mathematics

Article

A Bimodal Extension of the Log-Normal Distribution on the Real Line with an Application to DNA Microarray Data

Mai F. Alfahad [1], Mohamed E. Ghitany [1,*], Ahmad N. Alothman [1] and Saralees Nadarajah [2]

[1] Department of Statistics and Operations Research, Faculty of Science, Kuwait University, Kuwait City 13060, Kuwait; mai.alfahad@ku.edu.kw (M.F.A.); ahmad.alothman@ku.edu.kw (A.N.A.)
[2] Department of Mathematics, University of Manchester, Manchester M13 9PL, UK; saralees.nadarajah@manchester.ac.uk
* Correspondence: me.ghitany@ku.edu.kw

Abstract: A bimodal double log-normal distribution on the real line is proposed using the random sign mixture transform. Its associated statistical inferences are derived. Its parameters are estimated by the maximum likelihood method. The performance of the estimators and the corresponding confidence intervals is checked by simulation studies. Application of the proposed distribution to a real data set from a DNA microarray is presented.

Keywords: bimodality; log-normal distribution; maximum likelihood estimation; Monte Carlo simulations

MSC: 62E15; 62F12

check for updates

Citation: Alfahad, M.F.; Ghitany, M.E.; Alothman, A.N.; Nadarajah, S. A Bimodal Extension of the Log-Normal Distribution on the Real Line with an Application to DNA Microarray Data. *Mathematics* **2023**, *11*, 3360. https://doi.org/10.3390/math11153360

Academic Editors: Babak Shiri and Zahra Alijani

Received: 3 July 2023
Revised: 23 July 2023
Accepted: 25 July 2023
Published: 31 July 2023

1. Introduction

A log-normal distribution is perhaps the most popular model for skewed data [1]. However, a log-normal distribution is defined only on the positive real line. Many of its application areas involve data spanning the entire real line. One example is the modeling of stock returns. The log-normal distribution is a popular model for stock returns. However, stock returns can be positive or negative. Positive stock returns correspond to profits, while negative stock returns correspond to losses. Other application areas of log-normal distributions involving data spanning the entire real line are discussed later on. Hence, a double log-normal distribution is needed.

We follow the procedure presented in [2] to construct a double log-normal (DLN) distribution. Consider the following two transforms [2]:

(i) The random sign transform (RST) given by

$$W = (2Y - 1)X;$$

(ii) The random sign mixture transform (RSMT) given by

$$Z = YX_1 - (1 - Y)X_2,$$

where Y is a Bernoulli random variable (RV) with parameter β and X, X_1 and X_2 are non-negative RVs independent of Y. The probability density function (PDF) of W is

$$f_W(w; \beta, \boldsymbol{\theta}) = \begin{cases} \overline{\beta}\, f_X(|w|; \boldsymbol{\theta}), & w < 0, \\ \beta\, f_X(w; \boldsymbol{\theta}), & w \geq 0, \end{cases}$$

and the PDF of Z is

$$f_Z(z;\beta,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) = \begin{cases} \overline{\beta}\, f_{X_2}(|z|;\boldsymbol{\theta}_2), & z<0, \\ \beta\, f_{X_1}(z;\boldsymbol{\theta}_1), & z \geq 0, \end{cases}$$

where $f_X(\cdot;\boldsymbol{\theta})$, $f_{X_1}(\cdot;\boldsymbol{\theta}_1)$, $f_{X_2}(\cdot;\boldsymbol{\theta}_2)$ are the PDFs of non-negative RVs X, X_1 and X_2, respectively, with (vector) parameters $\boldsymbol{\theta}$, $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, respectively. If X is an RV from a family of distributions $\mathcal{F}_1$, then W is said to have a *double* $\mathcal{F}_1$ distribution. If X_1 and X_2 are independent RVs from a family of distributions $\mathcal{F}_2$, then Z is said to have a *double* $\mathcal{F}_2$ distribution.

There are many double continuous distributions on the real line. However, the words 'double' or 'reflection' are sometimes used to denote the distribution of the absolute value of a random variable. Some double continuous distributions based on the RST are:

1. Double exponential distribution (Laplace) [3].
2. Double generalized gamma distribution [4].
3. Double Weibull distribution [5].
4. Double gamma distribution [6].
5. Double generalized Pareto distribution [7].
6. Double Lomax distribution [8].
7. Double Lindley distribution [9].

Some double continuous distributions based on the RSMT are:

1. Double half-normal distribution [10,11].
2. Double exponential distribution [12].
3. Double inverse gamma distribution [13].
4. Double Gompertz distribution [14].
5. Double Pareto II distribution [15].
6. Double inverse Gaussian distribution [16].

We construct the DLN distribution using the RSMT, i.e., the distribution of Z when X_1 and X_2 independently follow the log-normal distribution.

The remainder of this paper is organized as follows. In Section 2, the statistical properties of the DLN distribution are presented. The maximum likelihood estimates (MLEs) of the parameters and their asymptotic distributions are studied in Section 3. Simulations to check the finite sample performance of the estimators of the parameters and the corresponding confidence intervals are presented in Section 4. An application of the proposed double distribution to a real data set from a DNA microarray is presented in Section 5. Finally, the conclusions and comments are stated in Section 6.

2. Statistical Properties

We present the statistical properties of the DLN distribution in this section.

2.1. Probability Density Function

The PDF of the DLN distribution is

$$f_Z(z) = \begin{cases} \overline{\beta}\, f_{X_2}(|z|;\mu_2,\sigma_2), & z<0, \\ \beta\ \ f_{X_1}(z;\mu_1,\sigma_1), & z \geq 0, \end{cases}$$

where for $-\infty < \mu_1, \mu_2 < \infty$ and $\sigma_1, \sigma_2 > 0$, and

$$f_{X_j}(x;\mu_j,\sigma_j) = \frac{1}{\sqrt{2\pi}\,\sigma\, x}\exp\left[-\frac{(\ln(x)-\mu_j)^2}{2\sigma_j^2}\right], \quad x>0, \quad j=1,2 \tag{1}$$

are the PDFs of the LN distributions.

Figure 1 shows the bimodality of the PDF of the DLN distribution for selected parameter values.

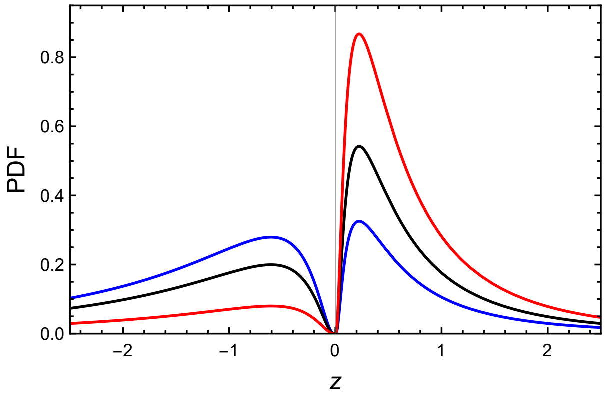

Figure 1. PDF of the DLN distribution: $(\beta, \mu_1, \sigma_1, \mu_2, \sigma_2)$: (0.3, −0.5, 1, 0.5, 1) (——), (0.5, −0.5, 1, 0.5, 1) (——), $(0.8, -0.5, 1, 0.5, 1)$(——).

The DLN distribution has two modes given by

$$\mathrm{Mode}(Z) = -\ \mathrm{Mode}(X_2) \ \text{ and } \ \mathrm{Mode}(X_1),$$

where

$$\mathrm{Mode}(X_j) = e^{\mu_j - \sigma_j^2}, \quad j = 1, 2 \tag{2}$$

are the modes of the LN distributions.

