Log-Epsilon-Skew Normal Distribution and Applications

Terry L. Mashtare Jr.

Department of Biostatistics

University at Buffalo, 249 Farber Hall, 3435 Main Street, Buffalo, NY 14214-3000, U.S.A.

Alan D. Hutson

Department of Biostatistics

University at Buffalo, 249 Farber Hall, 3435 Main Street, Buffalo, NY 14214-3000, U.S.A.

Roswell Park Cancer Institute, Elm & Carlton Streets, Buffalo, NY 14263, U.S.A.

Govind S. Mudholkar

Department of Statistics

University of Rochester, Rochester, NY 14627, U.S.A.

August 20, 2009

Abstract

In this note we introduce a new family of distributions called the log-epsilon-skew-normal (LESN) distribution. This family of distributions can trace its roots back to the epsilon-skew-normal (ESN) distribution developed by Mudholkar and Hutson (2000). A key advantage of our model is that the well known lognormal distribution is a subclass of the LESN distribution. We study its main properties, hazard function, moments, skewness and kurtosis coefficients, and discuss maximum likelihood estimation of model parameters. We summarize the results of a simulation study to examine the behavior of the maximum likelihood estimates, and we illustrate the maximum likelihood estimation of the LESN distribution parameters to a real world data set.

Keywords: Lognormal distribution, maximum likelihood estimation.

1 Introduction

In this note we introduce a new family of distributions called the log-epsilon-skew-normal (LESN) distribution. This family of distributions can trace its roots back to the epsilon-skew-normal (ESN) distribution developed by Mudholkar and Hutson (2000) [14]. A key advantage of our model is that the well known lognormal distribution is a subclass of the LESN distribution. The usage of the lognormal distribution spans across several fields such as astrophysics [10], environmental sciences [4], and radiology [15]. Eckhard (2001) compared the use of the lognormal distribution across several different science disciplines [7].

The lognormal distribution has a long and rich history. Galton (1879) suggested the use of the lognormal distribution to analyze data for which the geometric mean is better than the arthimatic mean for estimating central tendency [9]. McAlister (1879) derived the lognormal distribution at Galton's suggestion [13]. Finney (1941) examined the moments, moment estimation, and efficiency of the estimation [8]. Another interesting result was put forth by Heyde (1963). He showed the lognormal distribution is not uniquely determined by its moments [11]. It is known that the moment generating function for the lognormal distribution does not exist [16]. However, it is well-known that the moments of the lognormal distribution can be derived from the moment generating function of the normal distribution.

In terms of background, Let $T \sim LN(\theta, \sigma)$. Then the probability distribution function (pdf) and cumulative distribution function (cdf) for T are given by

$$f_T(t) = \frac{1}{t\sqrt{2\pi}\sigma} \exp\left(-\frac{(\log t - \theta)^2}{2\sigma^2}\right), \tag{1.1}$$

and

$$F_T(t) = \Phi\left(\frac{\log t - \theta}{\sigma}\right), \tag{1.2}$$

respectively, where $\Phi(\cdot)$ denotes the standard normal cdf.

In survival analysis, the lognormal distribution is used in accelerated failure time (AFT) models [6]. The form of the lognormal hazard function is given by

$$h_T(t) = \frac{\left(\frac{1}{t\sqrt{2\pi}\sigma}\right) \exp\left(-\frac{(\log t - \theta)^2}{2\sigma^2}\right)}{1 - \Phi\left(\frac{\log t - \theta}{\sigma}\right)} \tag{1.3}$$

The lognormal hazard function is unimodal starting at zero and increasing to a maximum, then decreases toward zero as $t \to \infty$. Sweet (1990) gives mathematical properties of the hazard rate for the lognormal distribution [17].

One competing approach towards generalizing the lognormal distribution would be to utilize the skew normal distribution of Azzalini (1985,1986) [2, 3]. This approach has been utilized in practice by Chai and Bailey (2008) for the purpose of modeling continuous data with a discrete component at zero [5]. However, the mathematical properties has not been developed.

In section 3, we introduce the LESN distribution and discuss its properties. In section 4, we examine the maximum likelihood estimates of the LESN parameters. In section 5 we summarize the results of a simulation study to examine the behavior of the maximum likelihood estimates. In section 6, we illustrate the maximum likelihood estimation of the LESN distribution parameters to a real world data set.

2 Notation

In order to simplify our presentation of our demonstrations we first introduce some notational devices. Towards this end let:

T, t $LESN$: random variable, value taken by T

θ, σ, ϵ: ESN location, scale, and skewness parameters

$Z_1 = \frac{\log T - \theta}{\sigma(1-\epsilon)}$ and $Z_2 = \frac{\log T - \theta}{\sigma(1+\epsilon)}$

$f_{Z_i}(\cdot), F_{Z_i}(\cdot), S_{Z_i}(\cdot), h_{Z_i}$: pdf, cdf, survival function, and hazard function for Z_i, $i = 1, 2$.

$\phi(\cdot), \Phi(\cdot)$: pdf and cdf of a standard normal random variable.

3 The log-epsilon-skew-normal distribution

We utilize the ESN model as a starting point for our development of the LESN distribution. Details of the epsilon-skew-normal (ESN) distribution [14], developed by Mudholkar and Hutson (2000), are given in appendix A. Toward this end let T be a random variable such that $\log(T) \sim ESN(\theta, \sigma, \epsilon)$. It follows that $T \sim LESN(\theta, \sigma, \epsilon)$. The pdf, cdf, and the quantile function of the LESN distribution are given as follows:

$$f_T(t) = \begin{cases} \frac{1}{t\sqrt{2\pi}\sigma} \exp\left(-\frac{(\log t - \theta)^2}{2\sigma^2(1-\epsilon)^2}\right), & \text{if } 0 < t < e^{\theta}, \\ \frac{1}{t\sqrt{2\pi}\sigma} \exp\left(-\frac{(\log t - \theta)^2}{2\sigma^2(1+\epsilon)^2}\right), & \text{if } t \geq e^{\theta}, \end{cases} \tag{3.1}$$

$$F_T(t) = \begin{cases} (1-\epsilon)\Phi\left(\frac{\log t - \theta}{\sigma(1-\epsilon)}\right), & \text{if } 0 < t < e^{\theta}, \\ -\epsilon + (1+\epsilon)\Phi\left(\frac{\log t - \theta}{\sigma(1+\epsilon)}\right), & \text{if } t \geq e^{\theta}, \end{cases} \tag{3.2}$$

and,

$$Q_T(u) = \exp[\theta + \sigma Q_0(u)], \tag{3.3}$$

respectively, where $-1 < \epsilon < 1$ and $Q_0(u)$ is given by equation (A.3). As can be seen at $\epsilon = 0$, equations (3.1) and (3.2) give equations (1.1) and (1.2).

