mathematics

MDPI

Article

An Alternative to the Log-Skew-Normal Distribution: Properties, Inference, and an Application to Air Pollutant Concentrations

Jaime Arrué [1], Reinaldo Boris Arellano-Valle [2], Osvaldo Venegas [3,*], Heleno Bolfarine [4] and Héctor W. Gómez [1]

1 Departamento de Matemáticas, Facultad de Ciencias Básicas, Universidad de Antofagasta, Antofagasta 1240000, Chile
2 Departamento de Estadística, Facultad de Matemáticas, Pontificia Universidad Católica de Chile, Santiago 7820436, Chile
3 Departamento de Ciencias Matemáticas y Físicas, Facultad de Ingeniería, Universidad Católica de Temuco, Temuco 4780000, Chile
4 Departamento de Estatítica, IME, Universidade de São Paulo, São Paulo 05508-090, Brazil
* Correspondence: ovenegas@uct.cl; Tel.: +54-45-220-5351

check for updates

Citation: Arrué, J.; Arellano-Valle, R.B.; Venegas, O.; Bolfarine, H; Gómez, H.W. An Alternative to the Log-Skew-Normal Distribution: Properties, Inference, and an Application to Air Pollutant Concentrations. *Mathematics* **2022**, *10*, 4336. https://doi.org/10.3390/math10224336

Academic Editor: Leonid V. Bogachev

Received: 28 September 2022
Accepted: 10 November 2022
Published: 18 November 2022

Abstract: In this study, we consider an alternative to the log-skew-normal distribution. It is called the modified log-skew-normal distribution and introduces greater flexibility in the skewness and kurtosis parameters. We first study several of the main probabilistic properties of the new distribution, such as the computation of its moments and the non-existence of the moment-generating function. Parameter estimation by the maximum likelihood approach is also studied. This approach presents an overestimation problem in the shape parameter, which in some cases, can even be infinite. However, as we demonstrate, this problem is solved by adapting bias reduction using Firth's approach. We also show that the modified log-skew-normal model likewise inherits the non-singularity of the Fisher information matrix of the untransformed model, when the shape parameter is null. Finally, we apply the modified log-skew-normal model to a real example related to pollution data.

Keywords: log-normal distribution; non-singular information matrix; modified likelihood; modified score; bias prevention

MSC: 62E15; 62E20

1. Introduction

The log-normal distribution is commonly used to model the behavior of data with positive asymmetry, in which most of the observations are concentrated near the minimum value. Some applications of the log-normal model are in species abundance patterns, environmental concentrations, stock prices, the distribution of the molecular weights of polymers, the production of copper nano-particles, etc.

The log-normal (LN) distribution arises from the transformation $Y = e^X$, where $X \sim N(\mu, \sigma^2)$ is the normal distribution with mean μ and variance σ^2 and has the density given by

$$f_Y(y) = \frac{1}{\sigma y}\phi\left(\frac{\log y - \mu}{\sigma}\right), \quad y > 0, \tag{1}$$

where $\phi(\cdot)$ is the standard normal density. The notation $Y \sim LN(\mu, \sigma^2)$ is typically used for the log-normal distribution with parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. More simply, we can say that a random variable $Y \sim LN(\mu, \sigma^2)$ if and only if $\log Y \sim N(\mu, \sigma^2)$. As noted by Jones and Arnold [1], this distribution is both log-symmetric about its mean/median

$E(Y) = e^{\mu}$ and R-symmetric about its mode $\delta = e^{\mu+\sigma^2}$, a condition called double symmetry. As defined by Mudholkar and Wang [2], in density terms, a non-negative random variable Y is said to be R-symmetric about δ if $f_Y(\delta y) = f_Y(\delta/y)$ for some $\delta > 0$ and all $y > 0$, which is equivalent to $\delta Y \overset{d}{=} \delta/Y$, where "$\overset{d}{=}$" means equal in distribution. When f_Y is unimodal, then δ is the (unique) mode of Y, which is less than the mean of Y. Thus, R-symmetric distributions are always positively skewed. Despite these good properties, there are data that are not adequately modeled by the log-normal distribution, since they have symmetry and kurtosis indices that are outside their natural range. A model that presents this characteristic is the log-skew-normal (LSN) distribution, introduced and studied by Azzalini et al. [3]. It is a version with positive support of the skew-normal (SN) distribution, defined as follows: for a skew-normal random variable $Z \sim SN(\lambda)$, where $\lambda \in \mathbb{R}$ is the skewness parameter, we say that the random variable $Y = e^Z$ follows a log-skew-normal distribution with skewness parameter λ, denoted by $Y \sim LSN(\lambda)$, if its probability density function (pdf) is given by

$$f_Y(y;\lambda) = \frac{2}{y}\phi(\log(y))\Phi(\lambda \log(y)), \quad y > 0,$$

where $\Phi(\cdot)$ denotes the cumulative distribution function (cdf) of the standard $\mathrm{N}(0,1)$.

Applications of this model to real datasets are reported in Azzalini et al. [3], Marchenko and Genton [4], and Bolfarine et al. [5], where a bimodal version of the model was used to fit a real (bimodal) dataset. Chai and Bailey [6] extended the SN model to a situation of continuous datasets with a discrete component at zero. On the other hand, the modified skew-normal (MSN) distribution is a particular case of the generalized skew-normal (GSN) distribution introduced by Arellano-Valle et al. [7], for $\lambda_2 = 1$. Later, Arellano-Valle et al. [8] investigated the Fisher information matrix (FIM) for the location-scale version of the GSN model. Thus, the MSN model is a fair competitor to the SN model since for both control asymmetry with a single scalar parameter, say $\lambda \in \mathbb{R}$, such that, if $\lambda = 0$, then the ordinary normal model results. Moreover, one advantage of the MSN model over the SN model in location-scale situations is that its FIM is nonsingular at $\lambda = 0$, which is not the case with the SN model (see Arellano-Valle et al. [8] and Arrué et al. [9]). The present paper focuses on investigating the possibility of developing a more flexible distribution for positive data, such that the regularity conditions for inference in large samples remain valid when a maximum likelihood approach is used. We considered it natural to use the log version of the MSN for this situation, calling it the modified log-skew-normal(MLSN) model.

This paper is organized as follows. Section 2 presents some probabilistic properties of the MSN model. It includes the derivation of the first few moments of the distribution and the moment-generating function. Plots and ranges for the asymmetry and kurtosis coefficients are reported. The score functions are derived, and the observed information matrix is presented. Section 3 is devoted to the definition of the new distribution, termed the MLSN distribution. The survival and the associated risk functions are derived. A general expression for the moments is obtained, and the non-existence of the moment-generating function is proven. Parameter estimation is conducted by the maximum likelihood (ML) approach. Observed and expected (Fisher) information matrices are derived. It is shown that the FIM for a location-scale extension of the model is non-singular so that large sample properties of the ML estimators (MLE) are satisfied. In Section 4, we present a brief introduction to Firth's approach (see Firth [10]) for bias reduction and some tables related to a small-scale simulation study, illustrating the amount of bias reduction in the asymmetry parameter. Finally, in Section 5, a real data illustration is presented indicating that the new model outperforms its most direct competitors. The closing section summarizes the main contributions of the paper.

