell}_n(\boldsymbol{\theta})$ gives

$$\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|<\delta_n} \frac{1}{n}\left\|\ddot{\ell}_n(\boldsymbol{\theta}) - \ddot{\ell}_n(\boldsymbol{\theta}_0)\right\| = O_{P_{\boldsymbol{\theta}_0}}(\delta_n). \tag{12}$$

Note that $\ddot{\ell}_n(\boldsymbol{\theta}_0)$ is a sum of independent and identically distributed random vectors with a finite covariance matrix. From the central limit theorem (c.f. Van der Vaart, 2000),

$$\frac{1}{n}\ddot{\ell}_n(\boldsymbol{\theta}_0) + \mathcal{I}(\theta_0) = O_{P_{\boldsymbol{\theta}_0}}\left(\frac{1}{\sqrt{n}}\right). \tag{13}$$

Combining (12) and (13), the result follows. □

Lemma 3 implies a uniform local asymptotic normality (LAN, Le Cam (1960)) condition below:

$$\sup_{\|\mathbf{h}\|\le C_n}\left|\ell_n\left(\boldsymbol{\theta}_0 + \frac{1}{\sqrt{n}}\mathbf{h}\right) - \ell_n(\boldsymbol{\theta}_0) - \frac{1}{\sqrt{n}}\mathbf{h}^{\mathrm{T}}\dot{\ell}_n(\boldsymbol{\theta}_0) + \frac{1}{2}\mathbf{h}^{\mathrm{T}}\mathcal{I}(\boldsymbol{\theta}_0)\mathbf{h}\right| = o_p(1) \tag{14}$$

for any $C_n = o(n^{1/6})$. This can be shown by using a Taylor's Theorem (Ferguson, 1996) that yields

$$\begin{aligned}\Delta_n(\mathbf{h}) &= \ell_n\left(\boldsymbol{\theta}_0 + \frac{1}{\sqrt{n}}\mathbf{h}\right) - \ell_n(\boldsymbol{\theta}_0) - \frac{1}{\sqrt{n}}\mathbf{h}^{\mathrm{T}}\dot{\ell}_n(\boldsymbol{\theta}_0) + \frac{1}{2}\mathbf{h}^{\mathrm{T}}\mathcal{I}(\boldsymbol{\theta}_0)\mathbf{h}\\ &= \mathbf{h}^{\mathrm{T}}\int_0^1\int_0^1 s\left\{\frac{1}{n}\ddot{\ell}_n\left(\boldsymbol{\theta} + st\frac{\mathbf{h}}{\sqrt{n}}\right) + \mathcal{I}(\boldsymbol{\theta}_0)\right\}\mathrm{d}t\mathrm{d}s\ \mathbf{h}\end{aligned}$$

and the right-hand side of the above equation can be bounded by

$$\frac{C_n^2}{2}\sup_{\|\mathbf{h}\|\le C_n}\left\|\frac{1}{n}\ddot{\ell}_n\left(\boldsymbol{\theta} + \frac{\mathbf{h}}{\sqrt{n}}\right) + \mathcal{I}(\boldsymbol{\theta}_0)\right\| = O_{P_{\boldsymbol{\theta}_0}}\left(\frac{C_n^3}{\sqrt{n}}\right) = o_p(1).$$

The uniform LAN condition (14) establishes the existence of a local maximizer of the log-likelihood. With this condition and Le Cam's Third Lemma (Section 12.3 of Lehmann and Romano, 2005), the asymptotic normality under contiguous alternatives follows from Theorem 1. However, The uniform LAN condition itself does not indicate that the maximizer satisfies the likelihood equation (5). We provide the following lemma to show this. It is the Lemma 5 of Smith (1985). We state it for completeness and skip the proof.