3.1 Median and mode

It is straightforward to verify that the median of $LESN(\theta, \sigma, \epsilon)$ is

$$Q_T(1/2) = \begin{cases} \exp\left[\theta + \sigma(1-\epsilon)\Phi^{-1}\left(\frac{1}{2(1-\epsilon)}\right)\right], & \text{if } \epsilon < 0 \\ \exp\left[\theta + \sigma(1+\epsilon)\Phi^{-1}\left(\frac{1/2+\epsilon}{1+\epsilon}\right)\right], & \text{if } \epsilon \geq 0. \end{cases} \tag{3.4}$$

Its mode is

$$e^{\theta - \sigma^2(1-\epsilon)^2}. \tag{3.5}$$

Figure 1 shows some typical LESN probability density functions with $\theta = 0$. A horizontal line is drawn at the splice point $t = e^\theta = 1$. We see that the LESN pdf is unimodal, with the mode before the splice point $t = 1$. This compares to the result at (3.5). We see a variety of shapes with high peaks for large σ. This will be further examined within the context of the hazard function given below.

3.2 Hazard Function

The hazard function for the LESN is given by the following function:

$$h_T(t) = \begin{cases} \frac{\left(\frac{1}{t\sigma}\right)\phi\left(\frac{\log t - \theta}{\sigma(1-\epsilon)}\right)}{1-(1-\epsilon)\Phi\left(\frac{\log t - \theta}{\sigma(1-\epsilon)}\right)}, & \text{if } 0 < t < e^\theta, \\ \frac{\left(\frac{1}{t\sigma}\right)\phi\left(\frac{\log t - \theta}{\sigma(1+\epsilon)}\right)}{(1+\epsilon)[1-\Phi\left(\frac{\log t - \theta}{\sigma(1+\epsilon)}\right)]}, & \text{if } t \geq e^\theta. \end{cases} \tag{3.6}$$

Using the notation given in section 2, we can rewrite $h_T(t)$ more compactly:

$$h_T(t) = \begin{cases} \frac{(1-\epsilon)f_{Z_1}(z_1)}{[S_{Z_1}(z_1)]+\epsilon F_{Z_1}(z_1)}, & \text{if } 0 < t < e^\theta, \\ \frac{f_{Z_2}(z_2)}{S_{Z_2}(z_2)}, & \text{if } t \geq e^\theta. \end{cases} \tag{3.7}$$

Note that the hazard function for $t \in [e^\theta, \infty)$ corresponds to the hazard function for a $LN(\theta, \sigma(1+\epsilon))$ random variable. However for $t \in (0, e^\theta)$ the form of the hazard function is more complicated. This yields some interesting results as developed below.

In order to investigate the possible hazard shapes we examine the derivative of $h_T(t)$, denoted by $h'_T(t)$. It can be shown that

$$h'_T(t) = h_T(t)[h_T(t) - g(t)], \tag{3.8}$$

where

$$g(t) = \begin{cases} \frac{1}{t}\left[\frac{\log t - \theta}{\sigma^2(1-\epsilon)^2} + 1\right], & \text{if } 0 < t < e^\theta, \\ \frac{1}{t}\left[\frac{\log t - \theta}{\sigma^2(1+\epsilon)^2} + 1\right], & \text{if } t \geq e^\theta. \end{cases} \tag{3.9}$$

Now through examination of $h_T(t)$ at (3.7) and $h'_T(t)$ we have the following are properties of the LESN hazard function:

1.
$$h_T(0) = 0. \tag{3.10}$$

2.
$$h'_T(0) = 0 \tag{3.11}$$

3. For $t \in (0, e^\theta)$, if $\epsilon < 0$ then
$$h_T(t) < h_Z(t) \tag{3.12}$$
where $h_Z(t)$ is the hazard function for $LN(\theta, \sigma)$.

   Similarly, if $\epsilon > 0$ then
$$h_T(t) > h_Z(t) \tag{3.13}$$

4.
$$\lim_{t\to\infty} h_T(t) = 0 \tag{3.14}$$

5.

$$\lim_{t\to\infty} h'_T(t) = 0 \tag{3.15}$$

6.

$$h_T(e^\theta) = \frac{2}{e^\theta \sigma(1+\epsilon)\sqrt{2\pi}} \tag{3.16}$$

7. $h'_T(e^\theta)$ exists and is given by

$$h'_T(e^\theta) = \frac{2 - \sigma(1+\epsilon)\sqrt{2\pi}}{e^{2\theta}\pi\sigma^2(1+\epsilon)^2} \tag{3.17}$$

8. At $t = e^\theta$, $h_T(t)$ is increasing if

$$\sigma(1+\epsilon) < \sqrt{\frac{2}{\pi}}. \tag{3.18}$$

9. At $t = e^\theta$, $h_T(t)$ is decreasing if

$$\sigma(1+\epsilon) > \sqrt{\frac{2}{\pi}}. \tag{3.19}$$

10. $t = e^\theta$ is a relative maximum of $h_T(t)$ if

$$\sigma(1+\epsilon) = \sqrt{\frac{2}{\pi}}. \tag{3.20}$$

In summary, some of the properties above are self explainatory. In addition, property 3 implies that for $t \in (0, e^\theta)$, if $\epsilon < 0$, then the hazard function for a LESN random variable is bounded above by the hazard function for a LN random variable. Similarly, if $\epsilon > 0$, then the hazard function for a LESN random variable is bounded below by the hazard function for a LN random variable. Property 6 implies that $h_T(t)$ is continuous at the splice point $t = e^\theta$. Property 8 combined with equation (3.7) imply that if $\sigma(1+\epsilon) < \sqrt{\frac{2}{\pi}}$, then the hazard function has a relative max for some $t \in [e^\theta, \infty)$. This is illustrated in Figure (2f). Similarly, property 9 combined with equation (3.7) imply that if $\sigma(1+\epsilon) > \sqrt{\frac{2}{\pi}}$, then the hazard function decreases on $t \in [e^\theta, \infty)$.

3.3 Derivations of properties of LESN hazard function

In general, it is well-known that we can write the hazard function and the derivative of the hazard function for T as

$$h_T(t) = \frac{f_T(t)}{1 - F_T(t)} \tag{3.21}$$

and

$$h'_T(t) = \frac{f'_T(t)}{1 - F_T(t)} + [h_T(t)]^2. \tag{3.22}$$

We first examine the properties of the hazard function over the interval $(0, e^\theta)$. Utilizing (3.21) and (3.22), we have $h_T(0) = f_T(0) = 0$ and $h'_T(0) = f'_T(0) = 0$. This proves properties 1 and 2.

If $\epsilon > 0$, we have

$$\frac{(1-\epsilon) f_{Z_1}(z_1)}{[S_{Z_1}(z_1)] + \epsilon F_{Z_1}(z_1)} \leq \frac{(1-\epsilon) f_{Z_1}(z_1)}{[S_{Z_1}(z_1)]} < \frac{f_{Z_1}(z_1)}{[S_{Z_1}(z_1)]}. \tag{3.23}$$

Similarly if $\epsilon < 0$ we have

$$\frac{(1-\epsilon) f_{Z_1}(z_1)}{[S_{Z_1}(z_1)] + \epsilon F_{Z_1}(z_1)} \geq \frac{(1-\epsilon) f_{Z_1}(z_1)}{[S_{Z_1}(z_1)]} > \frac{f_{Z_1}(z_1)}{[S_{Z_1}(z_1)]}, \tag{3.24}$$

Thus property 3 follows.