2. Preliminaries

We say that a random variable Z follows a standard MSN distribution with parameter λ, denoted $Z \sim MSN(\lambda)$, if its pdf is given by

$$f_Z(z;\lambda) = 2\phi(z)\Phi(\lambda u(z)), \quad z \in \mathbb{R}, \tag{2}$$

where $u(z) = z/\sqrt{1+z^2}$ and $\lambda \in \mathbb{R}$. If $\lambda = 0$, then the MSN pdf in (2) reduces to the pdf of the standard normal distribution. Non-null values of the parameter λ directly affect the model's asymmetry so that, in the limit, as $\lambda \to \infty$, the MSN model becomes the half-normal (HN) distribution. All that is required to obtain the location-scale version of the model is to make the transformation $X = \mu + \sigma Z$, where $\mu \in \mathbb{R}$ and $\sigma > 0$ are the location and scale parameters, respectively. We use the notation $X \sim MSN(\mu, \sigma, \lambda)$. As shown in Arellano-Valle et al. [8] the FIM for the location-scale version is nonsingular at $\lambda = 0$. Thus, the ordinary properties of the MLE (consistency and asymptotic normality) remain valid for the MSN model.

Figure 1 shows plots of the pdf for the standard MSN model for several values of λ. The subtle difference between the MSN and SN densities for the same parameter value can be observed, say for the case $\lambda = 1$ (dotted line and segmented line).

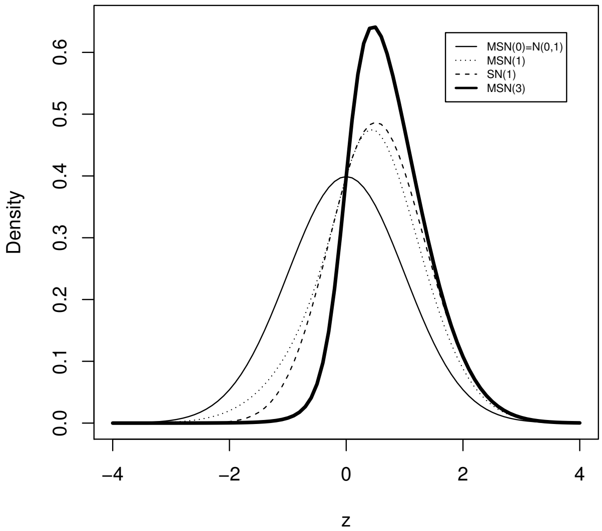

Figure 1. Plots of the MSN(λ) model for different values of λ: MSN(0) (solid line), MSN(1) (dotted line), SN(1) (dashed line), and MSN(3) (thick solid line) models.

The following properties follow from Arellano-Valle et al. [7].

2.1. Properties

Let $Z \sim MSN(\lambda)$; then:

1. $Z \sim \text{MSN}(\lambda) \Rightarrow -Z \sim \text{MSN}(-\lambda)$;
2. If $Z \sim \text{MSN}(\lambda)$, then $|Z| \sim \text{HN}(0,1)$;
3. If $Z|S = s \sim \text{SN}(s)$ and $S \sim \text{N}(\lambda, 1)$, then $Z \sim \text{MSN}(\lambda)$;
4. $\text{MSN}(0) \overset{d}{=} \text{N}(0,1)$, and $\text{MSN}(\lambda) \overset{d}{=} \text{MGSN}(\lambda, 1) \overset{d}{=} \text{GSN}(\lambda, 1)$.

According to Property 3, the MSN distribution can be represented as a mixture of the asymmetry parameter between the skew-normal and standard normal distributions. For the location-scale extension, that is $X = \mu + \sigma Z$, so that $X \sim \text{MSN}(\mu, \sigma, \lambda)$, with pdf

$$f_X(x;\mu,\sigma,\lambda) = \frac{2}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\lambda u\left(\frac{x-\mu}{\sigma}\right)\right). \tag{3}$$

2.2. Moments of the MSN Model

According to Property 3, the moments of the MSN distribution can be obtained by using the fact that it is a mixture between the SN and the $N(0,1)$ so that we can write $E(Z^k) = E(E(Z^k|S))$, where $E(Z^k|S)$, $k = 0,1,...$, are the moments for the SN model with asymmetry parameter s. Thus, for even values of k, the moments of the SN(s) distribution are constant, coinciding with the moments for the MSN distribution. For the odd moments, making use of the stochastic representation for the skew-normal distribution in Henze [11], we can write

$$\begin{aligned} E(Z^{2k+1}) &= E(E(Z^{2k+1}|S)) \\ &= E\left(\frac{b(2k+1)!}{2^k}\sum_{\nu=0}^{k}\frac{\nu!2^{2\nu}}{(2\nu+1)!(k-\nu)!}\frac{S^{2\nu+1}}{(1+S^2)^{(2k+1)/2}}\right) \\ &= \frac{b(2k+1)!}{2^k}\sum_{\nu=0}^{k}\frac{\nu!2^{2\nu}}{(2\nu+1)!(k-\nu)!}E\left(\frac{S^{2\nu+1}}{(1+S^2)^{(2k+1)/2}}\right), \end{aligned}$$

where $k = 0,1,2,...$, $b = \sqrt{\frac{2}{\pi}}$. Considering

$$\psi_{k,\nu} = \psi_{k,\nu}(\lambda) = E\left(\frac{S^{2\nu+1}}{(1+S^2)^{(2k+1)/2}}\right) = \int_{-\infty}^{\infty}\frac{s^{2\nu+1}}{(1+s^2)^{(2k+1)/2}}\phi(s-\lambda)ds$$

and $\psi_k = \psi_{k,k}$, we have that

$$E(Z^{2k+1}) = \frac{b(2k+1)!}{2^k}\sum_{\nu=0}^{k}\frac{\nu!2^{2\nu}}{(2\nu+1)!(k-\nu)!}\psi_{k,\nu}. \tag{4}$$

Therefore, the first four moments of the standard MSN distribution are given by

$$E(Z) = b\psi_0, \quad E(Z^2) = 1, \quad E(Z^3) = b(3\psi_0 - \psi_1) \quad \text{and} \quad E(Z^4) = 3.$$

The odd moments can also be obtained from Arellano-Valle et al. [7]:

$$E(Z^{2k+1}) = 2c_k - 2^k k! b,$$

for $k = 0,1,2,...$ where $c_k := c_k(\lambda) = \int_0^\infty x^k \phi(\sqrt{x})\Phi(\lambda u(\sqrt{x}))dx$.

