Article

On the Folded Normal Distribution

Michail Tsagris [1]**, Christina Beneki** [2] **and Hossein Hassani** [3,*]

[1] School of Mathematical Sciences, University of Nottingham, NG7 2RD, UK;
E-Mail: mtsagris@yahoo.gr
[2] School of Business and Economics, TEI of Ionian Islands, 31100 Lefkada, Greece;
E-Mail: christinabeneki@gmail.com
[3] Statistical Research Centre, Executive Business Centre, Bournemouth University, BH8 8EB, UK;

* Author to whom correspondence should be addressed; E-Mail: hhassani@bournemouth.ac.uk;
Tel.: +44-120-296-8708.

Abstract: The characteristic function of the folded normal distribution and its moment function are derived. The entropy of the folded normal distribution and the Kullback–Leibler from the normal and half normal distributions are approximated using Taylor series. The accuracy of the results are also assessed using different criteria. The maximum likelihood estimates and confidence intervals for the parameters are obtained using the asymptotic theory and bootstrap method. The coverage of the confidence intervals is also examined.

1. Introduction

Mainly studied in the 1960s, the folded normal distribution is a special case of the Gaussian distribution occurring when the sign of the variable is always positive. In 1961, a method of estimating the parameters based upon the estimating equations of the moments was discussed in [1], where they also gave some examples of its applications in the industrial sector. The folded normal distribution was used to study the magnitude of deviation of an automobile strut alignment [2]. The properties of the multivariate folded normal distribution with its possible applications were studied in [3]. In addition, tables with probabilities for a range of values of the vector of parameters were provided, and an application of the model with real data was illustrated. An alternative method using the second and fourth

moments of the distribution was proposed in [4], whilst [5] performed maximum likelihood estimation and calculated the asymptotic information matrix. Thereafter, the sequential probability ratio test for the null hypothesis of the location parameter being zero against a specific alternative was evaluated in [6] with the idea of illustrating the use of cumulative sum control charts for multiple observations.

In [7], the author dealt with the hypothesis testing of the zero location parameter regardless of the variance being known or not. The distribution formed by the ratio of two folded normal variables was studied and illustrated with a few applications in [8]. The folded normal distribution has been applied to many practical problems. For instance, introduced in [9] is an economic model to determine the process specification limits for folded normally distributed data.

Through this paper, we will examine the folded normal distribution from a different perspective. In the process, we will consider the study of some of its properties, namely the characteristic and moment generating functions, the Laplace and Fourier transformations and the mean residual life of this distribution. The entropy of this distribution and its Kullback–Leibler divergence from the normal and half normal distributions will be approximated via the Taylor series. The accuracy of the approximations are assessed using numerical examples.

Also reviewed here is the maximum likelihood estimates (for an introduction, see [1]), with examples from simulated data given for illustration purposes. Simulation studies will be performed to assess the validity of the estimates with and without bootstrap calibration in low sample cases. Numerical optimization of the log-likelihood will be carried out using the simplex method [10].

2. The Folded Normal

The folded normal distribution with parameters (μ, σ^2) stems from taking the absolute value of a normal distribution with the same vector of parameters. The density of Y, with $Y \sim N(\mu, \sigma^2)$ is given by:

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2} \tag{1}$$

Thus, $X = |Y|$, denoted by $Y \sim FN(\mu, \sigma^2)$, has the following density:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \left[e^{-\frac{1}{2\sigma^2}(x-\mu)^2} + e^{-\frac{1}{2\sigma^2}(x+\mu)^2} \right] \tag{2}$$

The density can be written in a more attractive form [5]:

$$f(x) = \sqrt{\frac{2}{\pi\sigma^2}} e^{-\frac{(x^2+\mu^2)^2}{2\sigma^2}} \cosh\left(\frac{\mu x}{\sigma^2}\right) \tag{3}$$

and by expanding the *cosh* via a Taylor series, we can also write the density as:

$$f(x) = \sqrt{\frac{2}{\pi\sigma^2}} e^{-\frac{(x^2+\mu^2)^2}{2\sigma^2}} \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} \left(\frac{\mu x}{\sigma^2}\right)^{2n} \tag{4}$$

We can see that the folded normal distribution is not a member of the exponential family. The cumulative distribution can be written as:

$$F(x) = \frac{1}{2} \left[erf\left(\frac{x-\mu}{\sqrt{2\sigma^2}}\right) + erf\left(\frac{x+\mu}{\sqrt{2\sigma^2}}\right) \right] \tag{5}$$

where erf is the error function:

$$erf(x) = \frac{2}{\sqrt{pi}} \int_0^x e^{-t^2} dx \tag{6}$$

The mean and the variance of Equation (2) is calculated using direct calculation of the integrals as follows [1]:

$$\mu_f = \sqrt{\frac{2}{\pi}} \sigma e^{-\frac{\mu^2}{2\sigma^2}} + \mu \left[1 - 2\Phi\left(-\frac{\mu}{\sigma}\right) \right] \tag{7}$$

$$\sigma_f^2 = \mu^2 + \sigma^2 - \mu_f^2 \tag{8}$$

where $\Phi(.)$ is the cumulative distribution function of the standard normal distribution. The third and fourth moments about the origin are calculated in [4]. We develop the calculation further by providing the characteristic function and the moment generating function of Equation (2). Figure (1) shows the densities of the folded normal for some parameter values.