Lemma 4. *Let h be a continuously differentiable real-valued function of $p+1$ real variables and let H denote the gradient vector of h. Suppose that the scalar product of $\mathbf{u}$ and $H(\mathbf{u})$ is negative whenever $\|\mathbf{u}\| = 1$. Then h has a local maximum, at which $H = 0$, for some $\mathbf{u}$ with $\|\mathbf{u}\| < 1$.*

Now we prove Theorem 1.

Proof of Theorem 1. From Lemma 3 and the mean value theorem (Ferguson, 1996), for $C_n' = o(n^{1/6})$,

$$\begin{aligned}&\sup_{\|\mathbf{h}\|\le C_n'}\mathbf{h}^{\mathrm{T}}\left\{\frac{1}{\sqrt{n}}\dot{\ell}_n\left(\boldsymbol{\theta}_0 + \frac{1}{\sqrt{n}}\mathbf{h}\right) - \frac{1}{\sqrt{n}}\dot{\ell}_n(\boldsymbol{\theta}_0) + \mathcal{I}(\boldsymbol{\theta}_0)\mathbf{h}\right\}\\ &= \sup_{\|\mathbf{h}\|\le C_n'}\mathbf{h}^{\mathrm{T}}\int_0^1\left\{\frac{1}{n}\ddot{\ell}_n\left(\boldsymbol{\theta} + s\frac{\mathbf{h}}{\sqrt{n}}\right) + \mathcal{I}(\boldsymbol{\theta}_0)\right\}\mathrm{d}s\ \mathbf{h}\\ &\le (C_n')^2 \times \sup_{\|\mathbf{h}\|\le C_n'}\left\|\frac{1}{n}\ddot{\ell}_n\left(\boldsymbol{\theta} + \frac{\mathbf{h}}{\sqrt{n}}\right) + \mathcal{I}(\boldsymbol{\theta}_0)\right\| = O_{P_{\boldsymbol{\theta}_0}}\left\{\frac{(C_n')^3}{\sqrt{n}}\right\} = o_p(1).\end{aligned}$$

So when $C_n' \to \infty$ slowly,

$$C_n'\sup_{\|\mathbf{u}\|=1}\mathbf{u}^{\mathrm{T}}\frac{1}{\sqrt{n}}\dot{\ell}_n\left(\boldsymbol{\theta}_0 + \frac{C_n'}{\sqrt{n}}\mathbf{u}\right) = C_n'\sup_{\|\mathbf{u}\|=1}\left\{\frac{1}{\sqrt{n}}\mathbf{u}^{\mathrm{T}}\dot{\ell}_n(\boldsymbol{\theta}_0) - C_n'\mathbf{u}^{\mathrm{T}}\mathcal{I}(\boldsymbol{\theta}_0)\mathbf{u}\right\} + o_p(1) \to -\infty, \tag{15}$$

in probability, which implies

$$\lim_{n\to\infty} \mathrm{P}_{\boldsymbol{\theta}_0}\left\{\sup_{\|\mathbf{u}\|=1} \mathbf{u}^{\mathrm{T}}\dot{\ell}_n\left(\boldsymbol{\theta}_0+\frac{C_n'}{\sqrt{n}}\mathbf{u}\right)<0\right\}=1. \tag{16}$$

From the results in (15), (16) and Lemma 4, with probability approaching 1 under $\boldsymbol{\theta}_0$, there is a root $\widehat{\boldsymbol{\theta}}_n$ to $\dot{\ell}_n(\boldsymbol{\theta})=0$ within radius $C_n'/\sqrt{n}$ of the true parameter. And since C_n' can goes to ∞ as slow as possible, this root must satisfy

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)=\{\mathcal{I}(\boldsymbol{\theta}_0)\}^{-1}\frac{1}{\sqrt{n}}\frac{\partial\ell_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}+o_P(1).$$

Noting that Model (1) is differentiable in quadratic mean, the asymptotic normality follows from Theorem 12.4.1 of Lehmann and Romano (2005).