2.3. Moment-Generating Function

As mentioned previously, if $Z|S = s \sim SN(s)$ and $S \sim N(\lambda,1)$, then $Z \sim MSN(\lambda)$, so that we can write

$$M_Z(t) = E(E(e^{Zt}|S)) = E(M_{Z|S}(t)) = 2e^{t^2/2}E\left(\Phi\left(\frac{S}{\sqrt{1+S^2}}t\right)\right). \tag{5}$$

Therefore, all the moments for the random variable Z are defined as follows:

$$E(Z^n) = \frac{\partial^n M_Z(t)}{\partial^n t}\bigg|_{t=0} = E\left(\frac{\partial^n M_{Z|S}(t)}{\partial^n t}\bigg|_{t=0}\right).$$

2.4. Observed Information Matrix for MSN Model

Consider a random sample $x_1, x_2, .., x_n$ from the MSN(θ) distribution, with $\theta = (\mu,\sigma,\lambda)$, so that the corresponding log-likelihood function is given by

$$l(\theta) = \frac{n}{2}\log\left(\frac{2}{\pi}\right) - n\log(\sigma) - \frac{1}{2}\sum_{i=1}^{n}z_i^2 + \sum_{i=1}^{n}\log(\Phi(\lambda u(z_i)),$$

where $z_i = (x_i - \mu)/\sigma$, $u(x) = x/\sqrt{1+x^2}$, $\mu \in \mathbb{R}$, $\sigma > 0$, $\lambda \in \mathbb{R}$, and $x \in \mathbb{R}$. The entries of the observed information matrix for the MSN model have the same structure as the corresponding moments for the MLSN model considered in Section 4.3, with the appropriate transformation.

3. The Modified Log-Skew-Normal Distribution

If $Z \sim \text{MSN}(\lambda)$, then we have that $Y = e^Z$ is distributed according to the standard MLSNdistribution with parameter λ, denoted by $Y \sim \text{MLSN}(\lambda)$, if its pdf is given by

$$f_Y(y;\lambda) = \frac{2}{y}\phi(\log(y))\Phi(\lambda u(\log(y))),$$

where $u(x) = x/\sqrt{1+x^2}$, $y \in \mathbb{R}^+$, and $\lambda \in \mathbb{R}$.

Figure 2 depicts plots of the pdf for the MLSN(λ) model for several values of λ. If $\lambda = 0$, then it coincides with the log-normal distribution (solid line), illustrating the fact that the MLSN model is an extension of the LN model. Concerning the location-scale situation, that is $Y \sim \text{MLSN}(\mu,\sigma,\lambda)$, where $\mu \in \mathbb{R}$, $\sigma > 0$, and $\lambda \in \mathbb{R}$, its density is given by

$$f_Y(y;\mu,\sigma,\lambda) = \frac{2}{y\sigma}\phi(z)\Phi(\lambda u(z)), \tag{6}$$

where $z = (\log(y) - \mu)/\sigma$, $u(x) = x/\sqrt{1+x^2}$, and $y \in \mathbb{R}^+$. In a survival analysis scenario, it is important to study the following functions: the survival function $S(t) = 1 - F(t)$ and the risk function $r(t) = f(t)/S(t)$, which for the model under study can be shown to be given by

$$S(t) = 1 - \int_0^t \frac{2}{w}\phi(\log(w))\Phi(\lambda u(\log(w)))dw$$

and

$$r(t) = \frac{\frac{2}{t}\phi(\log(t))\Phi(\lambda u(\log(t)))}{1 - \int_0^t \frac{2}{w}\phi(\log(w))\Phi(\lambda u(\log(w)))dw}.$$

Clearly, using L'Hopital´s rule, it follows that $r(t) \to 0$ as $t \to \infty$, as can also be appreciated graphically.

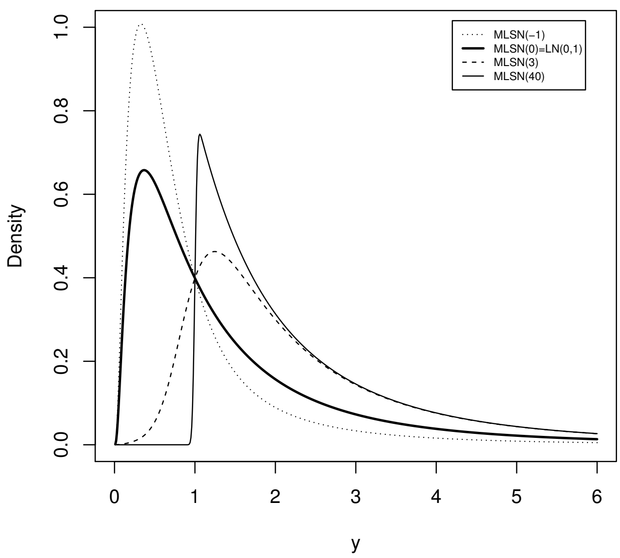

Figure 2. Plots of the MLSN(λ) model for different values of λ.

Figure 3 illustrates the behavior of the risk function $r(t)$ for some values of λ. The maximum values of the risk function for each λ decrease for $\lambda \in (-\infty, 1.4484)$ and increase otherwise. Moreover, $\lambda = -\infty$, $r(t)$ tends to a strictly increasing function defined in the interval $(0,1]$, otherwise being zero. On the other hand, $\lambda = \infty$, $r(t)$ is a strictly increasing function defined in the interval $[1,\infty)$, coinciding with the risk function of the log-normal model in this interval and taking the value zero in the interval $(0,1)$.

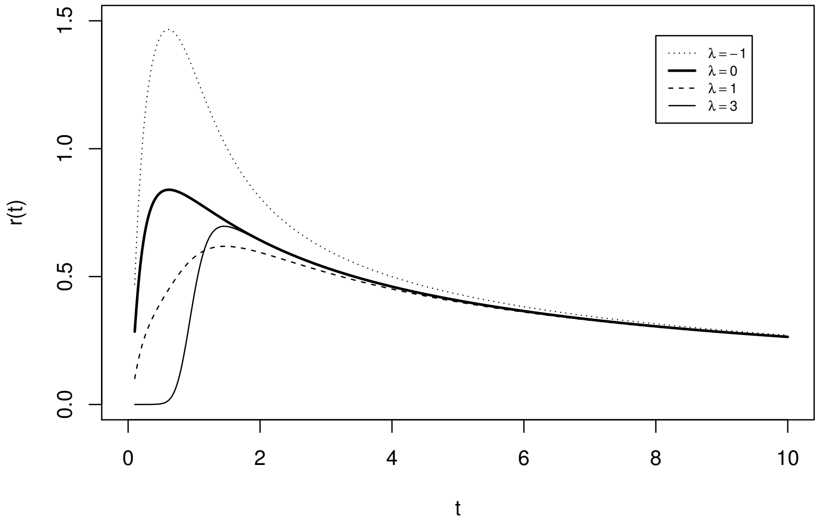

Figure 3. Plots of the risk function of the MLSN model for different values of λ.