Figure 1. The black line is the density of the $N(\mu, \sigma^2)$ and the red line of the $FN(\mu, \sigma^2)$. The parameters in the left figure (**a**) are $\mu = 2$ and $\sigma^2 = 3$ and in the right figure (**b**) $\mu = 2$ and $\sigma^2 = 4$.

(a)

(b)
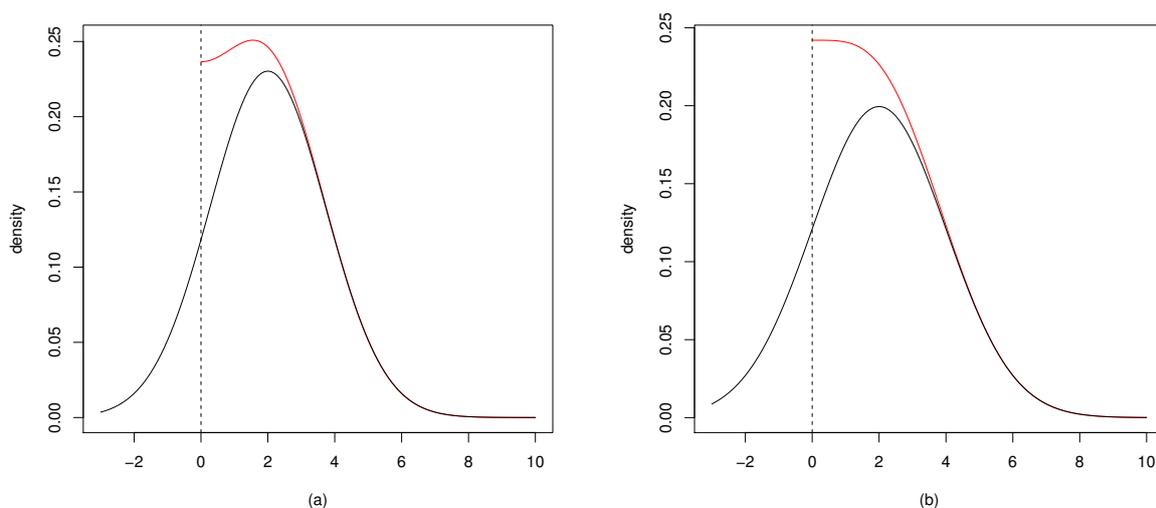
2.1. Relations to Other Distributions

The distribution of $Z = X/\sigma$ is a non-central χ distribution with one degree of freedom and non-centrality parameter equal to $(\mu/\sigma)^2$ [11]. It is clear that when $\mu = 0$, a central χ_1 is obtained. The half normal distribution is a special case of Equation (2), with $\mu = 0$ for which [12] showed that it is the limiting form of the folded (central) t distribution as the degrees of freedom of the latter go to infinity. Both distributions are further developed in the bivariate case in [13].

The folded normal distribution can also be seen as the the limit of the folded non-standarized t distribution as the degrees of freedom go to infinity. The folded non-standarized t distribution is the distribution of the absolute value of the non-standarized t distribution with v degrees of freedom:

$$g(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{v\pi\sigma^2}} \left\{ \left[1 + \frac{1}{v}\frac{(x-\mu)^2}{\sigma^2}\right]^{-\frac{v+1}{2}} + \left[1 + \frac{1}{v}\frac{(x+\mu)^2}{\sigma^2}\right]^{-\frac{v+1}{2}} \right\} \tag{9}$$

2.2. Mode of the Folded Normal Distribution

The mode of the distribution is the value of x for which the density is maximised. In order to find this value, we take the first derivative of the density with respect to x and set it equal to zero. Unfortunately, there is no closed form. We can, however, write the derivative in a better way and end up with a non-linear equation.

$$\frac{df(x)}{dx} = 0 \quad \Rightarrow \quad -\frac{(x-\mu)}{\sigma^2}e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} - \frac{(x+\mu)}{\sigma^2}e^{-\frac{1}{2}\frac{(x+\mu)^2}{\sigma^2}} = 0 \tag{10}$$

$$\Rightarrow \quad x\left[e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} + e^{-\frac{1}{2}\frac{(x+\mu)^2}{\sigma^2}}\right] - \mu\left[e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} - e^{-\frac{1}{2}\frac{(x+\mu)^2}{\sigma^2}}\right] = 0 \tag{11}$$

$$\Rightarrow \quad x\left(1 + e^{-\frac{2\mu x}{\sigma^2}}\right) - \mu\left(1 - e^{-\frac{2\mu x}{\sigma^2}}\right) = 0 \tag{12}$$

$$\Rightarrow \quad (\mu + x)e^{-\frac{2\mu x}{\sigma^2}} = \mu - x \tag{13}$$

$$\Rightarrow \quad x = -\frac{\sigma^2}{2\mu}\log\frac{\mu - x}{\mu + x} \tag{14}$$

We saw from numerical investigation that when $\mu < \sigma$, the maximum is met when $x = 0$. When $\mu \geq \sigma$, the maximum is met at $x > 0$, and when μ becomes greater than 3σ, the maximum approaches μ. This is of course something to be expected, since, in this case, the folded normal converges to the normal distribution.