2.2. Cumulative Distribution Function

The cumulative distribution function (CDF) of the DLN distribution is

$$F_Z(z) = P(Z \le z) = \begin{cases} \overline{\beta}\ \left[1 - F_{X_2}(|z|; \mu_2, \sigma_2)\right], & z < 0, \\ \overline{\beta} + \beta\ F_{X_1}(z; \mu_1, \sigma_1), & z \ge 0, \end{cases} \tag{3}$$

where

$$F_{X_j}(x; \mu_j, \sigma_j) = P(X_j \le x) = \Phi\left(\frac{\ln(x) - \mu_j}{\sigma_j}\right), \quad x > 0, \quad j = 1, 2$$

are the CDFs of the LN distributions and

$$\Phi(a) = P(Z \le a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2}\, dz, \qquad -\infty < a < \infty$$

is the CDF of the standard normal distribution.

Figure 2 shows the CDF of the DLN distribution for selected parameter values. We can observe that $F_Z(0) = \overline{\beta}$ and hence $F_Z(0)$ decreases as β increases.

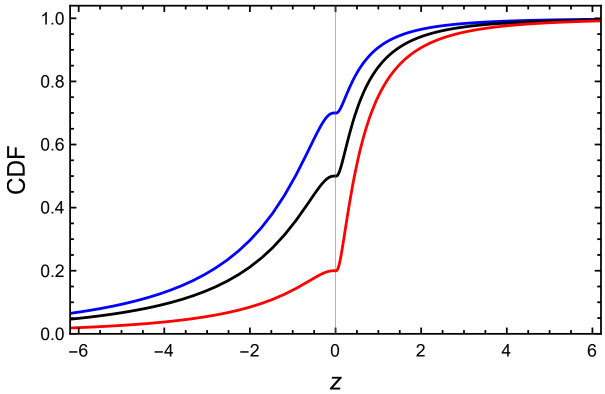

Figure 2. CDF of the DLN distribution: $(\beta, \mu_1, \sigma_1, \mu_2, \sigma_2)$: (0.3, −0.5, 1, 0.5, 1) (blue line), (0.5, −0.5, 1, 0.5, 1) (black line), $(0.8, -0.5, 1, 0.5, 1)$ (red line).

2.3. Hazard Rate Function

The survival function of the DLN distribution is

$$S_Z(z) = P(Z > z) = \begin{cases} 1 - \overline{\beta}\, S_{X_2}(|z|; \mu_2, \sigma_2), & z < 0, \\ \beta\, S_{X_1}(z; \mu_1, \sigma_1), & z \geq 0, \end{cases} \tag{4}$$

where

$$S_{X_j}(x; \mu_j, \sigma_j) = P(X_j > x) = 1 - \Phi\left(\frac{\ln(x) - \mu_j}{\sigma_j}\right), \quad x > 0, \quad j = 1, 2$$

are the SFs of the LN distributions.

The hazard rate function (HRF) of the DLN distribution is

$$h_Z(z) = \frac{f_Z(z)}{S_Z(z)} = \begin{cases} \dfrac{\overline{\beta}\, f_{X_2}(|z|; \mu_2, \sigma_2)}{1 - \overline{\beta}\, S_{X_2}(|z|; \mu_2, \sigma_2)}, & z < 0, \\ \dfrac{f_{X_1}(z)}{S_{X_1}(z; \mu_1, \sigma_1)}, & z \geq 0. \end{cases} \tag{5}$$

Figure 3 shows the HRF of the DLN distribution for selected parameter values. This figure shows that the HRF of the DLN distribution can be bimodal with one mode on each side of the origin.

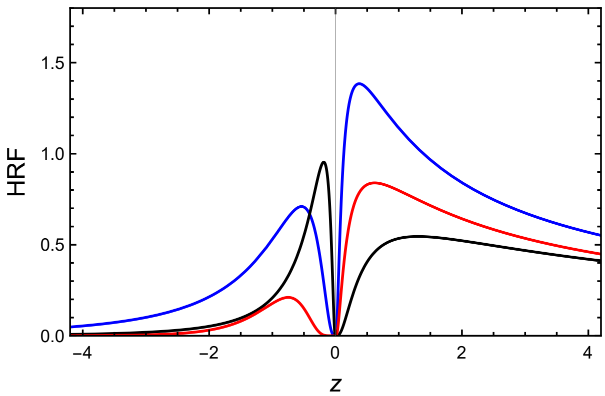

Figure 3. HRF of the DLN distribution: $(\beta, \mu_1, \sigma_1, \mu_2, \sigma_2)$: (0.3, −0.5, 1, 0.5, 0.9) (blue line), (0.5, 0.5, 0.9, −0.5, 1) (black line), $(0.8, 0, 1, 0, 0.5)$ (red line).

2.4. Moments and Associated Measures

The rth raw moment of the DLN distribution is

$$E(Z^r) = \beta\, E(X_1^r) + (-1)^r\, \overline{\beta}\, E(X_2^r), \qquad r \geq 1, \tag{6}$$

where

$$E\left(X_j^r\right) = e^{r\,\mu_j + r^2\,\sigma_j^2/2}, \qquad j = 1, 2 \tag{7}$$

are the rth moments of the LN distributions.

In particular, the first four raw moments of Z are

$$\begin{aligned}
E(Z) &= \beta\, e^{\mu_1+\sigma_1^2/2} - \overline{\beta}\, e^{\mu_2+\sigma_2^2/2}, \\
E\left(Z^2\right) &= \beta\, e^{2\mu_1+2\sigma_1^2} + \overline{\beta}\, e^{2\mu_2+2\sigma_2^2}, \\
E\left(Z^3\right) &= \beta\, e^{3\mu_1+9\sigma_1^2/2} - \overline{\beta}\, e^{3\mu_2+9\sigma_2^2/2}, \\
E\left(Z^4\right) &= \beta\, e^{4\mu_1+8\sigma_1^2} + \overline{\beta}\, e^{4\mu_2+8\sigma_2^2}.
\end{aligned}$$

The variance, skewness and kurtosis of the DLN distribution can be obtained using the well-known expressions:

$$\begin{aligned}
Variance(Z) &= E\left(Z^2\right) - [E(Z)]^2, \\
Skewness(Z) &= \frac{E(Z^3) - 3E(Z^2)E(Z) + 2[E(Z)]^3}{[Var(Z)]^{3/2}}, \\
Kurtosis(Z) &= \frac{E(Z^4) - 4E(Z^3)E(Z) + 6E(Z^2)[E(Z)]^2 - 3[E(Z)]^3}{[Var(Z)]^2}
\end{aligned}$$

upon substituting for the raw moments.

Figure 4 shows the mean, variance, skewness and kurtosis of the DLN distribution as a function of β for selected values of $(\mu_1, \sigma_1, \mu_2, \sigma_2)$. We can observe that the skewness can be negative or positive, i.e., the DLN distribution can be skewed to the left or skewed to the right.