Properties 4 and 5 follow from the fact that $h_T(t)$ for $t \in [e^\theta, \infty)$ corresponds to the hazard function for a $LN(\theta, \sigma(1+\epsilon))$ random variable. Sweet (1990) showed that the hazard function for a lognormal random variable tends to 0 as $t \to \infty$ [17].

Property 6 follows from substituting $t = e^\theta$ in equation (3.6). Note that $h_T(t)$ is continuous at $t = e^\theta$. Using equations (3.8) and (3.9) we have

$$\lim_{t \to e^{\theta -}} h'_T(t) = \lim_{t \to e^{\theta +}} h'_T(t) = \frac{2 - \sigma(1+\epsilon)\sqrt{2\pi}}{e^{2\theta} \pi \sigma^2 (1+\epsilon)^2}, \tag{3.25}$$

thus property 7 follows. Properties 8 – 10 follow directly from property 7.

3.4 Moments

By definition, the r-th moment of T from an $LESN(\theta, \sigma, \epsilon)$ distribution is given by

$$\mu_r' = E(T^r) = E(e^{r \log T}) = E(e^{rX}) = M_X(r) = e^{\theta r} M(\sigma r), \tag{3.26}$$

where $M_X(\cdot)$ and $M(\cdot)$ are the moment generating functions for $X \sim ESN(\theta, \sigma, \epsilon)$ and $ESN(0, 1, \epsilon)$ random variables, respectively. Using equations (3.26) and (A.6), we have

$$\mu_r' = E(T^r) = e^{\theta r}\{(1-\epsilon)e^{(1-\epsilon)^2\sigma^2 r^2/2}\Phi[-(1-\epsilon)\sigma r] + (1+\epsilon)e^{(1+\epsilon)^2\sigma^2 r^2/2}\Phi[(1+\epsilon)\sigma r]\}. \tag{3.27}$$

For the purposes of organizing our expressions let $\tau_1 = e^{(1-\epsilon)^2\sigma^2}$, $\tau_2 = e^{(1+\epsilon)^2\sigma^2}$, $\nu_1 = (1-\epsilon)\sigma$, and $\nu_2 = (1+\epsilon)\sigma$. We can then rewrite the moment equation (3.27) as

$$\mu_r' = e^{\theta r}\{\tau_1^{r^2/2}(1-\epsilon)\Phi(-\nu_1 r) + \tau_2^{r^2/2}(1+\epsilon)\Phi(\nu_2 r)\}. \tag{3.28}$$

Using equation (3.28), the first four central moments are then given by

$$\mu_1' = e^{\theta}\{\sqrt{\tau_1}(1-\epsilon)\Phi(-\nu_1) + \sqrt{\tau_2}(1+\epsilon)\Phi(\nu_2)\}, \tag{3.29}$$

$$\mu_2' = e^{2\theta}\{\tau_1^2(1-\epsilon)\Phi(-2\nu_1) + \tau_2^2(1+\epsilon)\Phi(2\nu_2)\}, \tag{3.30}$$

$$\mu_3' = e^{3\theta}\{\tau_1^{9/2}(1-\epsilon)\Phi(-3\nu_1) + \tau_2^{9/2}(1+\epsilon)\Phi(3\nu_2)\}, \tag{3.31}$$

$$\mu_4' = e^{4\theta}\{\tau_1^8(1-\epsilon)\Phi(-4\nu_1) + \tau_2^8(1+\epsilon)\Phi(4\nu_2)\}. \tag{3.32}$$

We note that the first four central moments $\mu_r = E[T - \mu_1']^r$ can be calculated from the first four central moments by

$$\mu_1 = \mu_1', \tag{3.33}$$

$$\mu_2 = \mu_2' - \mu_1^2, \tag{3.34}$$

$$\mu_3 = \mu_3' - 3\mu_1\mu_2' + 2\mu_1^3, \tag{3.35}$$

$$\mu_4 = \mu_4' - 4\mu_1\mu_3' + 6\mu_1^2\mu_2' - 3\mu_1^4. \tag{3.36}$$

Furthermore, it follows that the coefficients of skewness and kurtosis are

$$\sqrt{\beta_1} = \sqrt{\frac{\mu_3}{\mu_2^3}} \text{ and } \beta_2 = \frac{\mu_4}{\mu_2^2}. \tag{3.37}$$

It should be noted the skewness and kurtosis for the lognormal distribution involves only the parameter σ. As compared to the lognormal, we note that $\sqrt{\beta_1}$ and β_2 involve the parameters σ and ϵ.

4 Maximum likelihood estimation

Let $Y_1, Y_2, \cdots, Y_n$ denote i.i.d. $Y_i \sim LESN(\theta, \sigma, \epsilon)$. The loglikelihood for the i-th observation is given by

$$l_{LESN}(\theta, \sigma^2, \epsilon | y_i) = \begin{cases} \log y_i - \frac{1}{2}\log 2\pi\sigma - \frac{(\log y_i - \theta)^2}{2\sigma^2(1-\epsilon)^2}, & \text{if } \log y_i < \theta, \\ \log y_i - \frac{1}{2}\log 2\pi\sigma - \frac{(\log y_i - \theta)^2}{2\sigma^2(1+\epsilon)^2}, & \text{if } \log y_i \geq \theta, \end{cases} \tag{4.1}$$

Denote the loglikelihood for the i-th observation for the $ESN(\theta, \sigma, \epsilon)$ as $l_{ESN}(\theta, \sigma^2, \epsilon | x_i)$, then $l_{LESN}(\theta, \sigma^2, \epsilon | y_i) = \log y_i + l_{ESN}(\theta, \sigma^2, \epsilon | \log y_i)$. We denote the maximum likelihood estimates of $(\theta, \sigma^2, \epsilon)$ as $(\hat{\theta}, \hat{\sigma}^2, \hat{\epsilon})$.

Let $y_{(1)} \leq y_{(2)} \leq \cdots \leq y_{(n)}$ denote the order statistic of a sample from the $LESN(\theta, \sigma, \epsilon)$ population. Also denote $y_0 = 0$ and $y_{n+1} = \infty$. We apply some of the results from Mudholkar and Hutson (2000) to the $LESN(\theta, \sigma, \epsilon)$ distribution.