2.3. Characteristic Function and Other Related Functions of the Folded Normal Distribution

Forms for the higher moments of the distribution when the moment is an odd and even number is provided in [4]. Here, we derive its characteristic and, thus, the moment generating function.

$$\begin{aligned}
\varphi_x(t) &= E\left(e^{itX}\right) = \int_0^\infty e^{itx} f_X(x)\,dx = \int_0^\infty e^{itx}\frac{1}{\sqrt{2\pi\sigma^2}}\left[e^{-\frac{1}{2\sigma^2}(x-\mu)^2} + e^{-\frac{1}{2\sigma^2}(x+\mu)^2}\right]dx \\
&= \int_0^\infty \frac{e^{itx - \frac{1}{2\sigma^2}(x-\mu)^2}}{\sqrt{2\pi\sigma^2}}dx + \int_0^\infty \frac{e^{itx - \frac{1}{2\sigma^2}(x+\mu)^2}}{\sqrt{2\pi\sigma^2}}dx \\
&= \int_0^\infty \frac{e^A}{\sqrt{2\pi\sigma^2}}dx + \int_0^\infty \frac{e^B}{\sqrt{2\pi\sigma^2}}dx
\end{aligned} \tag{15}$$

We will work now with the forms A and B.

$$A = itx - \frac{1}{2\sigma^2}(x-\mu)^2 = \frac{2i\sigma^2 tx - x^2 + 2\mu x - \mu^2}{2\sigma^2} = -\frac{x^2 - 2x\left(i\sigma^2 t + \mu\right) + \mu^2}{2\sigma^2} \tag{16}$$

$$= -\frac{\left[x - (i\sigma^2 t + \mu)\right]^2 + \sigma^4 t^2 - 2i\sigma^2 t\mu}{2\sigma^2} = -\frac{(x-a)^2}{2\sigma^2} - \frac{\sigma^2 t^2}{2} + i\mu t \tag{17}$$

where $a = i\sigma^2 t + \mu$. Thus, the first part of Equation (15) becomes:

$$\int_0^\infty \frac{e^A}{2\pi\sigma^2}dx = e^{\frac{-\sigma^2 t^2}{2} + i\mu t}\int_0^\infty \frac{e^{-(x-a)^2}}{2\pi\sigma^2}dx = e^{\frac{-\sigma^2 t^2}{2} + i\mu t}\left[1 - P\left(X \leq 0\right)\right] \tag{18}$$

$$= e^{\frac{-\sigma^2 t^2}{2} + i\mu t}\left[1 - \Phi\left(-\frac{a}{\sigma}\right)\right] = e^{\frac{-\sigma^2 t^2}{2} + i\mu t}\left[1 - \Phi\left(-\frac{\mu}{\sigma} - i\sigma t\right)\right] \tag{19}$$

The second exponent, B, using similar calculations becomes:

$$B = itx - \frac{1}{2\sigma^2}(x+\mu)^2 = -\frac{\left[x - (i\sigma^2 t - \mu)\right]^2}{2\sigma^2} - \frac{\sigma^2 t^2}{2} - i\mu t \tag{20}$$

and, thus, the second part of Equation (15) becomes:

$$\int_0^\infty \frac{e^B}{2\pi\sigma^2}dx = e^{-\frac{\sigma^2 t^2}{2}-i\mu t}\left[1 - \Phi\left(\frac{\mu}{\sigma} - i\sigma t\right)\right] \tag{21}$$

Finally, the characteristic function becomes:

$$\varphi_x(t) = e^{\frac{-\sigma^2 t^2}{2}+i\mu t}\left[1 - \Phi\left(-\frac{\mu}{\sigma} + i\sigma t\right)\right] + e^{-\frac{\sigma^2 t^2}{2}-i\mu t}\left[1 - \Phi\left(\frac{\mu}{\sigma} + i\sigma t\right)\right] \tag{22}$$

Below, we list some more functions that include expectations.

1. The moment generating function of Equation (2) exists and is equal to:

$$M_x(t) = \varphi_x(-it) = e^{\frac{\sigma^2 t^2}{2}+\mu t}\left[1 - \Phi\left(-\frac{\mu}{\sigma} - \sigma t\right)\right] + e^{\frac{\sigma^2 t^2}{2}-\mu t}\left[1 - \Phi\left(\frac{\mu}{\sigma} - \sigma t\right)\right] \tag{23}$$

We can see that the characteristic generating function can be differentiated infinitely many times, since the first derivative contains the density of the normal distribution, and thus, it always contains some exponential terms. The folded normal distribution is not a stable distribution. That is, the distribution of the sum of its random variables do not form a folded normal distribution. We can see this from the characteristic (or the moment) generating function Equation (22) or Equation (23).