3.1. Moments for the MLSN Model

The r-th moment is given by

$$\mu_r = 2e^{r^2/2}E\left(\Phi\left(\frac{S}{\sqrt{1+S^2}}r\right)\right),$$

for $r = 0,1,2.....$ and $S \sim N(\lambda, 1)$. This expression is obtained directly from the moment-generating function of the MSN model, given in (5), since it is valid for all $t > 0$, particularly for $t = r$. Alternatively, the moments can be obtained from

$$\mu_r = E(Y^r) = \int_0^\infty y^r \frac{2}{y}\phi(\log(y))\Phi(\lambda u(\log(y)))dy.$$

3.1.1. Non-Existence of the Moment-Generating Function for the MLSN Distribution

Proposition 1. *For all $\lambda \in \mathbb{R}$, variable $Y \sim MLSN(\lambda)$ has no moment-generating function.*

Proof. Part of the proof parallels that of Lin and Stoyanov [12] for a similar proof. Thus, for each $t > 0$,

$$E(e^{tY}) = \int_0^\infty h_\lambda(y)dy,$$

where

$$h_\lambda(y) = \frac{e^{ty}}{y}\phi(\log(y))\Phi(\lambda u(\log(y))) > 0, \quad \forall y > 0,$$

so that, for

$$\lambda \geq 0, \quad \lim_{y\to\infty} \Phi(\lambda u(\log(y))) \geq \frac{1}{2}$$

and

$$\lambda < 0, \quad \lim_{y\to\infty} \Phi(\lambda u(\log(y))) \geq \Phi(\lambda),$$

in both cases, $h_\lambda(y) \to \infty$ as $y \to \infty$. Therefore, given $t > 0$, $E(e^{tY}) = \infty$, for any λ. □

3.1.2. Skewness and Kurtosis Coefficients

To obtain expressions for the kurtosis and skewness coefficients, we have to compute the central moments using the following relationships:

$$\mu_2' = \mu_2 - \mu_1^2, \quad \mu_3' = \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3 \quad \text{and} \quad \mu_4' = \mu_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_2 - 3\mu_1^4.$$

This allows us to compute the variance, standard deviation (SD), coefficient of variation (CV), asymmetry (γ_1), and kurtosis (γ_2) coefficients, respectively, given by:

$$\mu_2' = Var(Y), \quad CV(Y) = \frac{SD(Y)}{\mu_2}, \quad \gamma_1 = \frac{\mu_3'}{(\mu_2')^{3/2}} \quad \text{and} \quad \gamma_2 = \frac{\mu_4'}{(\mu_2')^2}.$$

Table 1 reveals that the variation ranges for the variance and CV are relatively short, while the variation range for the asymmetry and kurtosis coefficients are relatively long if compared with the ordinary LSN model.

Table 1. Range of values for statistics corresponding to variance, CV, skewness, and kurtosis.

Statistics	Variance	CV	γ_1	γ_2
Minimum	0.062	0.477	0.110	1.952
Maximum	6.986	1.395	10.985	524.351

Figure 4 shows the behavior of the variance, CV, skewness, and kurtosis as a function of the lambda parameter. It can be observed that the minimum values correspond to the asymptotes of the left tail. On the other hand, for small values of λ, say $\lambda > 4$, the plots of the right tail stabilize around the horizontal asymptotes with values of 6.74, 0.936, 5.83, and 97.93, respectively, for the corresponding indices.

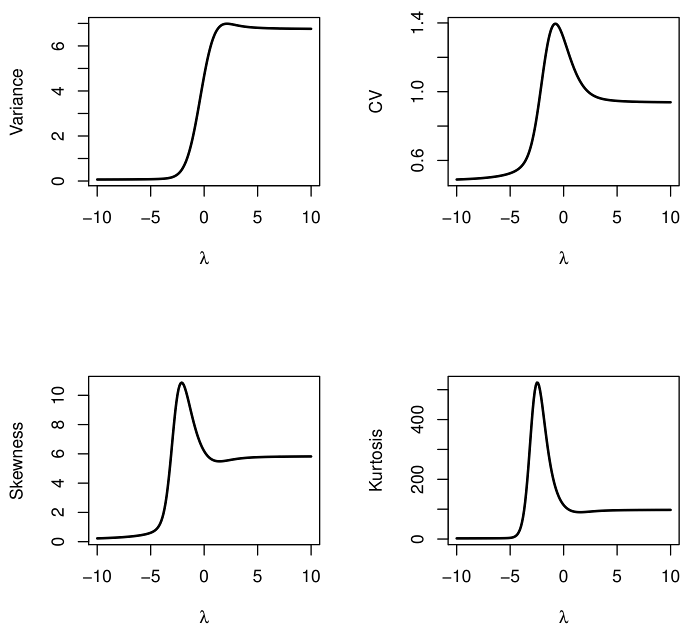

Figure 4. Plots of the variance, CV, skewness, and kurtosis.

3.2. MLE for the MLSN Model

Given a random sample $y_1, y_2, .., y_n$ from a random variable $Y \sim MLSN(\theta)$, with $\theta = (\mu, \sigma, \lambda)$, so that the corresponding log-likelihood is given by

$$l(\theta) = \frac{n}{2}\log\left(\frac{2}{\pi}\right) - n\log(\sigma) - \sum_{i=1}^{n}\log(y_i) - \frac{1}{2}\sum_{i=1}^{n} z_i^2 + \sum_{i=1}^{n}\log(\Phi(\lambda u(z_i))), \tag{7}$$

where $z_i = (\log(y_i) - \mu)/\sigma$, $u(x) = x/\sqrt{1+x^2}$, $\mu \in \mathbb{R}$, $\sigma > 0$, $\lambda \in \mathbb{R}$, and $y \in \mathbb{R}^+$. Using the following notation:

$$\rho_{nmi} = \frac{z_i^n}{(1+z_i^2)^{m/2}}\zeta(\lambda u(z_i)), \quad \eta_{nmi} = \frac{z_i^n}{(1+z_i^2)^m}\zeta^2(\lambda u(z_i)), \quad \overline{\eta}_{nm} = \frac{1}{n}\sum_{i=1}^{n}\rho_{nmi} \quad \text{and} \quad \zeta(x) = \frac{\phi(x)}{\Phi(x)},$$

we have that the associated scoring vector is given by

$$S_\mu = \frac{n}{\sigma}(\overline{z} - \lambda\overline{\rho}_{03}), \quad S_\sigma = \frac{n}{\sigma}\left(-1 + \overline{z^2} - \lambda\overline{\rho}_{13}\right) \quad \text{and} \quad S_\lambda = n\overline{\rho}_{11}.$$

Equating the scoring functions to zero, it follows that the likelihood equations are given as

$$\overline{z} = \lambda\overline{\rho}_{03}, \quad \overline{z^2} - 1 = \lambda\overline{\rho}_{13} \quad \text{and} \quad \overline{\rho}_{11} = 0.$$

Solving this system of equations, which require numerical procedures, leads to the MLE for μ, σ, and λ.