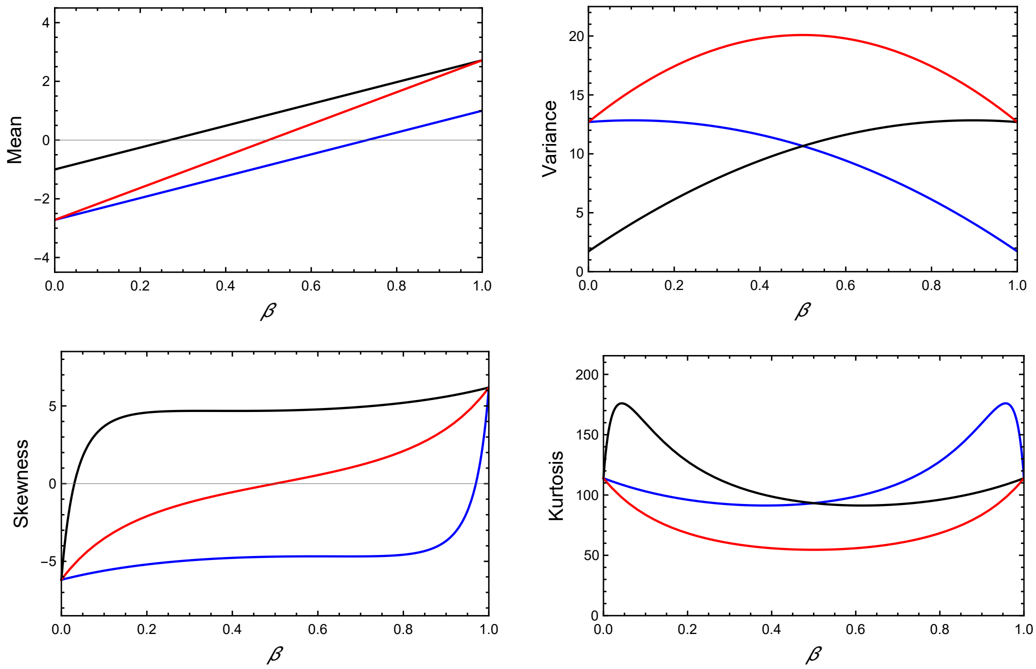

Figure 4. Mean, variance, skewness and kurtosis of the DLN distribution as a function of β: $(\mu_1, \sigma_1, \mu_2, \sigma_2) : (-0.5, 1, 0.5, 1)$(——), $(0.5, 1, -0.5, 1)$ (——), $(0.5, 1, 0.5, 1)$(——).

2.5. Harmonic Mean

The harmonic mean of an RV V is defined as

$$HM(V) = \frac{1}{E[1/V]},$$

provided $E[1/V]$ exists.

Proposition 1. *The harmonic mean of the RSMT Z is*

$$HM(Z) = \frac{1}{\frac{\beta}{HM(X_1)} - \frac{\overline{\beta}}{HM(X_2)}}.$$

Proof. Since

$$\begin{aligned} \frac{1}{HM(Z)} &= E[1/Z] \\ &= \int_0^\infty \frac{1}{z}\, \beta f_{X_1}(z) dz + \int_{-\infty}^0 \frac{1}{z}\, \overline{\beta}\, f_{X_2}(-z) dz \\ &= \beta\, E[1/X_1] - \overline{\beta}\, E[1/X_2] \\ &= \frac{\beta}{HM(X_1)} - \frac{\overline{\beta}}{HM(X_2)}, \end{aligned}$$

the proposition follows. □

Corollary 1. *The harmonic mean of the DLN distribution is*

$$HM(Z) = \frac{1}{\frac{\beta}{HM(X_1)} - \frac{\overline{\beta}}{HM(X_2)}},$$

where

$$HM(X_j) = e^{\mu_j - \sigma_j^2/2}, \ \ j = 1, 2$$

are the harmonic means of the LN distributions.

Figure 5 shows the harmonic mean of the DLN distribution as a function of β for selected parameter values.

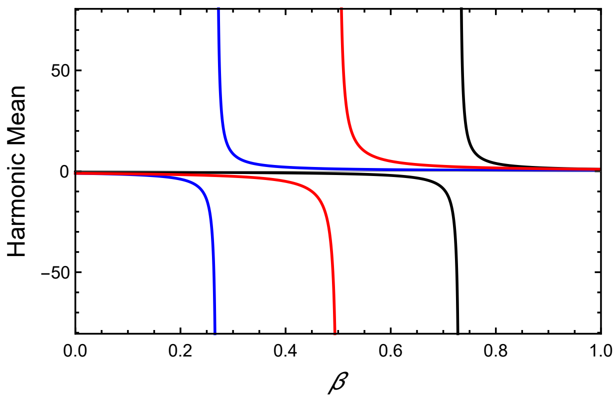

Figure 5. Harmonic mean of the DLN distribution as a function of β: $(\mu_1, \sigma_1, \mu_2, \sigma_2)$: $(-0.5, 1, 0.5, 1)$ (——), $(0.5, 1, -0.5, 1)$ (——), $(0.5, 1, 0.5, 1)$(——).

2.6. Entropies

Entropies are measures of a system's variation, instability or unpredictability. For an RV V with PDF $f_V(v)$, the following are two well-known entropies:

1. Tsallis entropy [17]:

$$T_\alpha(V) = \frac{1}{\alpha - 1}\left\{1 - E\left[f_V^{\alpha-1}(V)\right]\right\}, \quad 0 < \alpha \neq 1.$$

2. Shannon entropy [18]:

$$H(V) = -E[\ln f_V(V)] = \lim_{\alpha \to 1} T_\alpha(V).$$

Proposition 2. *The Tsallis entropy of the RSMT Z is*

$$T_\alpha(Z) = T_\alpha(Y) + \beta^\alpha \, T_\alpha(X_1) + \overline{\beta}^{\,\alpha} \, T_\alpha(X_2)$$

for $0 < \alpha \neq 1$, *where*

$$T_\alpha(Y) = \frac{1 - \beta^\alpha - \overline{\beta}^{\,\alpha}}{\alpha - 1}.$$

Proof. See [16]. □

Corollary 2. *The Shannon entropy of the RSMT Z is*

$$H(Z) = \lim_{\alpha \to 1} T_\alpha(Z) = H(Y) + \beta \, H(X_1) + \overline{\beta} \, H(X_2),$$

where

$$H(Y) = \lim_{\alpha \to 1} T_\alpha(Y) = -\beta \ln(\beta) - \overline{\beta} \ln(\overline{\beta}).$$

Proposition 3. *The Tsallis entropy of the LN distribution with parameters* (μ, σ) *is*

$$T_\alpha(X) = \frac{1}{\alpha-1}\left\{1 - \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{\alpha-1} \frac{1}{\sqrt{\alpha}} \exp\left[\mu(1-\alpha) + \frac{\sigma^2}{2\alpha}(1-\alpha)^2\right]\right\}$$

for $0 < \alpha \neq 1$.

Proof. Since

$$\begin{aligned} 1-(\alpha-1)T_\alpha(X) &= \int_0^\infty f_X^\alpha(x)dx \\ &= \int_0^\infty \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^\alpha \frac{1}{x^\alpha} \exp\left\{-\frac{\alpha}{2\sigma^2}[\ln(x)-\mu]^2\right\}dx \\ &= \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^\alpha \int_{-\infty}^\infty e^{(1-\alpha)y} \exp\left[-\frac{1}{2(\sigma/\sqrt{\alpha})^2}(y-\mu)^2\right]dy \\ &= \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^{\alpha-1} \frac{1}{\sqrt{\alpha}} \exp\left[\mu(1-\alpha) + \frac{\sigma^2}{2\alpha}(1-\alpha)^2\right], \end{aligned}$$

the proposition follows. □

Corollary 3. *The Shannon entropy of the LN distribution with parameters* (μ, σ) *is*

$$H(X) = \lim_{\alpha\to 1} T_\alpha(X) = \frac{1}{2} + \mu + \ln\left(\sqrt{2\pi}\,\sigma\right).$$

Proposition 4. *The Tsallis entropy of* $Z \sim \mathrm{DLN}(\beta, \mu_1, \sigma_1, \mu_2, \sigma_2)$ *is*

$$T_\alpha(Z) = T_\alpha(Y) + \beta^\alpha\, T_\alpha(X_1) + \overline{\beta}^{\,\alpha}\, T_\alpha(X_2)$$

for $0 < \alpha \neq 1$, *where*

$$T_\alpha(Y) = \frac{1-\beta^\alpha - \overline{\beta}^{\,\alpha}}{\alpha-1}$$

and

$$T_\alpha(X_j) = \frac{1}{\alpha-1}\left\{1 - \left(\frac{1}{\sqrt{2\pi}\,\sigma_j}\right)^{\alpha-1} \frac{1}{\sqrt{\alpha}} \exp\left[\mu_j(1-\alpha) + \frac{\sigma_j^2}{2\alpha}(1-\alpha)^2\right]\right\}, \quad j = 1, 2.$$

The proof of Proposition 4 follows directly from Propositions 2 and 3.