□

Proof of Theorem 2, equation (7). First we show that for any $\boldsymbol{\theta}_1\in S$, $E\ell_n(\boldsymbol{\theta}_1)<\infty$, so $E[\ell_n(\boldsymbol{\theta}_1)-\ell_n(\boldsymbol{\theta}_0)]<0$ by Jensen's inequality. This implies the existence of $\xi_{\boldsymbol{\theta}_1}$ such that

$$\lim_{n\to\infty}\mathrm{P}_{\boldsymbol{\theta}_0}\{\ell_n(\boldsymbol{\theta}_1)-\ell_n(\boldsymbol{\theta}_0)<-\xi_{\boldsymbol{\theta}_1}\}=1.$$

For $\boldsymbol{\theta}$ and η such that $\|\boldsymbol{\theta}-\boldsymbol{\theta}_1\|<\eta<\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_0\|-\delta$,

$$\begin{aligned}\frac{1}{n}|\ell_n(\boldsymbol{\theta})-\ell_n(\boldsymbol{\theta}_1)| \le |\log\sigma-\log\sigma_1| &+ \frac{1}{n}\sum_{i=1}^n|\log(Y_i-A)-\log(Y_i-A_1)|\\ &+\frac{1}{n}\sum_{i=1}^n\left|\frac{\{\log(Y_i-A)-\mu\}^2}{\sigma^2}-\frac{\{\log(Y_i-A_1)-\mu_1\}^2}{\sigma_1^2}\right|=\Delta_3+\Delta_4+\Delta_5.\end{aligned}$$

Δ_3 can be made smaller than $\xi_{\boldsymbol{\theta}_1}/4$ by choosing η small enough. By the mean value theorem,

$$\Delta_4=\frac{1}{n}\sum_{i=1}^n\left|\frac{A-A_1}{Y_i-A^*}\right|\le\frac{1}{n}\sum_{i=1}^n\frac{|A-A_1|}{Y_i-A_0},$$

for some A^* between A_0 and A_1. So $E(\Delta_4)$ can be make arbitrarily small by choosing small enough η, which implies

$$\lim_{n\to\infty}\mathrm{P}_{\boldsymbol{\theta}_0}\left(\Delta_4<\frac{\xi_{\boldsymbol{\theta}_1}}{4}\right)=1$$

for small enough η. Using a similar approach, we obtain

$$\lim_{n\to\infty}\mathrm{P}_{\boldsymbol{\theta}_0}\left(\Delta_5<\frac{\xi_{\boldsymbol{\theta}_1}}{4}\right)=1.$$

Combining results for Δ_3, Δ_4 and Δ_5,

$$\lim_{n\to\infty}\mathrm{P}_{\boldsymbol{\theta}_0}\left\{\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_1\|<\eta}\ell_n(\boldsymbol{\theta})-\ell_n(\boldsymbol{\theta}_0)<-\frac{\xi_{\boldsymbol{\theta}_1}}{4}\right\}=1.$$

For any compact set K, $S_\delta\cap K$ can be covered by a finite number of neighborhoods of points in S_δ, so it follows that

$$\lim_{n\to\infty}\mathrm{P}_{\boldsymbol{\theta}_0}\left\{\sup_{S_\delta\cap K}\ell_n(\boldsymbol{\theta})-\ell_n(\boldsymbol{\theta}_0)<-\xi_m\right\}=1.$$

□

Proof of Theorem 2, equation (8). First, if A_0 is known, model (1) can be transformed to a linear model with normal random error with unknown mean and variance. It follows that

$$\lim_{n\to\infty} \mathrm{P}_{\boldsymbol{\theta}_0} \left\{ \sup_{|\mu-\mu_0|>\delta,\, |\sigma-\sigma_0|>\delta} \ell_n(\mu,\sigma,A_0) - \ell_n(\boldsymbol{\theta}_0) < -\xi \right\} = 1. \tag{17}$$