4. Bias Prevention of the MLE

4.1. Abstract for Firth's Approach

One problem with the ML approach is that the likelihood function can be unbounded for the parameter λ. In the standard case (just λ unknown), the MLE of λ is infinite given that all observations are positive. In the location-scale situation, the same happens because all observations are greater than μ, the location parameter. Therefore, the MLE overestimates parameter λ, which is always of concern in a statistical analysis. The percentage of overestimation of the shape parameter λ depends on the sample size n and on the true parameter value. It is of interest to reduce the amount of bias in a statistical problem. One approach considered previously in the literature was proposed by Firth [10] and had previously been used by Sartori [13] for the skew-normal and skew-t (ST) distributions. The reduced bias estimator is of order $O(n^{-2})$ (see also Cox and Snell [14]). To arrive at the reducedbias estimator, the following notation is used: Let $l(\theta)$ be the log-likelihood function for a parametric regular family of distributions, and denote by $U(\theta) = l'(\theta)$ the score function, $j(\theta) = -l''(\theta)$ the second derivative of the log-likelihood function, and l' its first derivative. We also consider the expectation of those functions that are of order $O(n)$. Therefore, we have

$$i(\theta) = E_\theta\{j(\theta)\}, \quad \nu_{\theta,\theta,\theta}(\theta) = E_\theta\{l'(\theta)^3\}, \quad \nu_{\theta,\theta\theta}(\theta) = E_\theta\{l'(\theta)l''(\theta)\}.$$

Briefly speaking, Firth's approach consists of modifying the score function keeping the score unchanged, that is

$$U_M(\theta) = U(\theta) + M(\theta), \tag{8}$$

where $M(\theta)$ is $O(1)$ as $n \to \infty$. The solution to the modified likelihood equation $U_M(\theta) = 0$ produces the modified MLE, say $\widehat{\theta}_M$ with bias of order $O(n^{-2})$. The modification factor $M(\theta)$ is given by

$$M(\theta) = -i(\theta)b(\theta) = \frac{1}{2}i(\theta)^{-1}(\nu_{\theta,\theta,\theta} + \nu_{\theta,\theta\theta}). \tag{9}$$

From the modified score function, the modified quasi-log-likelihood function is defined as

$$l_M(\theta) = \int_c^\theta U_M(t)dt = l(\theta) - l(c) + \int_c^\theta M(t)dt,$$

where c is an arbitrary constant. This penalized log-likelihood function has a penalty of order $O(1)$. It is also possible to define a modified log-likelihood ratio statistic:

$$W_M(\theta) = 2\{l_M(\widehat{\theta}_M) - l_M(\theta)\},$$

which is asymptotically distributed as a χ_1^2 distribution, which can be used for testing hypotheses about θ, as well as for confidence interval construction.

4.2. Simulation Study

Below, we present a simulation study based on 5000 iterations for $\mu = 0$, $\sigma = 1$, $\lambda = 5, 10$, and several sample sizes. Table 2 shows the bias for the modified and ordinary MLE, the empirical coverage of the confidence intervals based on $W_M^P(\lambda)$, and the percentage of cases in which the un-modified MLE for λ is finite. It can be observed that the MLEs $\widehat{\mu}$ and $\widehat{\sigma}$ present satisfactory behavior concerning bias, and they seem to be unaffected by the sometimes erratic behavior of estimator $\widehat{\lambda}$, so that Firth's method is only applied for the parameter λ.

Table 2. Results are based on 5000 samples simulated from the MSN$(0, 1, \lambda)$ model.

λ	n	$\widehat{\mu}(sd)$	$\widehat{\sigma}(sd)$	$\widehat{\lambda}^{a}(sd)$	$\widehat{\lambda}_M(sd)$	$W_M^P(\lambda)IC$	$\%(\widehat{\lambda} < \infty)$
5	50	−0.00351 (0.10073)	0.99853 (0.12670)	7.00103 (4.63146)	4.35954 (1.18831)	96.16	86.48
	100	−0.00208 (0.07114)	0.99873 (0.08973)	6.44116 (2.91069)	4.80315 (0.93203)	94.18	97.84
	200	0.00038 (0.05080)	0.99838 (0.06355)	5.53622 (1.43447)	4.89744 (0.42190)	94.04	99.98
10	50	0.01599 (0.07556)	0.98026 (0.11462)	11.36052 (9.14682)	6.56160 (2.39887)	87.98	66.00
	100	0.00359 (0.04890)	0.99319 (0.08054)	13.52624 (8.63025)	8.71761 (2.92047)	92.08	87.74
	200	0.00095 (0.03409)	0.99748 (0.05698)	12.95004 (6.00251)	9.54944 (1.74399)	93.38	98.42

$\widehat{\mu}$, $\widehat{\sigma}$, $\widehat{\lambda}$, and $\widehat{\lambda}_M$ are the estimates of the true values of the parameters, and *sd* denotes the respective standard deviations; the empirical coverage of the 0.95 confidence interval is based on $W_M^P(\lambda)$ and the empirical (theoretical) percentage for the situation that $\widehat{\lambda}$ exists. [a] Computed as $\widehat{\lambda} < \infty$.

The notation $W_M^P(\lambda) = 2\{l_M^P(\widehat{\lambda}_M) - l_M^P(\lambda)\}$ defines the likelihood ratio statistic based on the profile log-likelihood function $l^P(\lambda) = l(\widehat{\mu}(\lambda), \widehat{\sigma}(\lambda), \lambda)$, where $\widehat{\mu}$ and $\widehat{\sigma}$ are the MLEs for μ and σ, respectively, for a fixed value of λ.

4.3. Observed Information Matrix for the MLSN Model

The entries for the observed information matrix of the MLSN distribution, corresponding to the log-likelihood given in (7), are $J_{\theta_i\theta_j} = -\frac{\partial^2 l(\theta)}{\partial\theta_i\partial\theta_j}$, for $i, j = 1, 2, 3$ (see Appendix A):

$$J_{\mu\mu} = -\frac{n}{\sigma^2}\Big(1 + 3\lambda\overline{\rho}_{15} + \lambda^3\overline{\rho}_{17} + \lambda^2\overline{\eta}_{03}\Big), \qquad J_{\mu\sigma} = \frac{n}{\sigma^2}\Big(-2\overline{z} + \lambda\overline{\rho}_{05} - 2\lambda\overline{\rho}_{25} - \widehat{\lambda^3}\overline{\rho}_{27} - \lambda^2\overline{\eta}_{13}\Big),$$

$$J_{\mu\lambda} = -\frac{n}{\sigma}\Big(\overline{\rho}_{03} - \lambda^2\overline{\rho}_{25} - \lambda\overline{\eta}_{12}\Big), \qquad J_{\sigma\sigma} = \frac{n}{\sigma^2}\Big(1 - 3\overline{z^2} + \lambda\overline{\rho}_{13} + \lambda\overline{\rho}_{15} - 2\lambda\overline{\rho}_{35} - \lambda^3\overline{\rho}_{37} - \lambda^2\overline{\eta}_{23}\Big),$$

$$J_{\sigma\lambda} = \frac{n}{\sigma}\Big(-\overline{\rho}_{13} + \lambda^2\overline{\rho}_{35} + \lambda\overline{\eta}_{22}\Big), \qquad J_{\lambda\lambda} = -n(\lambda\overline{\rho}_{33} + \overline{\eta}_{21}).$$