2. The cumulant generating function is simply the logarithm of the moment generating function:

$$K_x(t) = \log M_x(t) = \left(\frac{\sigma^2 t^2}{2} + \mu t\right)\log\left\{1 - \Phi\left(-\frac{\mu}{\sigma} - \sigma t\right) + e^{-2\mu t}\left[1 - \Phi\left(\frac{\mu}{\sigma} - \sigma t\right)\right]\right\} \tag{24}$$

3. The Laplace transformation can easily be derived from the moment generating function and is equal to:

$$E\left(e^{-tx}\right) = e^{\frac{\sigma^2 t^2}{2}-\mu t}\left[1 - \Phi\left(-\frac{\mu}{\sigma} + \sigma t\right)\right] + e^{\frac{\sigma^2 t^2}{2}+\mu t}\left[1 - \Phi\left(\frac{\mu}{\sigma} + \sigma t\right)\right] \tag{25}$$

4. The Fourier transformation is:

$$\hat{f}(t) = \int_{-\infty}^\infty e^{-2\pi i x t} f(x)\,dx = E\left(e^{-2\pi i X t}\right) \tag{26}$$

However, this is closely related to the characteristic function. We can see that $E\left(e^{-2\pi i x t}\right) = \phi_x(-2\pi t)$. Thus, Equation (26) becomes:

$$\hat{f}(t) = \phi_x(-2\pi t) = \quad e^{\frac{-4\pi^2\sigma^2 t^2}{2}-i2\pi\mu t}\left[1 - \Phi\left(-\frac{\mu}{\sigma} - i2\pi\sigma t\right)\right] \tag{27}$$

$$+ \quad e^{-\frac{4\pi^2\sigma^2 t^2}{2}+i2\pi\mu t}\left[1 - \Phi\left(\frac{\mu}{\sigma} - i2\pi\sigma t\right)\right] \tag{28}$$

5. The mean residual life is given by:

$$E(X - t|X > t) = E(X|X > t) - t \tag{29}$$

where $t \in \mathbb{R}_+$. The above conditional expectation is given by:

$$E(X|X > t) = \int_t^\infty \frac{xf(x)}{P(x > t)}dx = \int_t^\infty \frac{xf(x)}{1 - F(t)}dx \tag{30}$$

The denominator in Equation (30) is written as $1 - \frac{1}{2}\left[erf\left(\frac{x-\mu}{\sqrt{2\sigma^2}}\right) + erf\left(\frac{x+\mu}{\sqrt{2\sigma^2}}\right)\right]$. The contents within the integral in the numerator of Equation (30) could be replaced by $1 - F(t)$, as well, but we will not replace it. The calculation of the numerator is done in the same way as the calculation of the mean. Thus:

$$
\int_t^\infty x f(x)\, dx = \int_t^\infty x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx + \int_t^\infty x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x+\mu)^2} dx \tag{31}
$$

$$
= \frac{\sigma}{\sqrt{2\pi}} e^{\frac{(t-\mu)^2}{\sigma^2}} + \mu\left[1 - \Phi\left(\frac{t-\mu}{\sigma}\right)\right] + \frac{\sigma}{\sqrt{2\pi}} e^{\frac{(t-\mu)^2}{\sigma^2}} - \mu\Phi\left(\frac{t-\mu}{\sigma}\right) \tag{32}
$$

$$
= \sqrt{\frac{2}{\pi}}\sigma e^{\frac{(t-\mu)^2}{\sigma^2}} + \mu\left[1 - 2\Phi\left(\frac{t-\mu}{\sigma}\right)\right] \tag{33}
$$

Finally, Equation (30) can be written as:

$$
E(X - t|X > t) = \frac{\sqrt{\frac{2}{\pi}}\sigma e^{\frac{(t-\mu)^2}{\sigma^2}} + \mu\left[1 - 2\Phi\left(\frac{t-\mu}{\sigma}\right)\right]}{1 - \frac{1}{2}\left[erf\left(\frac{x-\mu}{\sqrt{2\sigma^2}}\right) + erf\left(\frac{x+\mu}{\sqrt{2\sigma^2}}\right)\right]} - t \tag{34}
$$

3. Entropy and Kullback–Leibler Divergence

When studying a distribution, the entropy and the Kullback–Leibler divergence from some other distributions are two measures that have to be calculated. In this case, we tried to approximate both of these quantities using a Taylor series. Numerical examples are displayed to show the performance of the approximations.

3.1. Entropy

The entropy is defined as the negative expectation of $-\log f(x)$.

$$
\begin{aligned}
E &= E[-\log f(x)] = -\int_0^\infty \log f(x) f(x)\, dx \\
&= -\int_0^\infty f(x) \log\left\{\frac{1}{\sqrt{2\pi\sigma^2}}\left[e^{-\frac{1}{2\sigma^2}(x-\mu)^2} + e^{-\frac{1}{2\sigma^2}(x+\mu)^2}\right]\right\} dx \\
&= \log\sqrt{2\pi\sigma^2}\int_0^\infty f(x)\, dx - \int_0^\infty f(x) \log\left[e^{-\frac{(x-\mu)^2}{2\sigma^2}}\left(1 + \frac{e^{\frac{(x+\mu)^2}{2\sigma^2}}}{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}\right)\right] dx \\
&= \log\sqrt{2\pi\sigma^2} + \int_0^\infty \frac{x^2 - 2\mu x + \mu^2}{2\sigma^2} f(x) - \int_0^\infty f(x) \log\left(1 + e^{-\frac{2\mu x}{\sigma^2}}\right) dx \tag{35}
\end{aligned}
$$