For μ_1, σ_1, η and $(\mu,\sigma,A) \in T$ such that $(\mu_1,\sigma_1,A) \in T$, $|\mu-\mu_1| < \eta$, $|\sigma-\sigma_1| < \eta$ and $\delta < \eta$,

$$\begin{aligned} \frac{1}{n}|\ell_n(\mu,\sigma,A) - \ell_n(\mu_1,\sigma_1,A_0)| \leq & |\log\sigma - \log\sigma_1| + \frac{1}{n}\sum_{i=1}^n |\log(Y_i - A) - \log(Y_i - A_0)| \\ & + \frac{1}{n}\sum_{i=1}^n \left| \frac{\{\log(Y_i - A) - \mu\}^2}{\sigma^2} - \frac{\{\log(Y_i - A_0) - \mu_1\}^2}{\sigma_1^2} \right| \\ = & \Delta_6 + \Delta_7 + \Delta_8. \end{aligned} \tag{18}$$

The term Δ_6 can be made smaller than $\xi/8$ by choosing η small enough. By the mean value theorem,

$$\Delta_7 = \frac{1}{n}\sum_{i=1}^n \left| \frac{A - A_0}{Y_i - A^*} \right| \leq \frac{|A - A_0|}{n} \sum_{i=1}^n \frac{1}{|Y_i - \max(A, A_0)|}$$

with probability tending to 1. If $A \leq A_0$,

$$\frac{1}{n}\sum_{i=1}^n \frac{1}{|Y_i - \max(A, A_0)|} = \frac{1}{n}\sum_{i=1}^n \frac{1}{(Y_i - A_0)},$$

and the right hand side of the upper inequality goes to $\exp(-\mu + \sigma^2/2)$ in probability. If $A_0 < A < A_0 + \delta_n$, Lemma 2 provides that there exists some constant M^* such that

$$\lim_{n\to\infty} \mathrm{P}_{\boldsymbol{\theta}_0} \left(\frac{1}{n}\sum_{i=1}^n \frac{1}{|Y_i - A|} < M^* \right) = 1$$

for small enough η. This implies that for small enough η,

$$\lim_{n\to\infty} \mathrm{P}_{\boldsymbol{\theta}_0} \left(\Delta_7 < \frac{\xi}{4} \right) = 1. \tag{19}$$

The same result can be found for Δ_8 using similar arguments.

$$\lim_{n\to\infty} \mathrm{P}_{\boldsymbol{\theta}_0} \left(\Delta_8 < \frac{\xi}{4} \right) = 1. \tag{20}$$

Combining (17), (18), (19) and (20), we have

$$\lim_{n\to\infty} \mathrm{P}_{\boldsymbol{\theta}_0} \left\{ \sup \ell_n(\mu,\sigma,A) - \ell_n(\boldsymbol{\theta}_0) < -\frac{3\xi}{4} \right\} = 1,$$

where the supremum is taken over all $\boldsymbol{\theta}$ satisfying $(\mu_1,\sigma_1,A) \in T$, $|\mu-\mu_1| < \eta$ and $|\sigma-\sigma_1| < \eta$ for fixed μ_1 and σ_1. This result can be extended directly to any finite set of values of μ_1 and σ_1, and then to any compact sets of values of μ_1 and σ_1. □

Acknowledgements

We are very grateful to a referee for constructive suggestions and detailed comments which have greatly improved the quality of this paper.

Aitchison, J. and Brown, J. A. C. (1957). *The Lognormal Distribution.* Cambridge University Press, England.

Basak, P., Basak, I., and Balakrishnan, N. (2009). Estimation for the three-parameter lognormal distribution based on progressively censored data. *Computational Statistics & Data Analysis* **53**, 3580 – 3592.

Cheng, R. C. H. and Amin, N. A. K. (1979). Maximum product of spacings estimation with application to the lognormal distribution. Tech. rep., University of Wales Institute of Science and Technology, Cardiff. Math. Report 79-1.