4.4. Fisher Information

For a random sample $y_1, y_2, .., y_n$ from the MLSN(θ) distribution, with $\theta = (\mu, \sigma, \lambda)$, the FIM associated with (7) is given by the following entries (see Appendix A):

$$I_{\mu\mu} = \frac{1}{\sigma^2}\left(1 + \lambda^2\eta_{03}\right), \quad I_{\sigma\mu} = -\frac{1}{\sigma^2}\left(-2E(z) + \lambda\rho_{05} - 2\lambda\rho_{25} - \lambda^3\rho_{27} - \lambda^2\eta_{13}\right),$$

$$I_{\mu\lambda} = \frac{1}{\sigma}\left(\rho_{03} - \lambda^2\rho_{25} - \lambda\eta_{12}\right), \quad I_{\sigma\sigma} = \frac{1}{\sigma^2}\left(2 + \lambda^2\eta_{23}\right), \quad I_{\lambda\sigma} = -\frac{\lambda}{\sigma}\eta_{22}, \quad I_{\lambda\lambda} = \eta_{21},$$

where

$$\begin{aligned} \rho_{nm} &= E\left(\frac{z^n}{(1+z^2)^{m/2}}\zeta\right), \quad \rho_{nm} = 0 \quad if \quad n = \text{odd number} \\ \eta_{nm} &= E\left(\frac{z^n}{(1+z^2)^m}\zeta^2\right), \quad \text{with } \zeta := \zeta(\lambda u(z)) = \frac{\phi(\lambda u(z))}{\Phi(\lambda u(z))}. \end{aligned}$$

For the case $\lambda = 0$, we have that

$$\text{FIM} = \begin{pmatrix} \frac{1}{\sigma^2} & 0 & \frac{d_1}{\sigma} \\ 0 & \frac{2}{\sigma^2} & 0 \\ \frac{d_1}{\sigma} & 0 & d_2 \end{pmatrix},$$

where $d_1 = 2(2/\pi)^{1/2}\int_0^\infty \frac{\phi(x)}{(1+x^2)^{3/2}}dx$ and $d_2 = \frac{2}{\pi}(1 - (2\pi)^{1/2}e^{1/2}\Phi(-1))$ have to be computed numerically. It is clearly seen that the above FIM is non-singular at $\lambda = 0$.

5. An Application

The dataset analyzed in this section was previously studied in Nadarajah [15] and Leiva et al. [16]. It consists of daily measurements of ozone concentration (in $ppb = ppm \times 1000$) in New York city between May and September 1973. The data were supposed to be independent, without the presence of tendencies or cyclical components (see Gokhale and Khare [17]). Table 3 presents the summary statistics, in particular the asymmetry and kurtosis coefficients, which are represented by the sample quantities (b_1) and (b_2), respectively.

Table 3. Descriptive statistics of ozone concentration level measurement data.

Data	n	Mean	sd	b_1	b_2
Ozone	116	42.129	32.987	1.209	4.112

Table 4 shows the MLE for the three parameters for the MLSN, LN, and LSN distributions, respectively, where values in parentheses correspond to standard errors. Note also from the table that the AIC for the LSN model is slightly smaller than that for the MLSN model; it is our opinion that the latter model should be preferred because the LSN model has a singular covariance matrix, and so, the likelihood ratio statistics for testing $H_0 : \lambda = 0$ is not distributed according to a central chi-squared distribution in large samples. On the other hand, under the MLSN model, for which the FIM is non-singular, we have that the hypothesis $H_0 : \lambda = 0$ is rejected at the 5% level, using the likelihood ratio statistics; this is given by

$$\Lambda = \frac{L_{LN(\hat{\mu},\hat{\sigma})}}{L_{MLSN(\hat{\mu},\hat{\sigma},\hat{\lambda})}}.$$

Replacing the MLE from Table 4, we have that $-2log\Lambda = -2(-543.883 + 542.105) = 9.516$, greater than the critical value $\chi_1^2 = 3.84$. Furthermore, from the table, a large sample 95% confidence interval for λ based on the MLSN model does not contain $\lambda = 0$.

Table 4. MLEs for the MLSN, LN, and LSN models.

MLEs	MLSN	LN	LSN
$\widehat{\mu}$	4.532 (0.105)	3.418 (0.080)	4.372 (0.234)
$\widehat{\sigma}$	1.411 (0.121)	0.861 (0.056)	1.285 (0.193)
$\widehat{\lambda}$	−4.095 (1.057)	-	−2.411 (1.396)
log-likelihood	−542.105	−543.883	−541.655
AIC	1090.211	1091.766	1089.31

Figure 5 presents the data histogram with the corresponding fitted fdp for the MLSN$(\widehat{\mu}, \widehat{\sigma}, \widehat{\lambda})$ (solid line) and LN$(\widehat{\mu}, \widehat{\sigma}, \widehat{\lambda})$ distributions (dotted line) and the fit cdf for the MLSN and LN models, jointly with the empirical cdf. It can be shown from the figure that the MLSN model seems to provide a (graphically) satisfactory fit.

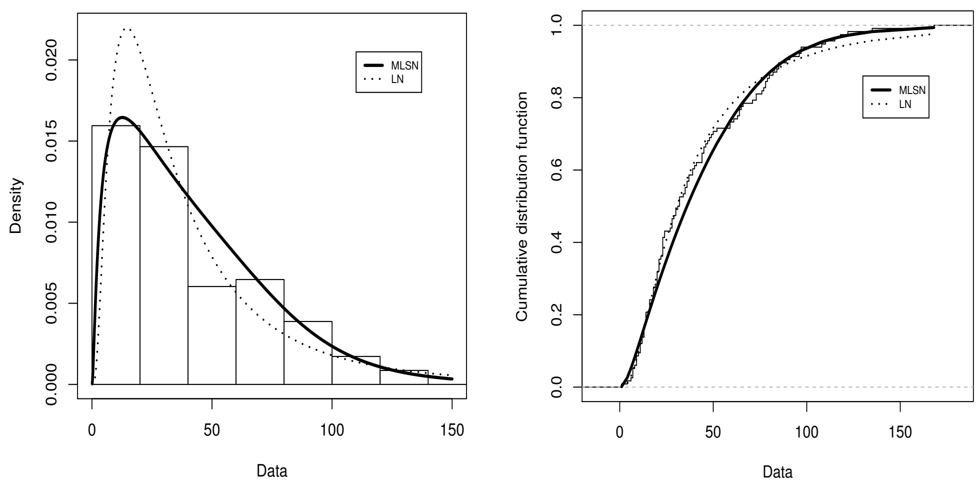

Figure 5. Left panel: Kernel density plot of the ozone concentration level data fit by the LN and MLSN distributions. **Right panel**: Empirical cdf of the LN (dotted line) and MLSN (solid line) models, whose parameters are estimated by the ML method for ozone concentration level data.

Graphical corroboration of the better fit of the MLSN model than the LN model is also illustrated with the QQ-plots in Figure 6.