Corollary 4. *The Shannon entropy of* $Z \sim \mathrm{DLN}(\beta, \mu_1, \sigma_1, \mu_2, \sigma_2)$ *is*

$$H(Z) = H(Y) + \beta\, H(X_1) + \overline{\beta}\, H(X_2),$$

where

$$H(Y) = -\beta\, \ln(\beta) - \overline{\beta}\, \ln(\overline{\beta})$$

and

$$H(X_j) = \frac{1}{2} + \mu_j + \ln\left(\sqrt{2\pi}\,\sigma_j\right), \qquad j = 1, 2.$$

Figure 6 shows the Tsallis and Shannon entropies of the DLN distribution as a function of β for selected parameter values.

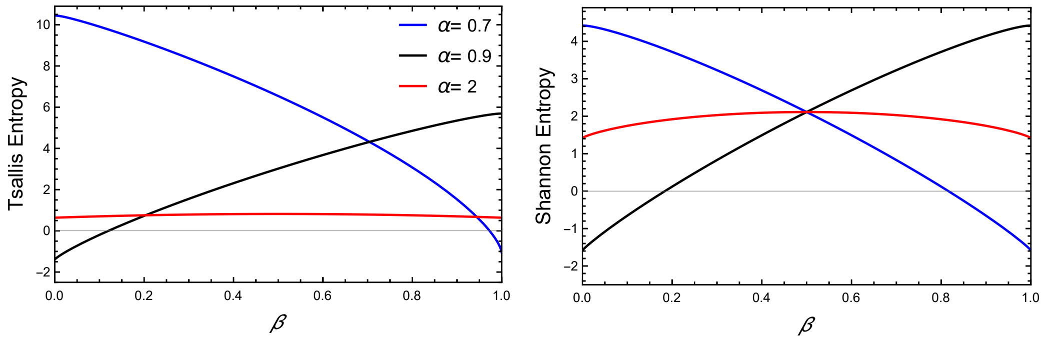

Figure 6. Tsallis and Shannon entropies of the DLN distribution as a function of β: $(\mu_1, \sigma_1, \mu_2, \sigma_2)$: $(-3, 1, 3, 1)$(——), (3, 1, −3, 1) (——), (0, 1, 0, 1) (——).

Note that the Tsallis and Shannon entropies can be negative for continuous distributions.

3. Maximum Likelihood Estimation

In this section, MLEs of the parameters of the DLN distribution and their asymptotic distributions are derived.

Let $z_1, z_2, \ldots, z_n$ be a random sample from the DLN$(\beta, \mu_1, \sigma_1, \mu_2, \sigma_2)$ distribution. The log-likelihood function is

$$\ln L(\beta, \mu_1, \sigma_1, \mu_2, \sigma_2) = \sum_{i=1}^{n} \ln\big[\beta\, f_{X_1}(z_i; \mu_1, \sigma_1)\big]\, \mathbf{1}_{\{z_i>0\}} + \sum_{i=1}^{n} \ln\big[\overline{\beta}\, f_{X_2}(|z_i|; \mu_2, \sigma_2)\big]\, \mathbf{1}_{\{z_i<0\}},$$

where $\mathbf{1}_A$ denotes the indicator function. The MLEs of the parameters $(\beta, \mu_1, \sigma_1, \mu_2, \sigma_2)$ are:

$$\begin{aligned}
\widehat{\beta} &= \frac{n_1}{n}, \\
\widehat{\mu}_1 &= \frac{1}{n_1} \sum_{i=1}^{n} \ln(z_i) \mathbf{1}_{(z_i>0)}, \\
\widehat{\sigma}_1 &= \sqrt{\frac{1}{n_1} \sum_{i=1}^{n} [\ln(z_i) - \widehat{\mu}_1]^2 \mathbf{1}_{(z_i>0)}}, \\
\widehat{\mu}_2 &= \frac{1}{n_2} \sum_{i=1}^{n} \ln(|z_i|)\, 1_{(z_i<0)}, \\
\widehat{\sigma}_2 &= \sqrt{\frac{1}{n_2} \sum_{i=1}^{n} [\ln(|z_i|) - \widehat{\mu}_2]^2 \mathbf{1}_{(z_i<0)}},
\end{aligned}$$

where

$$n_1 = \sum_{i=1}^{n} 1_{(z_i>0)}, \qquad n_2 = \sum_{i=1}^{n} 1_{(z_i<0)}, \qquad n_1 + n_2 = n.$$

The Fisher information matrix about $(\beta, \mu_1, \sigma_1, \mu_2, \sigma_2)$ is

$$\mathbf{I}(\beta, \mu_1, \sigma_1, \mu_2, \sigma_2) = \begin{bmatrix} I_Y(\beta) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \beta\, \mathbf{I}_{X_1}(\mu_1, \sigma_1) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \overline{\beta}\, \mathbf{I}_{X_2}(\mu_2, \sigma_2) \end{bmatrix},$$

where $I_Y(\beta) = \frac{1}{\beta\,\overline{\beta}}$ is the Fisher information matrix about β and $\mathbf{I}_{X_j}(\mu_j, \sigma_j) = \mathrm{diag}(\frac{1}{\sigma_j^2}, \frac{2}{\sigma_j^2})$, $j = 1, 2$ is the Fisher information matrix about (μ_j, σ_j).

Moreover, the asymptotic distribution of the MLEs as $n \to \infty$ is

$$\sqrt{n}\begin{bmatrix} \widehat{\beta}-\beta \\ \widehat{\mu}_1-\mu_1 \\ \widehat{\sigma}_1-\sigma_1 \\ \widehat{\mu}_2-\mu_2 \\ \widehat{\sigma}_2-\sigma_2 \end{bmatrix} \xrightarrow{d} MVN\Big(\mathbf{0}, \mathbf{I}^{-1}(\beta,\mu_1,\sigma_1,\mu_2,\sigma_2)\Big),$$

where $\xrightarrow{d}$ denotes *convergence in distribution*, MVN stands for *multivariate normal distribution* and

$$\mathbf{I}^{-1}(\beta,\mu_1,\sigma_1,\mu_2,\sigma_2) = \mathrm{diag}\left(\beta\bar{\beta}, \frac{\sigma_1^2}{\beta}, \frac{\sigma_1^2}{2\beta}, \frac{\sigma_2^2}{\bar{\beta}}, \frac{\sigma_2^2}{2\bar{\beta}}\right).$$

4. Simulations

This section details simulations to check the finite sample performance of the MLEs of the parameters of the DLN distribution. The performance is evaluated in terms of biases, mean squared errors of the MLEs and coverage probabilities of the corresponding 95% confidence intervals.