Cheng, R. C. H. and Amin, N. A. K. (1981). Maximum likelihood estimation of parameters in the inverse gaussian distribution, with unknown origin. *Technometrics* **23**, 257–263.

Cheng, R. C. H. and Amin, N. A. K. (1983). Estimating parameters in continuous univariate distributions with a shifted origin. *Journal of the Royal Statistical Society. Series B. Statistical Methodology* **45**, 394–403.

Cheng, R. C. H. and Iles, T. C. (1987). Corrected maximum likelihood in non-regular problems. *Journal of the Royal Statistical Society. Series B. Statistical Methodology* **49**, 95–101.

Cheng, R. C. H. and Traylor, L. (1995). Non-regular maximum likelihood problems. *Journal of the Royal Statistical Society. Series B. Statistical Methodology* **57**, 3–44.

Cohen, A. C. (1951). Estimating parameters of logarithmic-normal distributions by maximum likelihood. *Journal of the American Statistical Association* **46**, 206–212.

Cohen, A. C. and Whitten, B. J. (1980). Estimation in the three-parameter lognormal distribution. *Journal of the American Statistical Association* **75**, 399–404.

Cramér, H. (1946). *Mathematical Methods of Statistics.* Princeton University Press.

Crow, E. and Shimizu, K. (1998). *Lognormal Distributions: Theory and Applications.* CRC Press, New York.

Ferguson, T. S. (1996). *A Course in Large Sample Theory.* Chapman and Hall.

Galambos, J. (1978). *The asymptotic theory of extreme order statistics.* John Wiley & Sons, New York.

Giesbrecht, F. and Kempthorne, O. (1976). Maximum likelihood estimation in the three-parameter lognormal distribution. *Journal of the Royal Statistical Society. Series B. Statistical Methodology* **38**, 257–264.

Hall, P. (1982). On estimating the endpoint of a distribution. *The Annals of Statistics* **10**, 556–568.

Hall, P. and Wang, J. Z. (2005). Bayesian likelihood methods for estimating the end point of a distribution. *Journal of the Royal Statistical Society. Series B. Statistical Methodology* **67**, 717–729.

Harter, H. L. and Moore, A. H. (1966). Local-maximum-likelihood estimation of the parameters of three-parameter lognormal populations from complete and censored samples. *Journal of the American Statistical Association* **61**, 842–851.

Hill, B. M. (1963). The three-parameter lognormal distribution and Bayesian analysis of a point-source epidemic. *Journal of the American Statistical Association* **58**, 72–84.

Le Cam, L. (1960). Locally asymptotically normal families of distributions. *University of California Publications in Statistics* **3**, 37–98.

Le Cam, L. (1990). Maximum likelihood: An introduction. *International Statistical Review* **58**, 153–171.

LeCam, L. (1970). On the assumptions used to prove asymptotic normality of maximum likelihood estimates. *The Annals of Mathematical Statistics* **41**, 802–828.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses.* Springer, 3rd edn.

Munro, A. H. and Wixley, R. A. J. (1970). Estimators based on order statistics of small samples from a three-parameter lognormal distribution. *Journal of the American Statistical Association* **65**, 212–225.

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika* **72**, 67–90.

Smith, R. L. (1989). A survey of nonregular problems. *Proceedings of International Statistical Institute Conference 47 Session, Paris* 353–372.

Smith, R. L. (1994). Nonregular regression. *Biometrika* **81**, 173–183.

Van der Vaart, A. W. (2000). *Asymptotic Statistics.* Cambridge University Press.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *The Annals of Mathematical Statistics* **20**, 595–601.

Weiss, L. and Wolfowitz, J. (1973). Maximum likelihood estimation of a translation parameter of a truncated distribution. *The Annals of Statistics* **1**, 944–947.