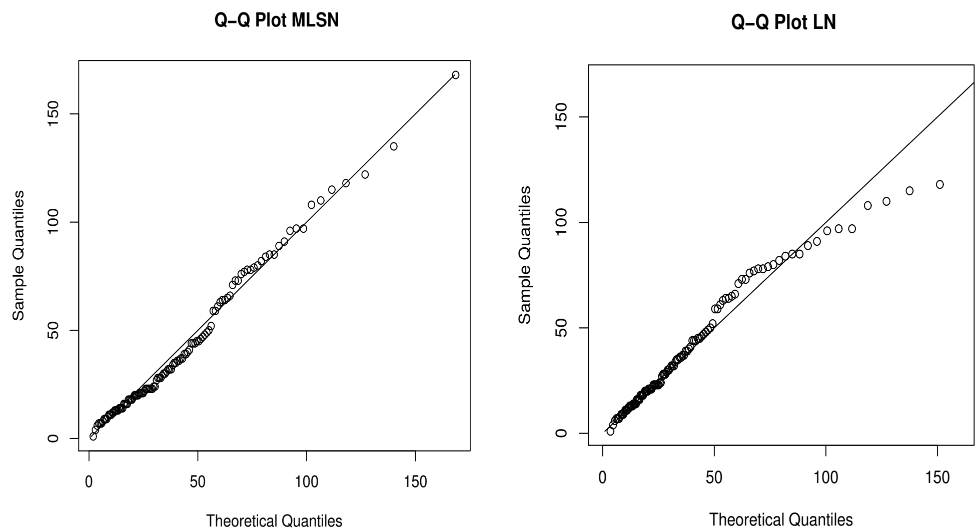

Figure 6. Left panel: QQ-plot for the MLSN model. **Right panel**: QQ-plot for the LN model.

Table 5 presents the MLEs $\widehat{\mu}$, $\widehat{\sigma}$, and $\widehat{\lambda}$ and the modified $\widehat{\lambda}_M$ with the corresponding standard errors (in parentheses), obtained by using the estimated FIM for the MLSN model, bearing in mind that the asymptotic distribution of $\widehat{\theta}$ is $N_3(\theta, I(\theta)^{-1}/n)$, where $\theta = (\mu, \sigma, \lambda)$ or $\theta = (\mu, \sigma, \lambda_M)$. The table indicates that the modified MLE $\widehat{\lambda}_M$ is greater than the ordinary MLE $\widehat{\lambda}$ and is expected to have smaller bias.

Table 5. MLEs of the μ, σ, λ, and λ_M parameters.

$\widehat{\mu}$	$\widehat{\sigma}$	$\widehat{\lambda}$	$\widehat{\lambda}_M$	$l(\widehat{\mu}, \widehat{\sigma}, \widehat{\lambda})$	$l(\widehat{\mu}, \widehat{\sigma}, \widehat{\lambda}_M)$
4.532(0.105)	1.411(0.121)	−4.095(1.057)	–	−542.105	
4.532(0.123)	1.411(0.130)	–	−3.342(0.791)	–	−542.663

Table 6 presents confidence intervals for $\widehat{\lambda}$ and $\widehat{\lambda}_M$ for several confidence coefficients. Comparing the lengths of the intervals for the two estimators, there is strong evidence that the modified estimator $\widehat{\lambda}_M$ presents shorter intervals.

Table 6. Confidence intervals for the estimated values of $\widehat{\lambda}$ and $\widehat{\lambda}_M$.

MLE	95%	98%	99%
$\widehat{\lambda}$	(−5.834, −2.356)	(−6.266, −1.924)	(−6.554, −1.636)
$\widehat{\lambda}_M$	(−4.644, −2.040)	(−4.968, −1.716)	(−5.183, −1.500)

6. Concluding Remarks

This paper focused on a transformation of the MSN model (Arellano-Valle et al. [8]), which led to a more flexible distribution (wider ranges for asymmetry and kurtosis). This model, which we call the MLSN distribution, is suitable for positive data, its main competitors being the LN and LSN distributions. One interesting aspect of the new model is that its FIM is non-singular so that the large sample theory for the MLE remains valid. This is not the case with the LSN model, for which the FIM is singular in the location-scale version. Thus, in particular for testing $H_0 : LN$ against $H_0 : MLSN$, under the MLSN model, the likelihood ratio statistics in large samples are distributed as in the chi-squared distribution. Large sample confidence intervals could also be constructed and used for testing the hypothesis $H_0 : \lambda = 0$, where λ is the skewness parameter. Rejection of H_0 indicates that the MLSN model should be preferred. It is also noticed with the simulation study that the MLE of λ can overestimate λ (which could be infinite for some samples). Thus, the bias-reducing approach of Firth [10] was used to derive a less-biased estimator. Estimations for the location and scale parameters remain stable, however, and do not need to be corrected. An application to a real dataset revealed that the new model can be a valuable alternative for modeling positive data.

Author Contributions: Conceptualization, J.A., R.B.A.-V., and H.W.G.; methodology, R.A.-V., H.B., and H.W.G.; software, J.A. and O.V.; validation, R.B.A.-V., H.B., and H.W.G.; formal analysis, R.A.-V. and H.W.G.; investigation, J.A., O.V., and H.W.G.; writing—original draft preparation, J.A. and R.B.A.-V.; writing—review and editing, J.A. and O.V.; funding acquisition, H.W.G. and O.V. All authors have read and agreed to the published version of the manuscript.

Funding: The research of Jaime Arrué and Héctor W. Gómez was supported by SEMILLERO UA-2022 (Chile).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset is available in [15,16].

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

Appendix A

Considering the notation:

$$\theta = (\mu, \sigma), \quad z = \frac{x-\mu}{\sigma}, \quad u(x) = \frac{x}{\sqrt{1+x^2}}, \quad \zeta = \zeta(\lambda u(z)) = \frac{\phi(\lambda u(z))}{\Phi(\lambda u(z))},$$

and the following derivatives:

$$\frac{\partial z}{\partial \mu} = -\frac{1}{\sigma}, \qquad \frac{\partial z}{\partial \sigma} = -\frac{z}{\sigma}, \qquad \frac{\partial u(z)}{\partial \theta} = \frac{1}{(1+z^2)^{3/2}}\frac{\partial z}{\partial \theta},$$

$$\begin{aligned}
\frac{\partial \zeta}{\partial \theta} &= \left(-\frac{\lambda^2 z}{(1+z^2)^2}\zeta - \frac{\lambda}{(1+z^2)^{3/2}}\zeta^2\right)\frac{\partial z}{\partial \theta}, \\
\frac{\partial}{\partial \theta}\left(\frac{z}{(1+z^2)^{3/2}}\right) &= \frac{1-2z^2}{(1+z^2)^{5/2}}\frac{\partial z}{\partial \theta}, \\
\frac{\partial \zeta}{\partial \lambda} &= -\frac{\lambda z^2}{(1+z^2)}\zeta - \frac{z}{(1+z^2)^{1/2}}\zeta^2,
\end{aligned}$$