Lemma 4.1. *There exists and integer $k = k(y_{(1)}, \ldots, y_{(n)})$, $0 \leq k \leq n$, such that the loglikelihood $l(\theta, \sigma^2, \epsilon)$ can be expressed as*

$$l(\theta, \sigma^2, \epsilon) = \begin{cases} -\frac{n}{2}\log 2\pi\sigma - \sum_{i=1}^{n}\log y_{(i)} - \frac{1}{8\sigma^2}\sum_{i=1}^{n}[(\log y_{(i)} - \theta)^2], & \textit{if } k = 0, n, \\ -\frac{n}{2}\log 2\pi\sigma - \sum_{i=1}^{n}\log y_{(i)} - \frac{1}{2\sigma^2}[\sum_{i=1}^{k}\frac{(\log y_i - \theta)^2}{(1-\epsilon)^2} + \sum_{i=k+1}^{n}\frac{(\log y_i - \theta)^2}{(1+\epsilon)^2}], & \textit{if } 1 \leq k < n, \end{cases} \tag{4.2}$$

Lemma 4.2. *If $k = 0$ or $k = n$ the maximum likelihood estimate $(\hat{\theta}, \hat{\sigma}^2, \hat{\epsilon})$ is given by*

$$(\hat{\theta}, \hat{\sigma}^2, \hat{\epsilon}) = \begin{cases} (\log y_{(n)}, s_n^2, -1) & \text{if } k = 0, \\ (\log y_{(1)}, s_0^2, 1) & \text{if } k = n, \end{cases} \tag{4.3}$$

where $s_0^2 = \sum_{i=2}^{n} (\log y_{(i)} - \log y_{(1)})^2/(4n)$ and $s_n^2 = \sum_{i=1}^{n-1} (\log y_{(i)} - \log y_{(n)})^2/(4n)$.

Lemma 4.3. *If $1 \le k < n$ then the local minima of*

$$g_j(\boldsymbol{y}, \theta) = \frac{1}{4} \left\{ \left[\sum_{i=1}^{j} (\log y_{(i)} - \theta)^2 \right]^{1/3} + \left[\sum_{i=j+1}^{n} (\log y_{(i)} - \theta)^2 \right]^{1/3} \right\}^3, \tag{4.4}$$

$j = 1, 2, \ldots, n-1$, if any, determines the local maxima of the loglikelihood.

Lemma 4.4. *Let θ_0 denote a solution of*

$$\begin{aligned} h_j'(\boldsymbol{y}, \theta) = & -\frac{2}{3} \left\{ \sum_{i=1}^{j} (\log y_{(i)} - \theta) \left[\sum_{i=1}^{j} (\log y_{(i)} - \theta)^2 \right]^{-2/3} \right. \\ & \left. + \sum_{i=j+1}^{n} (\log y_{(i)} - \theta) \left[\sum_{i=j+1}^{n} (\log y_{(i)} - \theta)^2 \right]^{-2/3} \right\} = 0, \end{aligned} \tag{4.5}$$

$j = 1, 2, \ldots, n-1$. Then the corresponding values $\sigma^2 = \sigma_0^2$ and $\epsilon = \epsilon_0$ which maximize the loglikelihood are given by

$$\epsilon_0 = \frac{[\sum_{i=j+1}^{n} (\log y_{(i)} - \theta_0)^2]^{1/3} - [\sum_{i=1}^{j} (\log y_{(i)} - \theta_0)^2]^{1/3}}{[\sum_{i=j+1}^{n} (\log y_{(i)} - \theta_0)^2]^{1/3} + [\sum_{i=1}^{j} (\log y_{(i)} - \theta_0)^2]^{1/3}}, \tag{4.6}$$

$$\sigma_0^2 = \frac{1}{4n} \left\{ \left[\sum_{i=1}^{j} (\log y_{(i)} - \theta_0)^2 \right]^{1/3} + \left[\sum_{i=j+1}^{n} (\log y_{(i)} - \theta_0)^2 \right]^{1/3} \right\}^3 \tag{4.7}$$

Lemma 4.5. *The global maximum of the loglikelihood is obtained by examining the following local maxima for the set $k = 0$, $k = n$, and the set $(\theta_0, \sigma_0^2, \epsilon_0)$ given by*

Lemmas 4.2 and 4.4:

$$Max\ l_k(\theta_0, \sigma_0^2, \epsilon_0) = \begin{cases} -\frac{n}{2}[1 + \log s_n^2] & \text{if } k = 0, \\ -\frac{n}{2}[1 + \log \sigma_0^2] & \text{if } 1 \leq k < n, \\ -\frac{n}{2}[1 + \log s_0^2] & \text{if } k = n, \end{cases} \tag{4.8}$$

where s_0^2 and s_1^2 are given by Lemma 4.2 and σ_0^2 is given by Lemma 4.4.

Theorem 4.6. *As $n \to \infty$, the maximum likelihood estimator $\sqrt{n}(\hat{\theta}_{ml} - \theta, \hat{\sigma}^2_{ml} - \sigma^2, \hat{\epsilon}_{ml} - \epsilon)$ is a centered trivariate normal distribution with variance-covariance matrix*

$$\mathbf{\Sigma}_{ml} = \begin{pmatrix} I^{\theta\theta} = \frac{3\pi(1-\epsilon^2)\sigma^2}{3\pi-8} & I^{\theta\sigma^2} = 0 & I^{\theta\epsilon} = \frac{-2\sqrt{2\pi}(1-\epsilon^2)\sigma}{3\pi-8} \\ & I^{\sigma^2\sigma^2} = 2\sigma^4 & I^{\sigma^2\epsilon} = 0 \\ & & I^{\epsilon\epsilon} = \frac{\pi(1-\epsilon^2)}{3\pi-8} \end{pmatrix}. \tag{4.9}$$

Theorem 4.7. *If $\epsilon = \epsilon_0$ then as $n \to \infty$, $\sqrt{n}(\hat{\theta}_{ml}(\epsilon_0) - \theta, \hat{\sigma}^2_{ml}(\epsilon_0) - \sigma^2)$ is a centered bivariate normal distribution with variance-covariance matrix*

$$\begin{pmatrix} (1-\epsilon_0^2)\sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix}. \tag{4.10}$$

In the context of LESN modeling the parameter ϵ may be interpreted as a measure of distance of the distribution from lognormality with respect to LESN alternatives. Hence, the test

$$\begin{aligned} H_0 &: \epsilon = 0 \\ H_1 &: \epsilon \neq 0 \end{aligned} \tag{4.11}$$

may be used as a diagnostic tool with respect to the appropriateness of testing the appropriateness of an underlying lognormal model versus a broad class of LESN alternatives.

5 Simulation

We conducted two simulation studies. The first study was to examine the behavior of the maximum likelihood based estimates of the LESN parameters. The second part of our study examined the lognormal versus LESN in terms of confidence intervals for the population mean under various LESN parameterizations.

For the first part we let $T \sim LESN(0, 1, \epsilon)$ with sample sizes $n = 25, 50, 100, 200$, and 500. We utilized 10000 simulations under each scenario. We let ϵ range from -0.75 to 0.75 by 0.25, which includes the special case of $\epsilon = 0$ corresponding to the lognormal distribution.

Tables 1 and 2 summarize the maximum likelihood parameter estimation of the LESN model parameters θ, σ, and ϵ. We see there is finite bias in the parameter estimates. As expected, the bias approaches zero and the variance of the parameter estimates decrease as the sample size increases. Hence, via simulation we have that the maximum likelihood estimates of the LESN parameters are well-behaved, thus validating the results at (4.9).

for the second part, for each data set simulated, we estimated $E(Y) = e^{\theta+\sigma^2/2}$ for the lognormal and $E(Y)$ for LESN as given in equation (3.29). 95% confidence intervals were computed for each sample using standard errors calculated via the delta method. Using the true mean, we were able to obtain estimated coverage probabilities of the confidence intervals. Results are summarized in tables 3 and 4.