Let us now take the second term of Equation (35) and see what is equal to:

$$
\frac{1}{2\sigma^2}\int_0^\infty x^2 f(x) = \frac{\mu^2 + \sigma^2}{2\sigma^2} \quad \text{by exploiting the knowledge of variance Equation (8)} \tag{36}
$$

$$
\frac{-2\mu}{2\sigma^2}\int_0^\infty x f(x) = -\mu\frac{\mu_f}{\sigma^2} \quad \text{since the first moment is given in Equation (7) and} \tag{37}
$$

$$\frac{\mu^2}{2\sigma^2} \int_0^\infty f(x) = \frac{\mu^2}{2\sigma^2} \tag{38}$$

Finally, the third term of Equation (35) is equal to:

$$A_n = -\int_0^\infty f(x) \sum_{n=1}^\infty \frac{(-1)^{n+1}}{n} e^{-\frac{2n\mu x}{\sigma^2}} dx \tag{39}$$

by making use of the Taylor expansion for $\log(1+x)$ around zero, but instead of x, we have $e^{-\frac{2\mu x}{\sigma^2}}$. Thus, we have managed to "break" the second integral of entropy Equation (35) down to smaller pieces of:

$$
\begin{aligned}
A_n &= -\int_0^\infty \sum_{n=1}^\infty \frac{(-1)^{n+1}}{n} e^{a_n x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx - \int_0^\infty \sum_{n=1}^\infty \frac{(-1)^{n+1}}{n} e^{a_n x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x+\mu)^2} dx \\
&= -\sum_{n=1}^\infty \frac{(-1)^{n+1}}{n} e^{\frac{(\mu+a_n\sigma^2)^2 - \mu^2}{2\sigma^2}} \left[1 - \Phi\left(-\frac{\mu}{\sigma} - \frac{a_n}{\sigma} \right) \right] \\
&\quad - \sum_{n=1}^\infty \frac{(-1)^{n+1}}{n} e^{\frac{(\mu-a_n\sigma^2)^2 - \mu^2}{2\sigma^2}} \left[1 - \Phi\left(\frac{\mu}{\sigma} - \frac{a_n}{\sigma} \right) \right]
\end{aligned}
$$

by interchanging the order of the summation and the integration, filling up the square in the same way to the characteristic function and with $a_n = -\frac{2n\mu}{\sigma^2}$. The final form of the entropy is given in Equation (40):

$$
\begin{aligned}
E &\simeq \log\sqrt{2\pi\sigma^2} + \frac{1}{2} + \frac{\mu^2 - \mu\mu_f}{\sigma^2} - \sum_{n=1}^\infty \frac{(-1)^{n+1}}{n} e^{\frac{(\mu-2n\mu)^2 - \mu^2}{2\sigma^2}} \left[1 - \Phi\left(-\frac{\mu}{\sigma} + \frac{-\frac{2n\mu}{\sigma^2}}{\sigma} \right) \right] \\
&\quad - \sum_{n=1}^\infty \frac{(-1)^{n+1}}{n} e^{\frac{(\mu-2n\mu)^2 - \mu^2}{2\sigma^2}} \left[1 - \Phi\left(\frac{\mu}{\sigma} + \frac{-\frac{2n\mu}{\sigma^2}}{\sigma} \right) \right]
\end{aligned} \tag{40}
$$

Figure 2 shows the true value of Equation (40), when $\sigma = 5$ and μ ranges from zero to 25, thus for values of $\theta = \frac{\mu}{\sigma}$ from zero to five. The true value was calculated using numerical integration. R provides this option with the command *integrate*. The second and third order approximations (using the first two and three terms of the infinite sums in Equation (40)), are also displayed for comparison.

Figure 2. Entropy values for a range of values of $\theta = \frac{\mu}{\sigma}$ with $\sigma = 1$ (**a**) and $\sigma = 5$ (**b**).
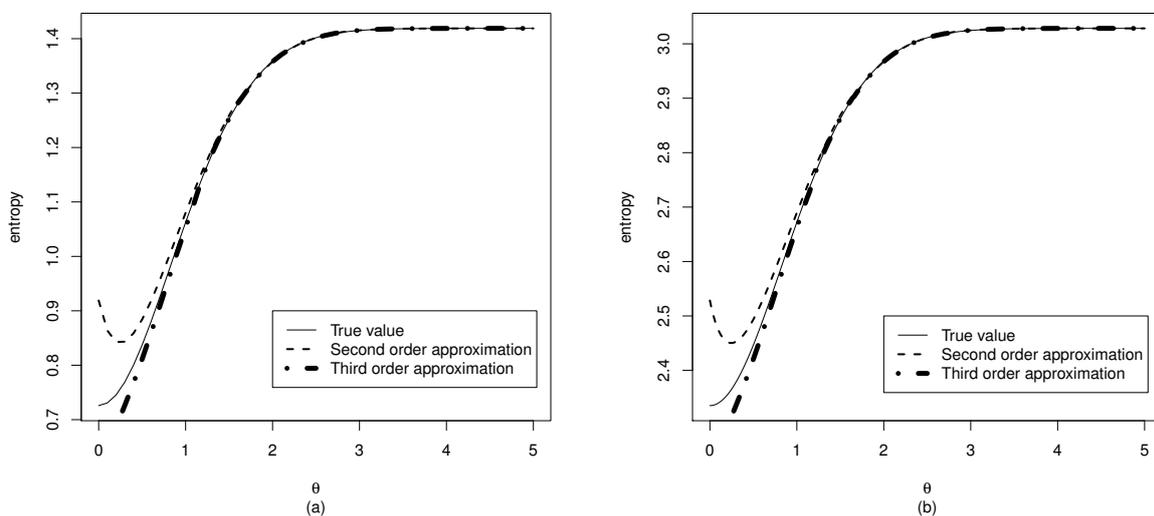
We can see that the second order approximation is not as good as the third order, especially for small values of θ. The Taylor approximation of Equation (40) is valid when the value, a_n, is close to zero. As with the logarithm approximation, the expansion is around zero; thus, when we start going further away from zero, the approximation loses its accuracy. The same is true in our case. When the values of θ are small, then the value of $\log\left(1 + e^{-\frac{2\mu x}{\sigma^2}}\right)$ is far from zero. As θ increases, and, thus, the exponential term decreases, the Taylor series approximates true value better. This is why we see a small discrepancy of the approximations on the left of Figure 2, which become negligible later on.