we have that the score functions score for μ, σ, and λ for the MSN(λ) model are given by

$$S_\mu = \frac{1}{\sigma}\left(z - \frac{\lambda}{(1+z^2)^{3/2}}\zeta\right), \quad S_\sigma = \frac{1}{\sigma}\left(-1 + z^2 - \frac{\lambda z}{(1+z^2)^{3/2}}\zeta\right), \quad S_\lambda = \frac{z}{(1+z^2)^{1/2}}\zeta,$$

so that the entries of the FIM are given by:

$$\begin{aligned}
I_{\mu\mu} &= -E(S_{\mu\mu}) = -\frac{1}{\sigma^2}E\left(-1 - \frac{3\lambda z}{(1+z^2)^{5/2}}\zeta - \frac{\lambda^3 z}{(1+z^2)^{7/2}}\zeta - \frac{\lambda^2}{(1+z^2)^3}\zeta^2\right) \\
&= \frac{1}{\sigma^2}\left(1 + \lambda^2\eta_{03}\right), \\
I_{\mu\sigma} &= -E(S_{\sigma\mu}) = -\frac{1}{\sigma^2}E\left(-2z + \frac{\lambda(1-2z^2)}{(1+z^2)^{5/2}}\zeta - \frac{\lambda^3 z^2}{(1+z^2)^{7/2}}\zeta - \frac{\lambda^2 z}{(1+z^2)^3}\zeta^2\right) \\
&= -\frac{1}{\sigma^2}\left(-2E(z) + \lambda\rho_{05} - 2\lambda\rho_{25} - \lambda^3\rho_{27} - \lambda^2\eta_{13}\right), \\
I_{\mu\lambda} &= -E(S_{\mu\lambda}) = -\frac{1}{\sigma}E\left(-\frac{1}{(1+z^2)^{3/2}}\left\{\zeta - \frac{\lambda^2 z^2}{(1+z^2)}\zeta - \frac{\lambda z}{(1+z^2)^{1/2}}\zeta^2\right\}\right) \\
&= \frac{1}{\sigma}\left(\rho_{03} - \lambda^2\rho_{25} - \lambda\eta_{12}\right), \\
I_{\sigma\sigma} &= -E(S_{\sigma\sigma}) \\
&= -\frac{1}{\sigma^2}E\left(1 - 3z^2 + \frac{\lambda z}{(1+z^2)^{3/2}}\zeta + \frac{\lambda(z-2z^3)}{(1+z^2)^{5/2}}\zeta - \frac{\lambda^3 z^3}{(1+z^2)^{7/2}}\zeta - \frac{\lambda^2 z^2}{(1+z^2)^3}\zeta^2\right) \\
&= \frac{1}{\sigma^2}\left(2 + \lambda^2\eta_{23}\right), \\
I_{\sigma\lambda} &= -E(S_{\lambda\sigma}) = -\frac{1}{\sigma}E\left(-\frac{z}{(1+z^2)^{3/2}}\zeta + \frac{\lambda^2 z^3}{(1+z^2)^{5/2}}\zeta + \frac{\lambda z^2}{(1+z^2)^2}\zeta^2\right) \\
&= -\frac{\lambda}{\sigma}\eta_{22}, \\
I_{\lambda\lambda} &= -E(S_{\lambda\lambda}) = -E\left(-\frac{\lambda z^3}{(1+z^2)^{3/2}}\zeta - \frac{z^2}{1+z^2}\zeta^2\right) \\
&= \eta_{21}.
\end{aligned}$$

References

1. Jones, M.C.; Arnold, B.C. Distributions that are both log-symmetric and R-symmetric. *Electron. J. Stat.* **2008**, *2*, 1300–1308. [CrossRef]
2. Mudholkar, G.S.; Wang, H. IG-symmetry and R-symmetry: interrelations and applications to the inverse Gaussian theory. *J. Stat. Plan. Inference* **2007**, *137*, 3655–3671. [CrossRef]
3. Azzalini, A.; Cappello, D.; Kotz, S. Log-skew-normal and log-skew-t distributions as models for family income data. *J. Income Distrib.* **2003**, *11*, 12–20. [CrossRef]
4. Marchenko, Y.V.; Genton, M.G. Multivariate log-skew-elliptical distributions with applications to precipitation data. *Environmetrics* **2010**, *21*, 318–340. [CrossRef]
5. Bolfarine, H.; Gómez, H.W.; Rivas, L. The log-bimodal-skew-normal model. A geochemical application. *J. Chemom.* **2011**, *25*, 329–332. [CrossRef]
6. Chai, H.; Bailey, K. Use of log-normal distribution in analysis of continuous data with a discrete component at zero. *Stat. Med.* **2008**, *27*, 3643–3655. [CrossRef] [PubMed]
7. Arellano-Valle, R.B.; Gómez, H.W.; Quintana, F.A. A new class of skew-normal distributions. *Commun. Stat.-Theory Methods* **2004**, *33*, 1465–1480. [CrossRef]
8. Arellano-Valle, R.B.; Gómez, H.W.; Salinas, H.S. A note on the Fisher information matrix for the skew-generalized-normal model. *SORT* **2013**, *37*, 19–28.
9. Arrué, J.; Arellano-Valle, R.B.; Gómez, H.W. Bias reduction of maximum likelihood estimates for a modified skew normal distribution. *J. Stat. Comput. Simul.* **2016**, *86*, 2967–2984. [CrossRef]
10. Firth, D. Bias reduction of maximum likelihood estimates. *Biometrika* **1993**, *80*, 27–38; Amended in *Biometrika* **1995**, *82*, 667. [CrossRef]
11. Henze, N. A probabilistic representation of the skew-normal distribution. *Scand. J. Stat.* **1986**, *13*, 271–275.
12. Lin, G.D.; Stoyanov, J. The logarithmic skew-normal distributions are moment-indeterminate. *J. Appl. Prob.* **2009**, *46*, 909–916. [CrossRef]
13. Sartori, N. Bias prevention of maximum likelihood estimates for scalar skew normal and skew t distributions. *J. Stat. Plan. Inference* **2006**, *136*, 4259–4275. [CrossRef]
14. Cox, D.R.; Snell, E.J. A general definition of residuals. *J. R. Stat. Soc. Ser. B* **1968**, *30*, 248–275. [CrossRef]
15. Nadarajah, S. A truncated inverted beta distribution with application to air pollution data. *Stoch. Environ. Res. Risk Assess.* **2008**, *22*, 285–289. [CrossRef]
16. Leiva, V.; Vilca-Labra, F.; Balakrishnan, N.; Sanhueza, A. A Skewed Sinh-Normal Distribution and its Properties and Application to Air Pollution. *Commun. Stat.-Theory Methods* **2010**, *39*, 426–443. [CrossRef]
17. Gokhale, S.; Khare, M. Statistical behavior of carbon monoxide from vehicular exhausts in urban environments. *Environ. Model. Softw.* **2007**, *22*, 526–535. [CrossRef]