We see that when simulating under the lognormal distribution ($\epsilon = 0$), both models perform well with the coverage probabilities approaching 95% as the sample size increases. Once $\epsilon \neq 0$, we see that the estimated coverage probabilities for the lognormal fit deteriorate rapidly and tend toward 0 as the sample size increases. More specifically, when $\epsilon = -0.25$, which is near lognormal, the coverage probability for the lognormal fit falls below 90% when $n = 500$. When $\epsilon = 0.25$, the coverage probability for the lognormal fit is always below 90% and falls as low as 70%. The coverage probability for the LESN is very well-behaved and approaches 95% when n gets larger.

6 Example

For the purpose of illustration we look at the Mayo Clinic primary biliary cirrhosis data that can be downloaded from *http://biostat.mc.vanderbilt.edu/wiki/Main/DataSets*. In particular, we examine 418 serum bilirubin (mg/dl) measurements. Some basic statistics are reported in the next table

n	$\overline{x}$	s	$\sqrt{\beta_1}$
418	3.22	4.41	2.72

In the table that follows, maximum likelihood estimators are presented for lognormal and LESN. It is interesting to note the p-value for testing equation 4.11 is $p < 0.0001$. This suggests that the lognormal model may not be appropriate.

Parameters	LESN	Lognormal
θ	-0.4169	0.5715
σ	0.9273	1.0238
ϵ	0.6713	–
mean	3.52	2.99
standard deviation	8.72	4.07
skewness	0.23	0.31

Figure (4a) depicts histogram of the serum bilirubin measurements and model fitting for lognormal and LESN models. We see that the LESN model provides a better fit. This is supported by figure (4b) which depicts a histogram of the natural logarithm of the serum bilirubin measurements and model fitting for normal and ESN models. Even after the log transformation, the data are clearly right skewed.

Appendix A

Some basic properties of the ESN distribution, developed by Mudholkar and Hutson (2000) [14] used in this note are below. The pdf, cdf, and qf of its canonical form $ESN(0,1,\epsilon)$ are respectively:

$$f_0(x) = \begin{cases} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2(1-\epsilon)^2}\right), & \text{if } x < 0, \\ \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2(1+\epsilon)^2}\right), & \text{if } x \geq 0, \end{cases} \tag{A.1}$$

$$F_0(x) = \begin{cases} (1-\epsilon)\Phi\left(\frac{x}{1-\epsilon}\right), & \text{if } x < 0, \\ -\epsilon + (1+\epsilon)\Phi\left(\frac{x}{1+\epsilon}\right), & \text{if } x \geq 0, \end{cases} \tag{A.2}$$

and,

$$Q_0(u) = F_0^{-1}(u) = \begin{cases} (1-\epsilon)\Phi^{-1}\left(\frac{u}{1-\epsilon}\right), & \text{if } 0 < u < (1-\epsilon)/2 \\ (1+\epsilon)\Phi^{-1}\left(\frac{u+\epsilon}{1+\epsilon}\right), & \text{if } (1-\epsilon)/2 \leq u < 1, \end{cases} \tag{A.3}$$

where $-1 < \epsilon < 1$, and $\Phi(x)$ denotes the standard normal c.d.f.

The general form for the p.d.f., denoted $ESN(\mu,\sigma,\epsilon)$, is $f_0(\frac{x-\mu}{\sigma})/\sigma$, where $f_0(\cdot)$ is given by (A.1). Similarly, the general form for the c.d.f. of $ESN(\mu,\sigma,\epsilon)$ is $F_0(\frac{x-\mu}{\sigma})$, where $F_0(\cdot)$ is given by (A.2). Its quantile function, $Q(u) = \mu + \sigma Q_0(u)$, can be used to generate samples from the $ESN(\mu,\sigma,\epsilon)$ population. Note the relationship $Q(\frac{1-\epsilon}{2}) = \mu$.

The mean of $ESN(\mu,\sigma,\epsilon)$is given by

$$E(X) = \mu + \frac{4\sigma\epsilon}{\sqrt{2\pi}}, \tag{A.4}$$

and the variance is given by

$$\text{Var}(X) = \frac{\sigma^2}{\pi}[(3\pi-8)\epsilon^2 + \pi]. \tag{A.5}$$

The moment generating function of the standard epsilon-skew-normal distribution $ESN(0,1,\epsilon)$ is

$$M(t) = (1-\epsilon)e^{(1-\epsilon)^2t^2/2}\Phi[-(1-\epsilon)t] + (1+\epsilon)e^{(1+\epsilon)^2t^2/2}\Phi[(1+\epsilon)t]. \tag{A.6}$$

If $X \sim ESN(\mu, \sigma, \epsilon)$ then $M_X(t) = e^{\mu t} M(\sigma t)$.

References

[1] R.B. Arellano-Valle, H.W. Gómez, F.A. Quintana, Statistical inference for a general class of asymmetric distributions. J.Statist.Plan.Inference.128 (2005) 427-443.

[2] A. Azzalini, A class of distributions which includes the normal ones. Scand.J.Statist.12 (1985) 171-178.

[3] A. Azzalini, Further results on a class of distributions which includes the normal ones. Statistica.46 (1986) 199-208.

[4] J.L. Benning, D.L. Barnesa, The effects of scale and spatial heterogeneities on diffusion in volcanic breccias and basalts: Amchitka Island, Alaska. J.Contam.Hydrology.106 (2009) 150-165.

[5] H.S. Chai, K.R. Bailey, Use of log-skew-normal distribution in analysis of continuous data with a discrete component at zero. Stat.Med.27 (2008) 3643-3655.

[6] D. Collett, Modeling Survival Data in Medical Research, 2nd ed. Chapman & Hall, Boca Raton, Florida, 2003.

[7] L. Eckhard, W.A. Stahel, M. Abbt, Log-normal distributions across the sciences: keys and clues. BioScience.51 (2001) 341-352.

[8] D.J. Finney, On the distribution of a variate whose logarithm is normally distributed. J.Roy.Stat.Soc.Suppl.7 (1941) 155-161.

[9] F. Galton, The geometric mean, in vital and social statistics. Proc.R.Soc.Lond.29 (1879) 365-367.

[10] Gandhi P. The flux-dependent RMS variability of x-ray binaries in the optical. Astrophys.J.697 (2009) L167-L172.

[11] C.C. Heyde, On a property of the lognormal distribution. J.R.Stat.Soc.Ser.B.Stat.Methodol.25 (1963) 392-393.

[12] A.D. Hutson, Utilizing the flexibility of the epsilon-skew-normal normal distribution for common regression problems. J.Appl.Statist.Sci.31 (2004) 673-683.

[13] D. McAlister, The law of the geometric mean. Proc.R.Soc.Lond.29 (1879) 367-376.