Woodroofe, M. (1972). Maximum likelihood estimation of a translation parameter of a truncated distribution. *The Annals of Mathematical Statistics* **43**, 113–122.

Woodroofe, M. (1974). Maximum likelihood estimation of translation parameter of truncated distribution ii. *The Annals of Statistics* **2**, 474–488.

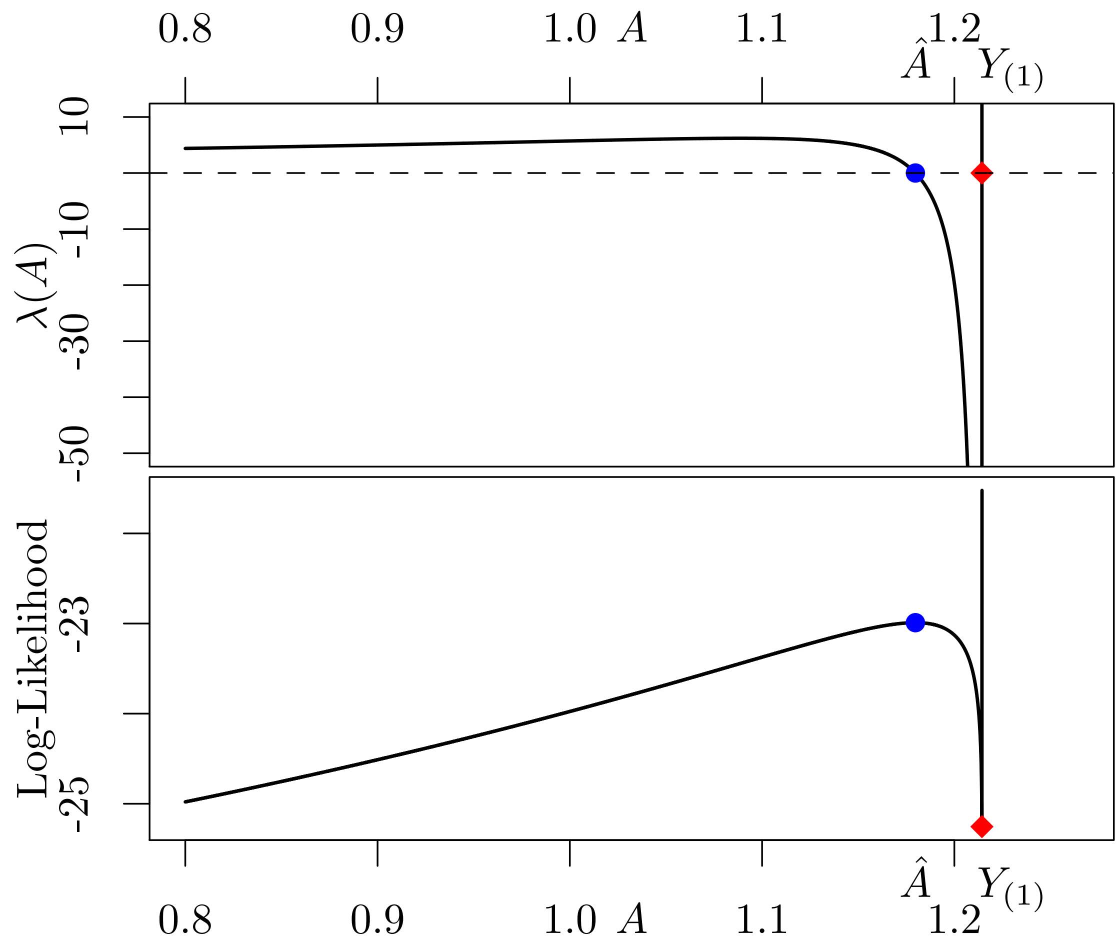

Figure 1: $\lambda(A)$ and Log-likelihood for a sample of size $n = 15$ with $\mu_0 = 0$, $\sigma = 1$ and $A_0 = 1$.