The simulation was repeated $M = 10{,}000$ times. In each of the M repetitions, a random sample of size $n = 50, 100, \ldots, 500$ was drawn from the DLN distribution with selected parameter values $(\beta,\mu_1,\sigma_1,\mu_2,\sigma_2) = (0.3,-2,1,-1,2)$, $(0.5,0,1,1,2)$, $(0.8,2,1,-1,2)$ and $(0.547,-2.812,1.016,-2.224,0.764)$, using the following algorithm:

1. Generate $Y_i \sim Bernoulli(\beta),\ i = 1,2,\ldots,n;$
2. Generate $X_{1,i} \sim \mathrm{LN}(\mu_1,\lambda_1),\ i = 1,2,\ldots,n;$
3. Generate $X_{2,i} \sim \mathrm{LN}(\mu_2,\lambda_2),\ i = 1,2,\ldots,n;$
4. Set $Z_i = Y_i\, X_{1,i} - (1-Y_i)\, X_{2,i},\ i = 1,2,\ldots,n.$

The parameter values $(\beta,\mu_1,\sigma_1,\mu_2,\sigma_2) = (0.547,-2.812,1.016,-2.224,0.764)$ are those estimated in the real data application in Section 5.

The measures examined in this simulation study are:

1. The bias of the MLEs:

$$Bias\left(\widehat{\theta}\right) = \frac{1}{M}\sum_{j=1}^{M}\left(\widehat{\theta}_j-\theta\right), \qquad \theta = \beta,\mu_1,\sigma_1,\mu_2,\sigma_2.$$

2. The mean squared error (MSE) of the MLEs:

$$MSE\left(\widehat{\theta}\right) = \frac{1}{M}\sum_{j=1}^{M}\left(\widehat{\theta}_j-\theta\right)^2, \qquad \theta = \beta,\mu_1,\sigma_1,\mu_2,\sigma_2.$$

3. The coverage probability (CP) of the 95% confidence interval of each parameter:

$$CP(\theta) = \frac{1}{M}\sum_{j=1}^{M}\mathbf{1}_{\left\{\left(\widehat{\theta}_j-1.96\,S.E.\left(\widehat{\theta}_j\right),\ \widehat{\theta}_j+1.96\,S.E.\left(\widehat{\theta}_j\right)\right)\right\}}, \qquad \theta = \beta,\mu_1,\sigma_1,\mu_2,\sigma_2.$$

The results of the simulation study are reported in Figures 7–9.

1. Figure 7 shows that the *absolute* biases of the MLEs are small and approach zero as n increases.
2. Figure 8 shows that the MSEs of the MLEs are small and decrease as n increases.
3. Figure 9 shows that the coverage probabilities of the 95% confidence intervals are close to the nominal level.

These conclusions show that the MLEs of the DLN distribution are well behaved for point as well as interval estimation.

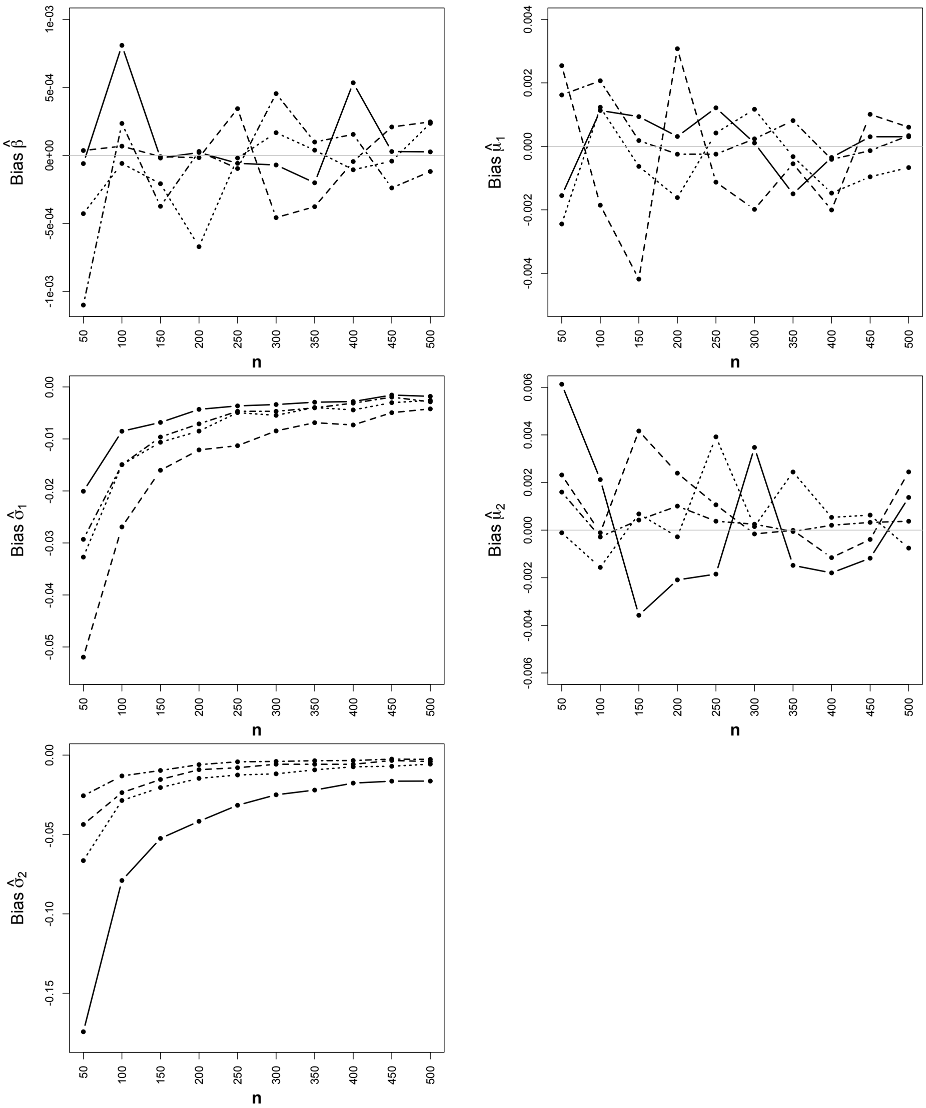

Figure 7. Bias of the MLEs of the parameters of the DLN distribution: $(\beta, \mu_1, \sigma_1, \mu_2, \sigma_2)$: $(0.3, -2, 1, -1, 2)$ (___), $(0.5, 0, 1, 1, 2)$ (- - -), $(0.8, 2, 1, -1, 2)$ (.....), $(0.547, -2.812, 1.016, -2.224, 0.764)$ $(-.-)$.