[14] G.S. Mudholkar, A.D. Hutson, The epsilon-skew-normal distribution for analyzing near-normal data. J.Statist.Plan.Inference.83 (2000) 291-309.

[15] P. Neti, R.W. Howell, Lognormal distribution of cellular uptake of radioactivity: statistical analysis of α-particle track autoradiography. J.Nucl.Med.49 (2008) 1009-1016.

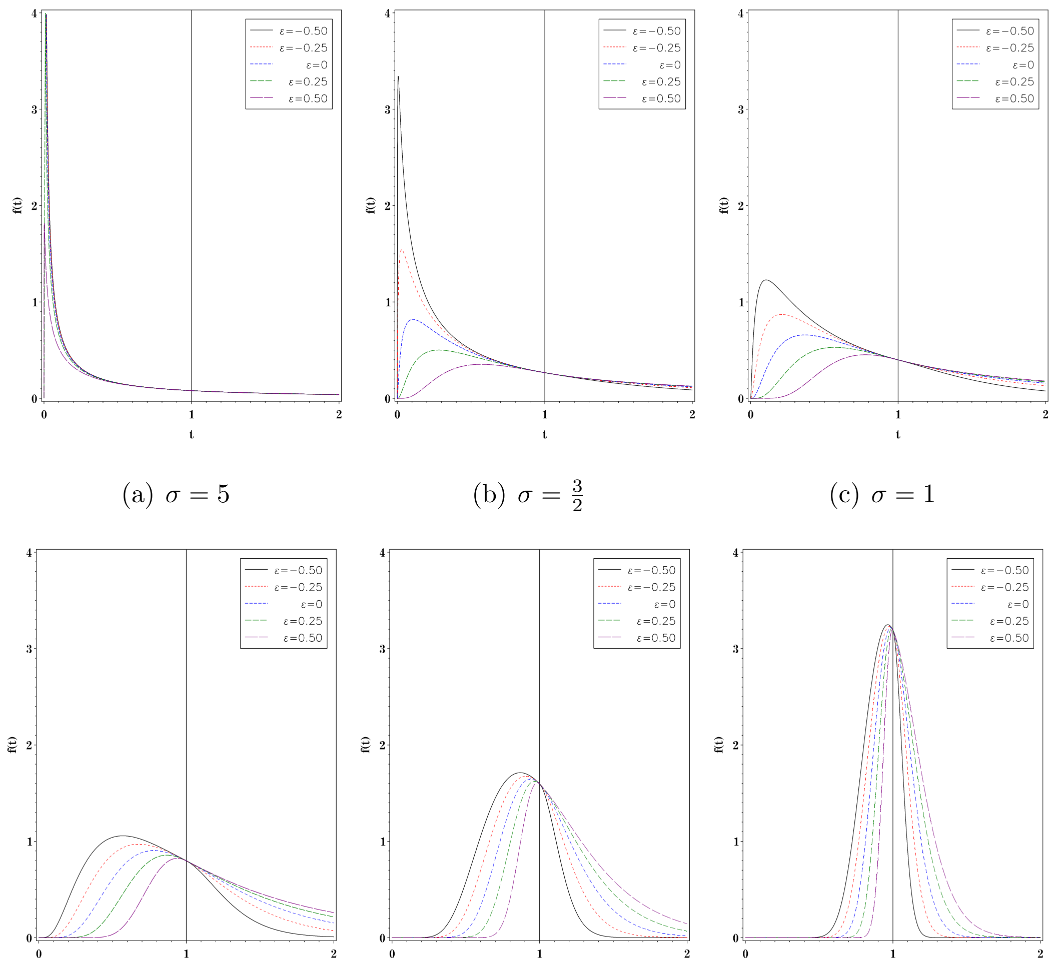
[16] J.P. Romano, A.F. Siegel, Counterexamples in Probability and Statistics. CRC Press, 1986.

[17] A.L. Sweet, On the hazard rate of the lognormal distribution. IEEE.T.Reliab.39 (1990) 325-328.

Table 1: MLE of LESN model fit $\epsilon < 0$

ϵ	n	$\theta = 0$	$\sigma = 1$	ϵ
-0.75	25	-0.031 ± 0.45	0.932 ± 0.14	-0.797 ± 0.29
	50	0.025 ± 0.31	0.967 ± 0.10	-0.797 ± 0.19
	100	0.027 ± 0.21	0.985 ± 0.07	-0.778 ± 0.13
	200	0.016 ± 0.14	0.993 ± 0.05	-0.765 ± 0.08
	500	0.005 ± 0.08	0.998 ± 0.03	-0.755 ± 0.05
-0.50	25	0.015 ± 0.59	0.934 ± 0.14	-0.553 ± 0.39
	50	0.039 ± 0.40	0.971 ± 0.10	-0.541 ± 0.25
	100	0.020 ± 0.26	0.986 ± 0.07	-0.519 ± 0.16
	200	0.008 ± 0.17	0.994 ± 0.05	-0.508 ± 0.10
	500	0.004 ± 0.10	0.997 ± 0.03	-0.503 ± 0.06
-0.25	25	0.024 ± 0.68	0.938 ± 0.14	-0.285 ± 0.45
	50	0.025 ± 0.44	0.972 ± 0.10	-0.273 ± 0.27
	100	0.009 ± 0.28	0.987 ± 0.07	-0.259 ± 0.16
	200	0.003 ± 0.18	0.993 ± 0.05	-0.253 ± 0.11
	500	0.000 ± 0.11	0.998 ± 0.03	-0.251 ± 0.07