3.2. Kullback–Leibler Divergence from the Normal Distribution

The Kullback–Leibler divergence [14] of one distribution from another in general is defined as the expectation of the logarithm of the ratio of the two distributions with respect to the first one:

$$KL\left(f|g\right) = E_f\left[\log\frac{f}{g}\right] = \int f\left(x\right)\log\frac{f\left(x\right)}{g\left(x\right)}dx$$

The divergence of the folded normal distribution from the normal distribution is equal to:

$$
\begin{aligned}
KL(FN||N) &= \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}}\left[e^{-\frac{1}{2\sigma^2}(x-\mu)^2} + e^{-\frac{1}{2\sigma^2}(x+\mu)^2}\right]\log\frac{\frac{1}{\sqrt{2\pi\sigma^2}}\left[e^{-\frac{1}{2\sigma^2}(x-\mu)^2} + e^{-\frac{1}{2\sigma^2}(x+\mu)^2}\right]}{\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}}dx \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}}\left[e^{-\frac{1}{2\sigma^2}(x-\mu)^2} + e^{-\frac{1}{2\sigma^2}(x+\mu)^2}\right]\log\left(1 + e^{-\frac{2\mu x}{\sigma^2}}\right)dx
\end{aligned}
$$

which is the same as the second integral of Equation (35). Thus, we can approximate this divergence by the same Taylor series:

$$
\begin{aligned}
KL(FN||N) &\simeq \sum_{n=1}^\infty \frac{(-1)^{n+1}}{n}e^{\frac{(\mu-2n\mu)^2-\mu^2}{2\sigma^2}}\left[1 - \Phi\left(-\frac{\mu}{\sigma} + \frac{-\frac{2n\mu}{\sigma^2}}{\sigma}\right)\right] \\
&+ \sum_{n=1}^\infty \frac{(-1)^{n+1}}{n}e^{\frac{(\mu-2n\mu)^2-\mu^2}{2\sigma^2}}\left[1 - \Phi\left(\frac{\mu}{\sigma} + \frac{-\frac{2n\mu}{\sigma^2}}{\sigma}\right)\right]
\end{aligned}
$$