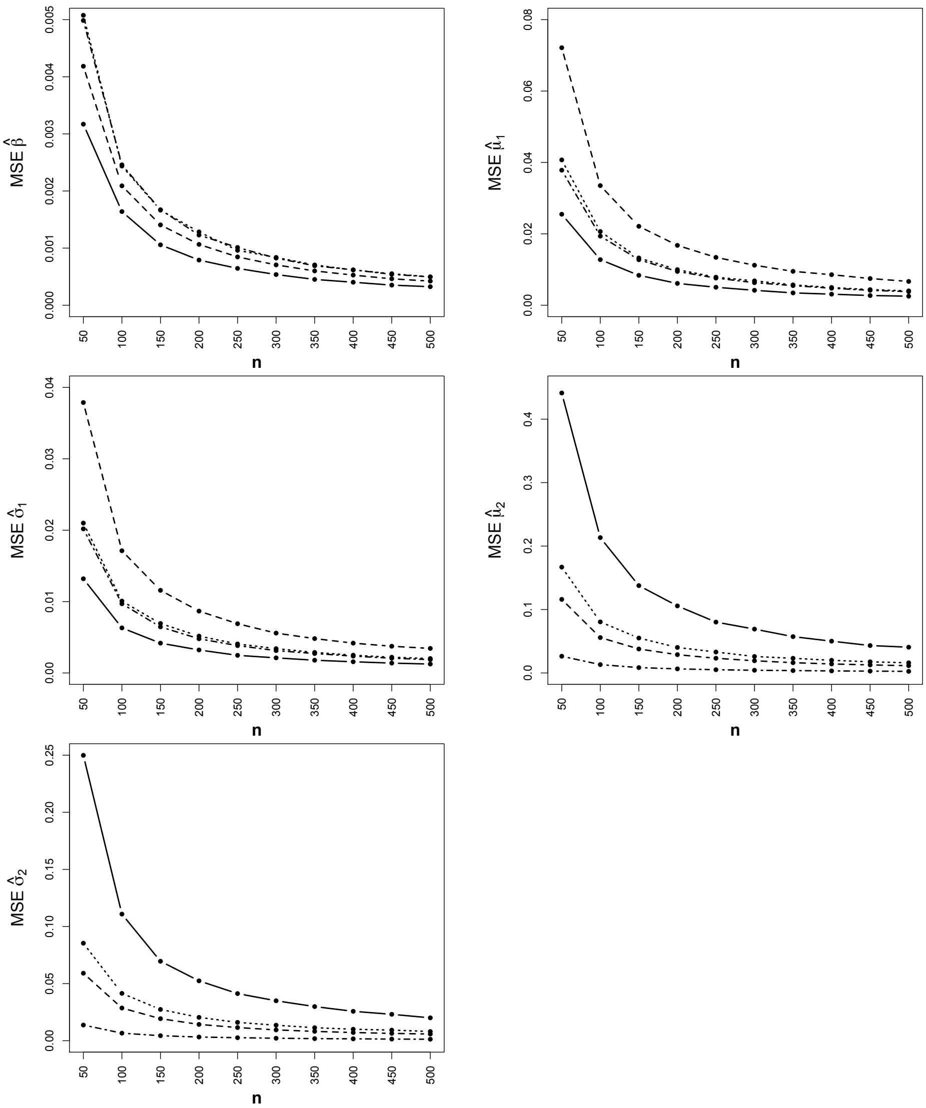

Figure 8. MSE of the MLEs of the parameters of the DLN distribution: $(\beta, \mu_1, \sigma_1, \mu_2, \sigma_2)$: $(0.3, -2, 1, -1, 2)$ (___), (0.5, 0, 1, 1, 2) (- - -), $(0.8, 2, 1, -1, 2)$ (.....), (0.547, −2.812, 1.016, −2.224, 0.764) (−.−).

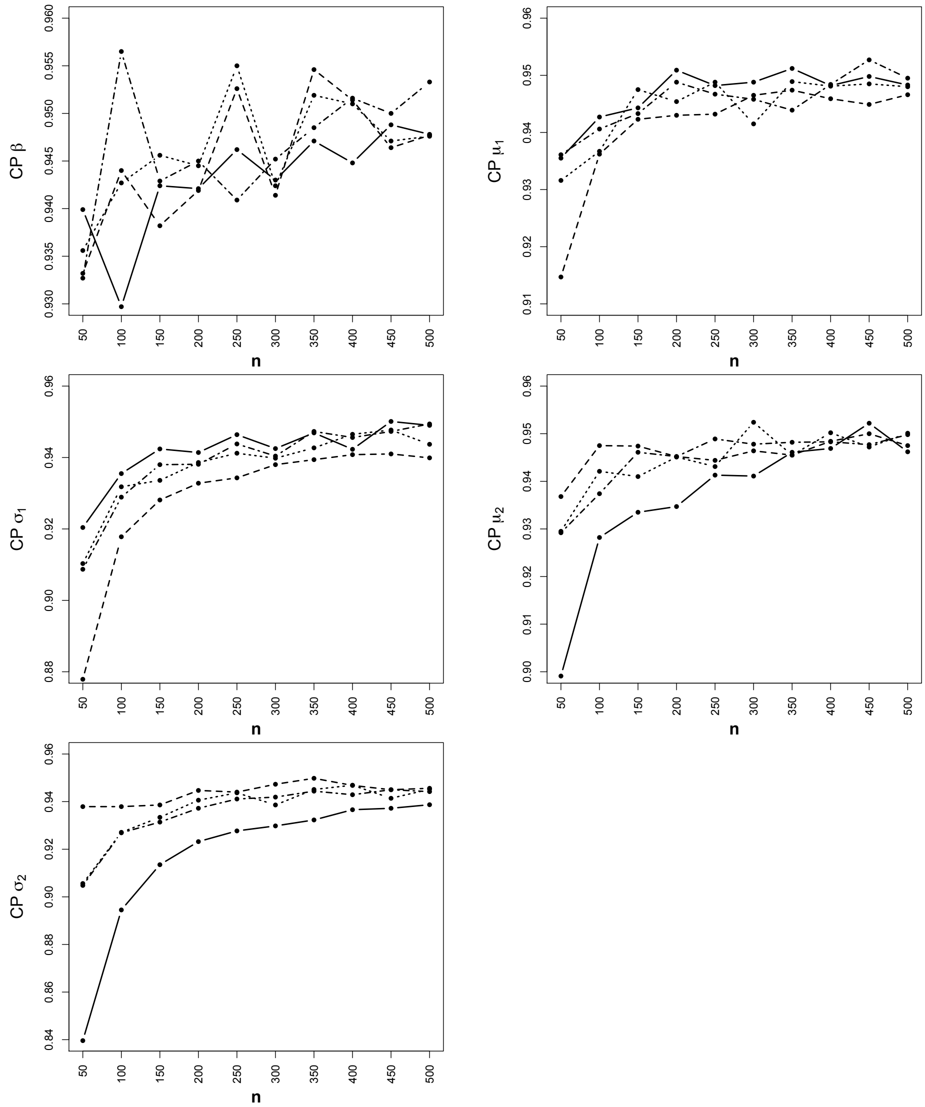

Figure 9. CP of the 95% confidence intervals of the parameters of the DLN distribution: $(\beta, \mu_1, \sigma_1, \mu_2, \sigma_2)$: (0.3, −2, 1, −1, 2) (___), (0.5, 0, 1, 1, 2) (- - -), (0.8, 2, 1, −1, 2) (.....), (0.547, −2.812, 1.016, −2.224, 0.764) (−.−).