Figure 1: Some typical LESN probability density functions



(a) $\sigma = 5$

(b) $\sigma = \frac{3}{2}$

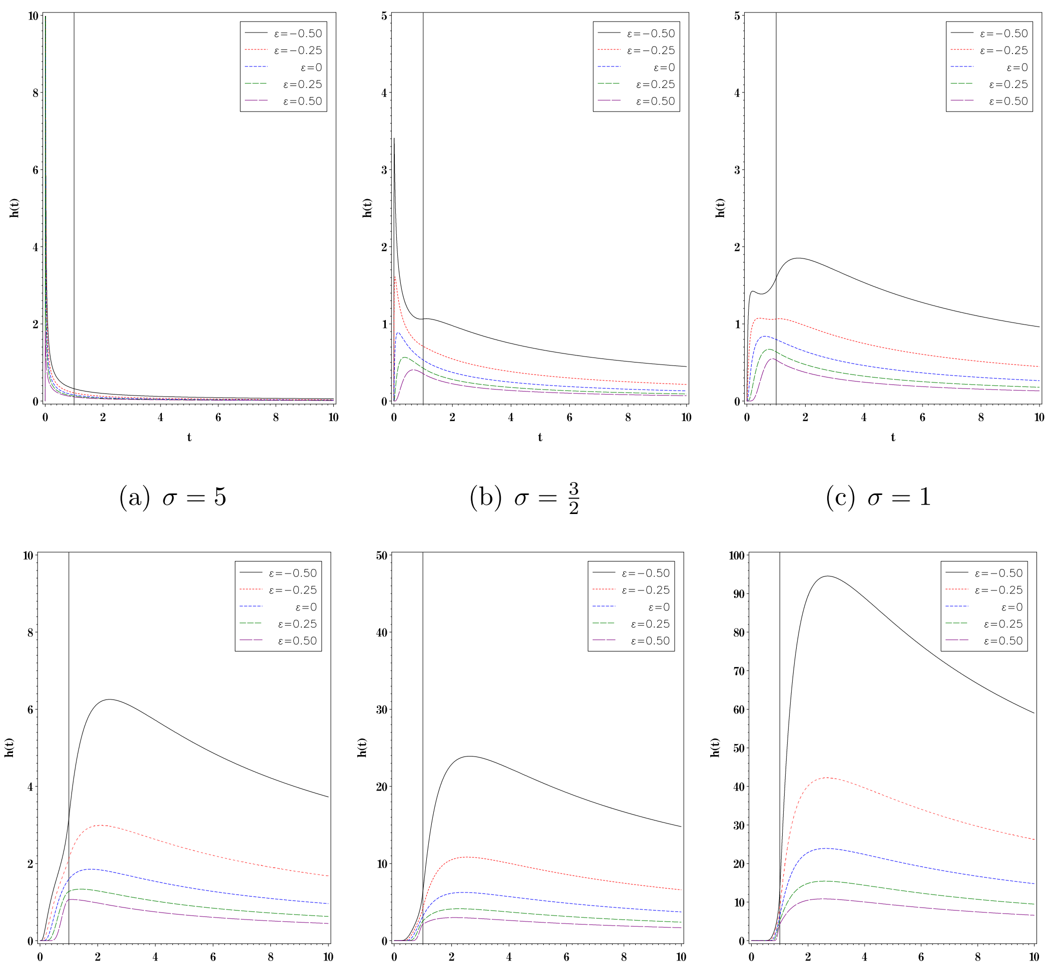
(c) $\sigma = 1$

(d) $\sigma = \frac{1}{2}$

(e) $\sigma = \frac{1}{4}$

(f) $\sigma = \frac{1}{8}$

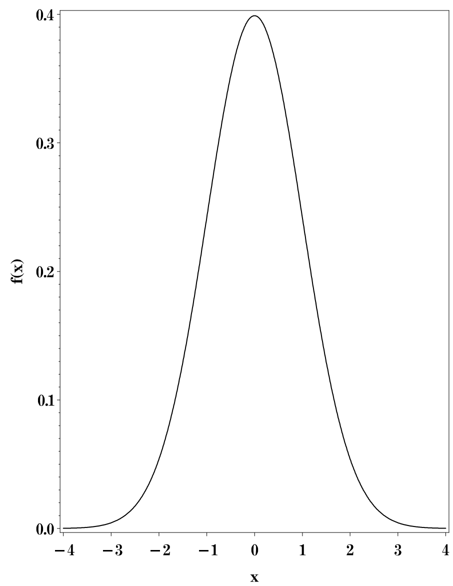
Figure 2: Some typical LESN hazard functions



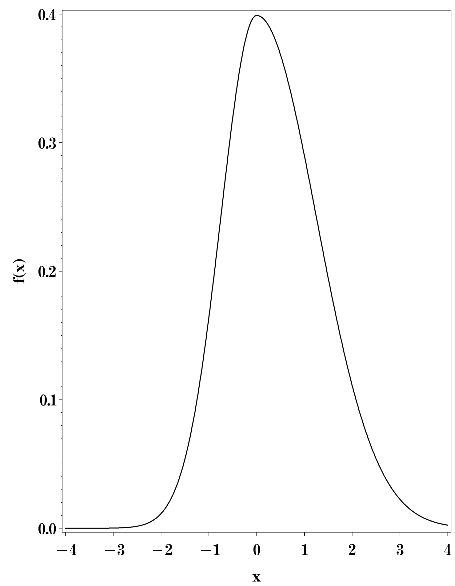
(a) $\sigma = 5$ (b) $\sigma = \frac{3}{2}$ (c) $\sigma = 1$

(d) $\sigma = \frac{1}{2}$ (e) $\sigma = \frac{1}{4}$ (f) $\sigma = \frac{1}{8}$

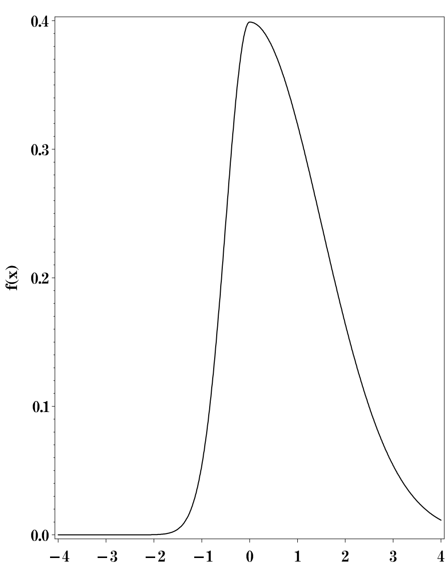
Figure 3: Some typical $ESN(0, 1, \epsilon)$ probability density functions.