Figure 3. Kullback–Leibler divergence from the normal for a range of values of $\theta = \frac{\mu}{\sigma}$ with $\sigma = 1$ (**a**) and $\sigma = 5$ (**b**).
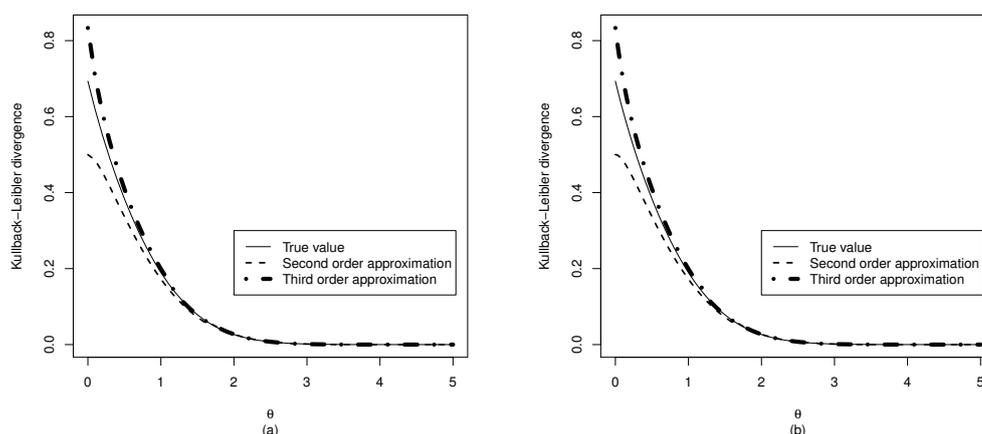
Figure 3 presents two cases of the Kullback–Leibler divergence, for illustration purposes, when the first two and three terms of the infinite sum have been used. In the first graph, the standard deviation is equal to one, and in the second case, it is equal to five. The divergence seems independent of the variance. The change occurs as a result of the value of θ. It becomes clear that when the value of the mean to the standard deviation increases, the folded normal converges to the normal distribution.
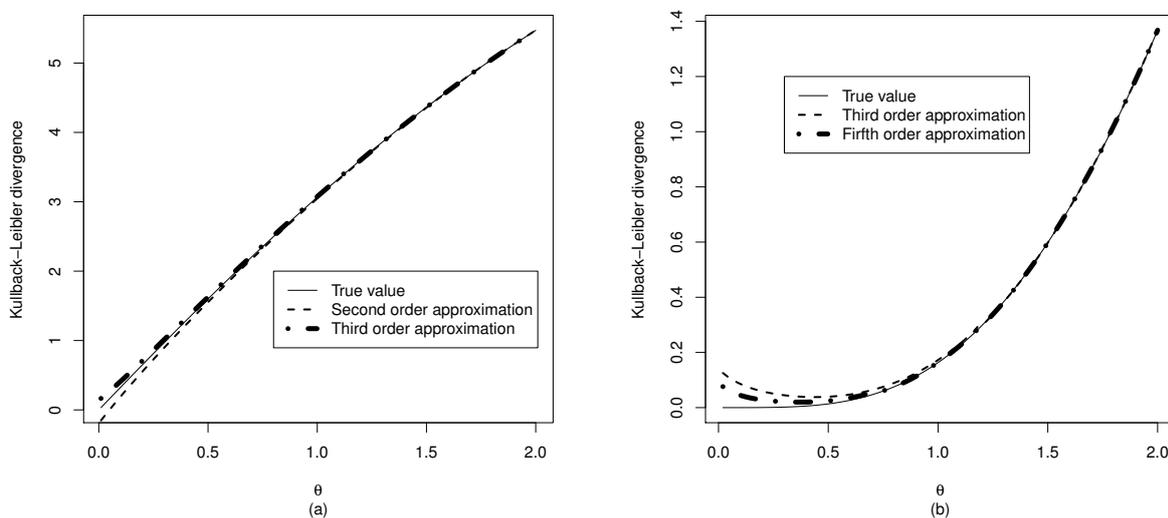
3.3. Kullback–Leibler Divergence from the Half Normal Distribution

As mentioned in Section 2.1, the half normal distribution is a special case of the folded normal distribution with $\mu = 0$. The Kullback–Leilber divergence of the folded normal from the half normal distribution is equal to:

$$
KL(FN\left(\mu, \sigma^2\right)||FN\left(\mu = 0, \sigma^2\right)) =
$$

$$
= \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}} \left[e^{-\frac{1}{2\sigma^2}(x-\mu)^2} + e^{-\frac{1}{2\sigma^2}(x+\mu)^2} \right] \log \frac{\frac{1}{\sqrt{2\pi\sigma^2}} \left[e^{-\frac{1}{2\sigma^2}(x-\mu)^2} + e^{-\frac{1}{2\sigma^2}(x+\mu)^2} \right]}{\frac{2}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}x^2}} dx
$$

$$
= -\log 2 \int_0^\infty f\left(x; \mu, \sigma^2\right) dx + \int_0^\infty f\left(x; \mu, \sigma^2\right) \log \left(e^{-\frac{\mu^2}{2\sigma^2} + \frac{\mu x}{\sigma^2}} + e^{-\frac{\mu^2}{2\sigma^2} - \frac{\mu x}{\sigma^2}} \right) dx
$$

$$
= -\log 2 + \int_0^\infty \left(\frac{\mu x}{\sigma^2} - \frac{\mu^2}{2\sigma^2} \right) f\left(x; \mu, \sigma^2\right) dx + \int_0^\infty f\left(x; \mu, \sigma^2\right) \log \left(1 + e^{-\frac{2\mu x}{\sigma^2}} \right) dx
$$

$$
= -\log 2 + \frac{2\mu\mu_f - \mu^2}{2\sigma^2} + KL(FN||N)
$$

where $f\left(x; \mu, \sigma^2\right)$ stands for the folded normal Equation (2) and μ_f is the expected value given in Equation (7). Figure 4 shows the approximations to the true value when $\sigma = 1$ and $\sigma = 5$. This time, we used the third and fifth order approximations, but even then, for small values of θ, the approximations were not satisfactory.

Figure 4. Kullback–Leibler divergence from the half normal for a range of values of $\theta = \frac{\mu}{\sigma}$ with $\sigma = 1$ (**a**) and $\sigma = 5$ (**b**).

The previous result cannot lead to an inequality regarding the Kullback–Leibler divergences from the two other distributions. When $\mu > \sigma$, then the divergence from the half normal will be greater than the divergence from the normal, and when $\mu < \sigma$, the opposite is true. However, this is not strict, since it can be the case for either inequality that the relationship between the divergences is not true. Instead, we can use it as a rule of thumb in general.

4. Parameter Estimation

We will show two ways of estimating the parameters. The first one can be found in [1], but we review it and add some more details. Both of them are essentially the maximum likelihood estimation procedure, but in the first case, we perform maximization, whereas in the second case, we seek the root of an equation.