5. Application

In this section, we apply the proposed DLN distribution to a real data set from a DNA microarray reported in [19]. According to Wikipedia, "A DNA microarray (also commonly known as DNA chip or biochip) is a collection of microscopic DNA spots attached to a solid surface. Scientists use DNA microarrays to measure the expression levels of large numbers of genes simultaneously or to genotype multiple regions of a genome". The data labelled as "SID 377353, ESTs [5':, 3':AA055048]" consist of the following 118 observations: 0.029, 0.062, 0.011, 0.009, 0.065, −0.128, 0.133, 0.116, 0.184, 0.111, −0.066, −0.049, 0.05, 0.137, 0.162, 0.173, 0.033, 0.107, 0.11, 0.147, 0.118, 0.172, 0.284, −0.137, 0.038, −0.145, −0.181, −0.155, 0.198, 0.024, 0.079, −0.252, 0.062, 0.097, 0.032, 0.026, 0.195, 0.019, 0.138, −0.3, −0.105, −0.11, −0.168, −0.173, −0.15, 0.078, 0.113, −0.047, 0.024, 0.001, −0.075, 0.014, 0.058, −0.083, −0.339, −0.177, −0.073, −0.044, −0.106, −0.159, −0.101, −0.074, −0.126, −0.131, −0.22, −0.184, −0.105, 0.173, 0.151, 0.064, −0.007, −0.005, −0.189, −0.219, −0.301, −0.212, −0.088, 0.157, 0.042, 0.184, 0.114, 0.102, 0.119, −0.064, −0.075, 0.073, 0.038, 0.017, −0.134, −0.118, −0.097, 0.059, 0.025, −0.102, −0.096, −0.035, 0.057, −0.055, 0.015, −0.23, −0.115, 0.255, 0.034, 0.078, 0.129, 0.081, 0.032, 0.047, −0.145, 0.012, −0.224, 0.074, −0.06, −0.137, 0.034, 0.009, −0.139, −0.141.

Figure 10 shows the histogram of the data, which indicates bimodality around the origin.

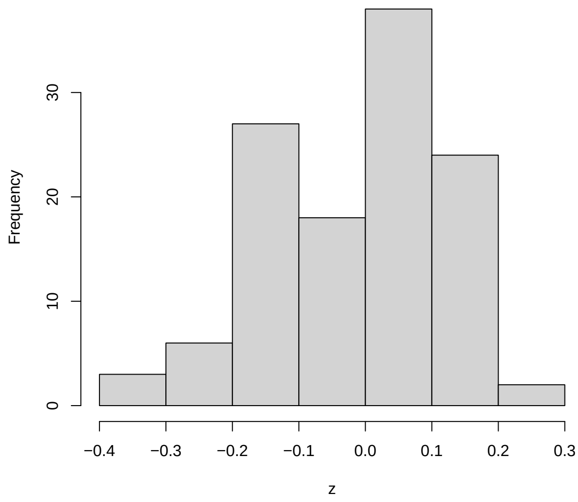

Figure 10. Histogram of the microarray data.

For the sake of comparing the bimodal DLN distribution with other bimodal distributions, we consider the double inverse Gaussian (DIG) distribution proposed in [16]. The PDF of the DIG distribution is

$$f_Z(z) = \begin{cases} \overline{\beta}\, f_{X_2}(|z|;\nu_2,\lambda_2), & z < 0, \\ \beta\ \, f_{X_1}(z;\nu_1,\lambda_1), & z \geq 0, \end{cases} \tag{8}$$

where

$$f_{X_j}(x;\nu_j,\lambda_j) = \sqrt{\frac{\lambda_j}{2\pi}}\, x^{-3/2}\ \exp\left[-\frac{\lambda_j (x-\nu_j)^2}{2\nu_j^2 x}\right], \qquad x > 0, \quad \nu_j,\lambda_j > 0, \quad j = 1,2 \tag{9}$$

are the PDFs of inverse Gaussian distributions.

Table 1 gives the MLEs, their standard errors (S.E.s), estimated log-likelihoods and Kolmogrov–Smirnov (KS), Anderson–Darling (AD) and Cramér–von Mises (CVM) goodness-of-fit tests of the fitted DIG and DLN distributions. This table shows that the MLE of β and its S.E. are the same for both the fitted DIG and DLN distributions, since the Bernoulli parameter β is estimated independently in the RSMT. In addition, this table shows that the MLEs of μ_1 and μ_2 in the fitted DLN distribution are both negative.

Table 1. Summary of the fitted DIG and DLN distributions for DNA microarray data.

Model	Parameter	MLE	S.E.	$\ln \widehat{L}$	KS (p-Value)	AD (p-Value)	CVM (p-Value)
DIG	β	0.542	0.046	39.249	0.126 (0.046)	3.285 (0.020)	0.545 (0.030)
	ν_1	0.087	0.017				
	λ_1	0.036	0.006				
	ν_2	0.132	0.018				
	λ_2	0.126	0.024				
DLN	β	0.542	0.046	64.829	0.065 (0.709)	0.851 (0.446)	0.103 (0.570)
	μ_1	−2.812	0.127				
	σ_1	1.016	0.090				
	μ_2	−2.224	0.104				
	σ_2	0.764	0.074				

Table 1 shows that the three goodness-of-fit tests have much smaller (larger) test statistics for the fitted DLN (DIG) distribution. This table also shows that the three goodness-of-fit tests reject (accept) the DIG (DLN) distribution for the given data. This conclusion is supported by the diagnostic plots in Figures 11 and 12. In these figures, (i) the PDF and CDF plots indicate, in an informal way, that the fitted DIG (DLN) distribution may not be suitable for the given data; (ii) the quantile–quantile (Q–Q) plots show that the fitted DIG and DLN distributions inappropriately describe the tails of the distributions; (iii) the probability–probability (P–P) plots show that the fitted DIG (DLN) distribution inappropriately (appropriately) describes the center of the distribution.

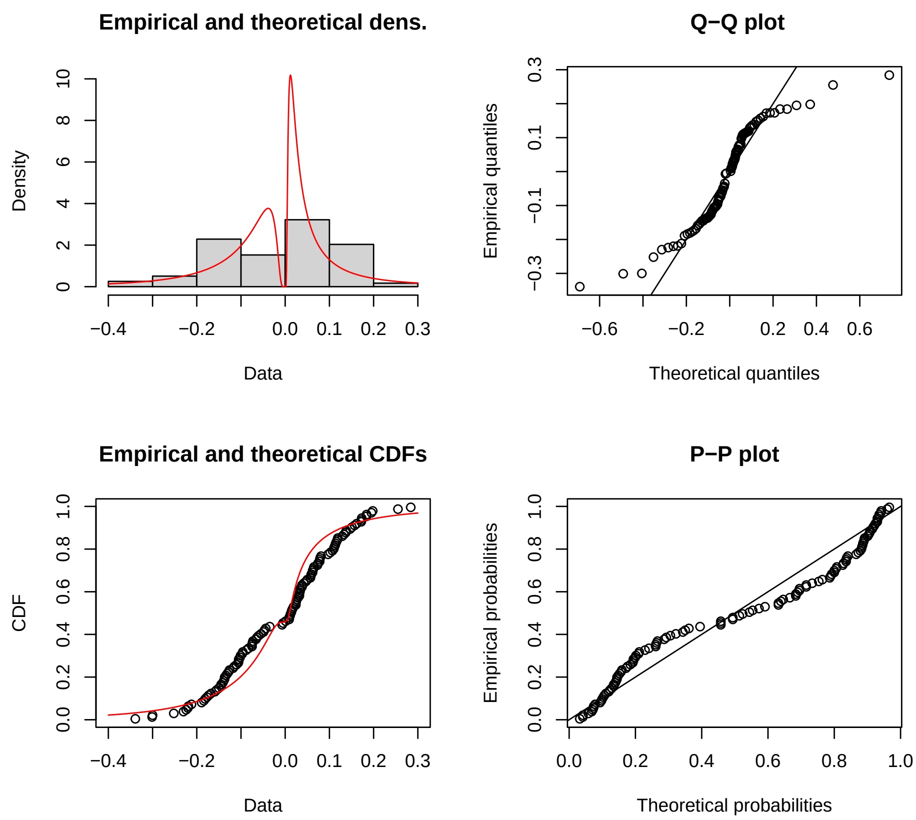

Figure 11. Diagnostic plots of the fitted DIG distribution.