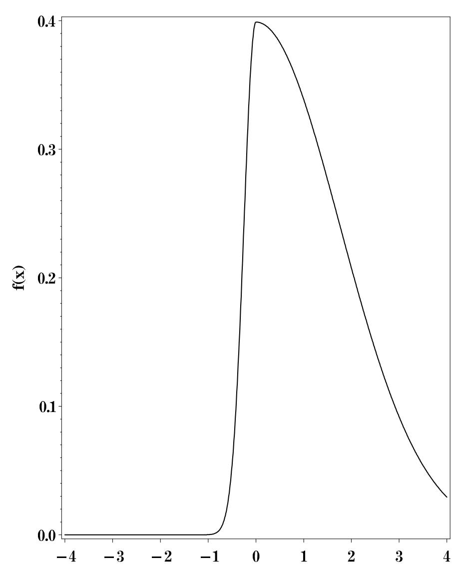
(a) $\epsilon = 0$



(b) $\epsilon = 0.25$

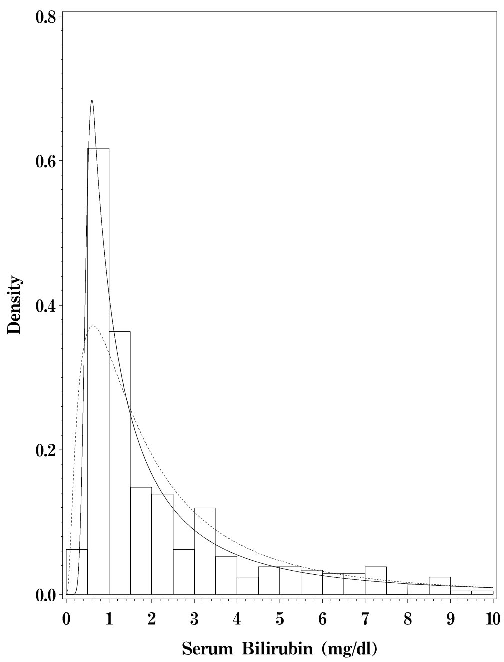


(c) $\epsilon = 0.50$

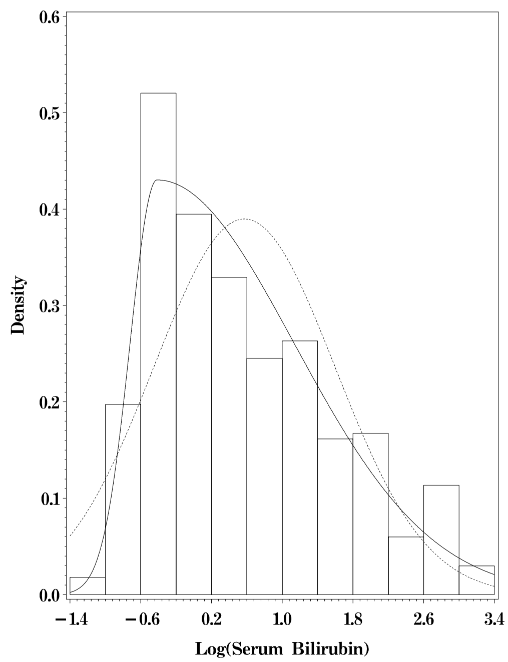


(d) $\epsilon = 0.75$

Figure 4: Histogram of Serum Bilirubin (mg/dl) data.



(a) Model fitting LESN (solid line) and log-normal (dotted line)



(b) Model fitting ESN (solid line) and normal (dotted line)

Table 2: MLE of LESN model fit $\epsilon \geq 0$

ϵ	n	$\theta = 0$	$\sigma = 1$	ϵ
0	25	0.003 ± 0.71	0.938 ± 0.14	-0.001 ± 0.47
	50	0.004 ± 0.45	0.972 ± 0.10	0.000 ± 0.28
	100	0.004 ± 0.28	0.988 ± 0.07	-0.002 ± 0.17
	200	0.005 ± 0.19	0.994 ± 0.05	-0.003 ± 0.11
	500	0.002 ± 0.12	0.997 ± 0.03	-0.001 ± 0.07
0.25	25	-0.032 ± 0.68	0.934 ± 0.14	0.291 ± 0.45
	50	-0.029 ± 0.44	0.973 ± 0.10	0.275 ± 0.27
	100	-0.005 ± 0.28	0.986 ± 0.07	0.257 ± 0.16
	200	-0.001 ± 0.18	0.993 ± 0.05	0.253 ± 0.11
	500	-0.002 ± 0.11	0.997 ± 0.03	0.252 ± 0.07
0.50	25	-0.028 ± 0.58	0.933 ± 0.14	0.564 ± 0.38
	50	-0.048 ± 0.40	0.969 ± 0.10	0.547 ± 0.25
	100	-0.025 ± 0.26	0.987 ± 0.07	0.521 ± 0.16
	200	-0.005 ± 0.17	0.994 ± 0.05	0.506 ± 0.10
	500	-0.001 ± 0.10	0.997 ± 0.03	0.502 ± 0.06
0.75	25	0.024 ± 0.44	0.930 ± 0.14	0.800 ± 0.28
	50	-0.015 ± 0.31	0.966 ± 0.10	0.790 ± 0.19
	100	-0.027 ± 0.21	0.986 ± 0.07	0.779 ± 0.13
	200	-0.014 ± 0.14	0.992 ± 0.05	0.763 ± 0.08
	500	-0.004 ± 0.08	0.997 ± 0.03	0.754 ± 0.05

Table 3: Estimates of Expected Value and Coverage Probabilities $\epsilon < 0$

ϵ	n	$E(Y)$	$E(Y)$ lognormal	CP	$E(Y)$ LESN	CP
	25		0.563 ± 0.12	99.2%	0.464 ± 0.07	86.2%
	50		0.565 ± 0.08	99.6%	0.475 ± 0.05	89.6%
-0.75	100	0.479	0.566 ± 0.06	96.8%	0.479 ± 0.04	91.4%
	200		0.566 ± 0.04	80.6%	0.479 ± 0.03	93.1%
	500		0.566 ± 0.03	22.6%	0.479 ± 0.02	94.6%
	25		0.784 ± 0.16	98.7%	0.698 ± 0.13	87.6%
	50		0.785 ± 0.11	99.2%	0.703 ± 0.09	91.0%
-0.50	100	0.700	0.784 ± 0.08	98.0%	0.702 ± 0.06	92.8%
	200		0.786 ± 0.06	91.1%	0.701 ± 0.04	94.2%
	500		0.785 ± 0.04	58.8%	0.701 ± 0.03	94.4%
	25		1.128 ± 0.25	96.5%	1.078 ± 0.24	89.1%
	50		1.122 ± 0.17	97.9%	1.064 ± 0.16	92.1%
-0.25	100	1.057	1.122 ± 0.12	97.9%	1.060 ± 0.11	94.0%
	200		1.123 ± 0.09	96.8%	1.058 ± 0.08	94.6%
	500		1.123 ± 0.05	89.1%	1.058 ± 0.05	95.0%

Table 4: Estimates of Expected Value and Coverage Probabilities $\epsilon \geq 0$

ϵ	n	$E(Y)$	$E(Y)$ lognormal	CP	$E(Y)$ LESN	CP
0	25	1.649	1.670 ± 0.42	91.2%	1.717 ± 0.50	88.1%
	50		1.660 ± 0.29	93.2%	1.677 ± 0.32	92.2%
	100		1.656 ± 0.21	94.0%	1.661 ± 0.22	93.1%
	200		1.652 ± 0.14	94.7%	1.653 ± 0.15	94.1%
	500		1.649 ± 0.09	95.0%	1.650 ± 0.09	94.7%
0.25	25	2.667	2.529 ± 0.74	82.7%	2.814 ± 1.05	86.7%
	50		2.517 ± 0.51	84.0%	2.750 ± 0.70	90.5%
	100		2.502 ± 0.36	82.5%	2.693 ± 0.45	92.4%
	200		2.498 ± 0.25	79.0%	2.680 ± 0.31	93.7%
	500		2.494 ± 0.16	70.0%	2.673 ± 0.20	94.4%
0.50	25	4.486	4.005 ± 1.40	73.7%	4.846 ± 2.35	86.2%
	50		3.935 ± 0.95	71.3%	4.696 ± 1.50	89.5%
	100		3.904 ± 0.64	66.0%	4.587 ± 0.98	92.2%
	200		3.891 ± 0.45	55.4%	4.528 ± 0.65	93.6%
	500		3.879 ± 0.29	30.4%	4.498 ± 0.41	94.2%
0.75	25	7.871	6.500 ± 2.78	63.7%	8.600 ± 5.47	84.0%
	50		6.339 ± 1.79	59.1%	8.286 ± 3.27	88.7%
	100		6.290 ± 1.21	50.2%	8.156 ± 2.12	91.8%
	200		6.217 ± 0.83	32.4%	7.968 ± 1.42	93.1%
	500		6.212 ± 0.53	8.7%	7.916 ± 0.88	94.2%