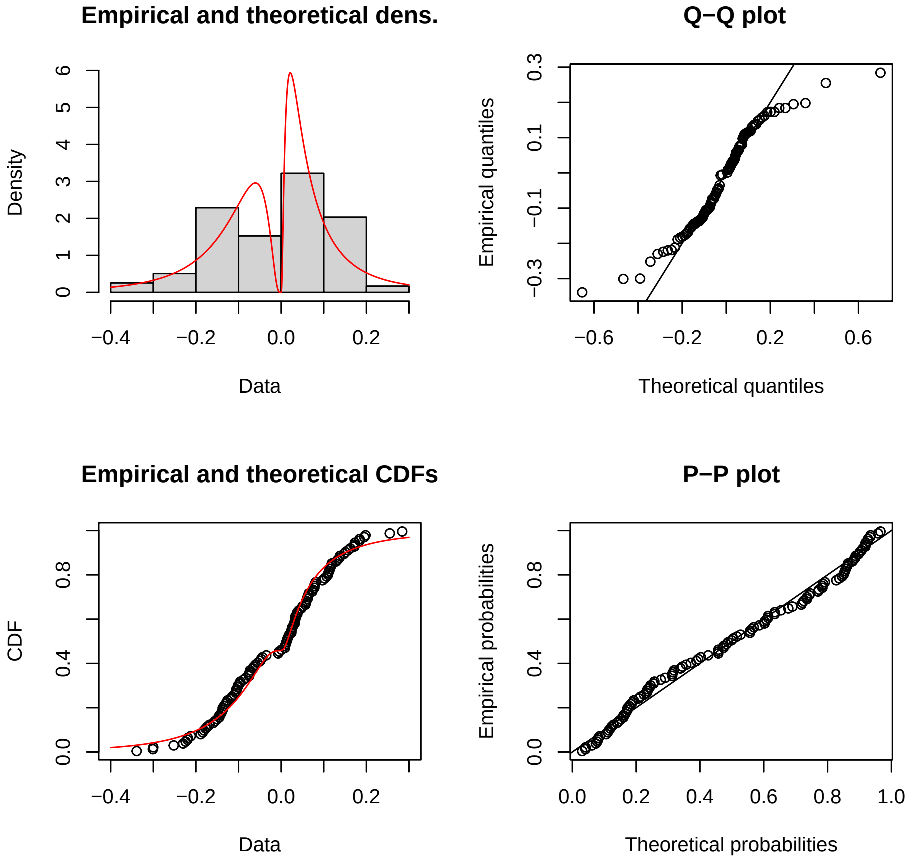

Figure 12. Diagnostic plots of the fitted DLN distribution.

6. Conclusions and Comments

We have proposed a bimodal distribution on the real line, referred to as the double log-normal distribution. We have derived its statistical properties, including the probability density, cumulative distribution and hazard rate functions, the moments and associated measures and harmonic mean, as well as Tsallis and Shannon entropies. Additionally, maximum likelihood estimates of the parameters and their asymptotic distribution are provided. Simulation studies showed that the maximum likelihood estimation performed well in terms of the bias, mean squared error and coverage probability of confidence intervals. Application to a DNA microarray data set showed that the proposed distribution is flexible and competitive for modeling bimodal data around the origin.

Instead of the log-normal distribution, one can consider the length biased log-normal distribution developed in [20]. It will be interesting to formulate a double length biased log-normal distribution.

Author Contributions: Conceptualization, M.E.G.; Methodology, M.F.A. and M.E.G.; Software, M.F.A. and A.N.A.; Validation, M.F.A., A.N.A. and S.N.; Formal analysis, M.F.A.; Data curation, M.F.A. and A.N.A.; Writing—original draft, M.E.G.; Writing—review & editing, S.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are given in the paper. The code used can be obtained from the corresponding author.

Acknowledgments: The authors would like to thank the Editor and the three referees for careful reading and comments which greatly improved the paper.

Conflicts of Interest: The authors have no conflict of interest.

References

1. Crow, E.L.; Shimizu, K. *Lognormal Distributions: Theory and Applications*; Statistics Textbooks and Monographs; Routledge: London, UK, 2018.
2. Aly, E. A unified approach for developing Laplace-type distributions. *J. Indian Soc. Probab. Stat.* **2018**, *19*, 245–269. [CrossRef]
3. Laplace, P.S. Memoire sur la probabilite des causes par les evenements. *Mem. L'Academie R. Des Sci. Present. Par Divers. Savan* **1774**, *6*, 621–656.
4. Plucinska, A. On a general form of the probability density function and its application to the investigation of the distribution of rheostat resistence. *Zastosow. Mat.* **1966**, *9*, 9–19.
5. Balakrishnan, N.; Kocherlakota, S. On the double Weibull distribution: Order statistics and estimation. *Sankhya Indian J. Stat. Ser. B* **1985**, *47*, 161–178.
6. Kantam, R.R.L.; Narasimham, V.L. Linear estimation in reflected gamma distribution. *Sankhya Indian J. Stat. Ser. B* **1991**, *53*, 25–47.
7. Nadarajah, S.; Afuecheta, E.; Chan, S. A double generalized Pareto distribution. *Stat. Probab. Lett.* **2013**, *83*, 2656–2663. [CrossRef]
8. Bindu, P.; Sangita, K. Double Lomax distribution and its applications. *Statistica* **2015**, *75*, 331–342.
9. Kumar, S.C.; Jose, R. On double Lindley distribution and some of its properties. *Am. J. Math. Manag. Sci.* **2019**, *38*, 23–43.
10. John, S. The three-parameter two-piece normal family of distributions and its fitting. *Commun. Stat. Theory Methods* **1982**, *11*, 879–885. [CrossRef]
11. Kimber, A. Methods for the two-piece normal distribution. *Commun. Stat. Theory Methods* **1985**, *14*, 235–245. [CrossRef]
12. Lingappaiah, G. On two-piece double exponential distribution. *J. Korean Stat. Soc.* **1988**, *17*, 46–55.
13. Abdulah, E.; Elsalloukh, H. Bimodal Class based on the Inverted Symmetrized Gamma Distribution with Applications. *J. Stat. Appl. Probab.* **2014**, *3*, 1–7. [CrossRef]
14. Hoseinzadeh, A.; Maleki, M.; Khodadadi, Z.; Contreras-Reyes, J.E. The skew-reflected-Gompertz distribution for analyzing symmetric and asymmetric data. *J. Comput. Appl. Math.* **2019**, *349*, 132–141. [CrossRef]
15. Halvarsson, D. Maximum likelihood estimation of asymmetric double type II Pareto distributions. *J. Stat. Theory Pract.* **2020**, *14*, 22. [CrossRef]
16. Almutairi, A.; Ghitany, M.; Alothman, A.; Gupta, R.C. Double Inverse-Gaussian Distributions and Associated Inference. *J. Indian Soc. Probab. Stat.* **2023**, 1–32. [CrossRef]
17. Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* **1988**, *52*, 479–487. [CrossRef]
18. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]
19. Cankaya, M. Asymmetric Bimodal Exponential Power Distribution on the Real Line. *Entropy* **2018**, *20*, 23. [CrossRef] [PubMed]
20. Ratnaparkhi, M.V.; Naik-Nimbalkar, U.V. The Length-biased Lognormal Distribution and its application in the Analysis of data from oil Field Exploration studies. *J. Mod. Appl. Stat. Methods* **2012**, *11*, 225–260. [CrossRef]