The log-likelihood of Equation (2) can be written in the following way:

$$
\begin{aligned}
l &= -\frac{n}{2}\log 2\pi\sigma^2 + \sum_{i=1}^{n}\log\left[e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} + e^{-\frac{(x_i+\mu)^2}{2\sigma^2}}\right] \Rightarrow \\
l &= -\frac{n}{2}\log 2\pi\sigma^2 + \sum_{i=1}^{n}\log\left[e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}\left(1 + e^{-\frac{(x_i+\mu)^2}{2\sigma^2}}e^{\frac{(x_i-\mu)^2}{2\sigma^2}}\right)\right] \Rightarrow \\
l &= -\frac{n}{2}\log 2\pi\sigma^2 - \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2} + \sum_{i=1}^{n}\log\left(1 + e^{-\frac{2\mu x_i}{\sigma^2}}\right)
\end{aligned}
\tag{41}
$$

where n is the sample size of the x_i values. The partial derivatives of Equation (41) are:

$$
\begin{aligned}
\frac{\partial l}{\partial \mu} &= \frac{\sum_{i=1}^{n}(x_i-\mu)}{\sigma^2} - \frac{2}{\sigma^2}\sum_{i=1}^{n}\frac{x_i e^{\frac{-2\mu x_i}{\sigma^2}}}{1 + e^{\frac{-2\mu x_i}{\sigma^2}}} = \frac{\sum_{i=1}^{n}(x_i-\mu)}{\sigma^2} - \frac{2}{\sigma^2}\sum_{i=1}^{n}\frac{x_i}{1 + e^{\frac{2\mu x_i}{\sigma^2}}}, \quad \text{and} \\
\frac{\partial l}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^4} + \frac{2\mu}{\sigma^4}\sum_{i=1}^{n}\frac{x_i e^{-\frac{2\mu x_i}{\sigma^2}}}{1 + e^{-\frac{2\mu x_i}{\sigma^2}}} \Rightarrow \\
\frac{\partial l}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^4} + \frac{2\mu}{\sigma^4}\sum_{i=1}^{n}\frac{x_i}{1 + e^{\frac{2\mu x_i}{\sigma^2}}}
\end{aligned}
$$

By equating the first derivative of the log-likelihood to zero, we obtain a nice relationship:

$$
\sum_{i=1}^{n}\frac{x_i}{1 + e^{\frac{2\mu x_i}{\sigma^2}}} = \frac{\sum_{i=1}^{n}(x_i-\mu)}{2}
\tag{42}
$$

Note that Equation (42) has three solutions, one at zero and two more with the opposite sign. The example in Section 4.1 will show graphically the three solutions. By substituting Equation (42), to the derivative of the log-likelihood w.r.t σ^2 and equating to zero, we get the following expression for the variance:

$$
\sigma^2 = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n} + \frac{2\mu\sum_{i=1}^{n}(x_i-\mu)}{n} = \frac{\sum_{i=1}^{n}(x_i^2-\mu^2)}{n} = \frac{\sum_{i=1}^{n}x_i^2}{n} - \mu^2
\tag{43}
$$

The above relationships Equations (42) and (43) can be used to obtain maximum likelihood estimates in an efficient recursive way. We start with an initial value for σ^2 and find the positive root of

Equation (42). Then, we insert this value of μ in Equation (43) and get an updated value of σ^2. The procedure is being repeated until the change in the log-likelihood value is negligible.

Another easier and more efficient way is to perform a search algorithm. Let us write Equation (42) in a more elegant way.

$$2\sum_{i=1}^{n} \frac{x_i}{1 + e^{\frac{2\mu x_i}{\sigma^2}}} - \sum_{i=1}^{n} \frac{x_i\left(1 + e^{\frac{2\mu x_i}{\sigma^2}}\right)}{1 + e^{\frac{2\mu x_i}{\sigma^2}}} + n\mu \;=\; 0 \Rightarrow$$

$$\sum_{i=1}^{n} \frac{x_i\left(1 - e^{\frac{2\mu x_i}{\sigma^2}}\right)}{1 + e^{\frac{2\mu x_i}{\sigma^2}}} + n\mu \;=\; 0$$

where σ^2 is defined in Equation (43). It becomes clear that the optimization the log-likelihood Equation (41) with respect to the two parameters has turned into a root search of a function with one parameter only. We tried to perform maximization via the E-M algorithm, treating the sign as the missing information, but it did not prove very good in this case.

4.1. An Example with Simulated Data

We generated 100 random values from the $FN(2, 9)$ in order to illustrate the maximum likelihood estimation procedure. The estimated parameter values were equal to $(\hat{\mu} = 2.183, \hat{\sigma}^2 = 8.065)$. The corresponding 95% confidence intervals for μ and σ^2 were $(0.782, 3.585)$ and $(2.022, 14.108)$ respectively. Figure 5 shows graphically the existence of the three extrema of the log-likelihood Equation (41), one minimum (always at zero) and two maxima at the maximum likelihood estimates of μ.

Figure 5. The left graph (**a**) shows the three solutions of the log-likelihood. The right three-dimensional figure (**b**) shows the values of the log-likelihood for a range of mean and variance values.
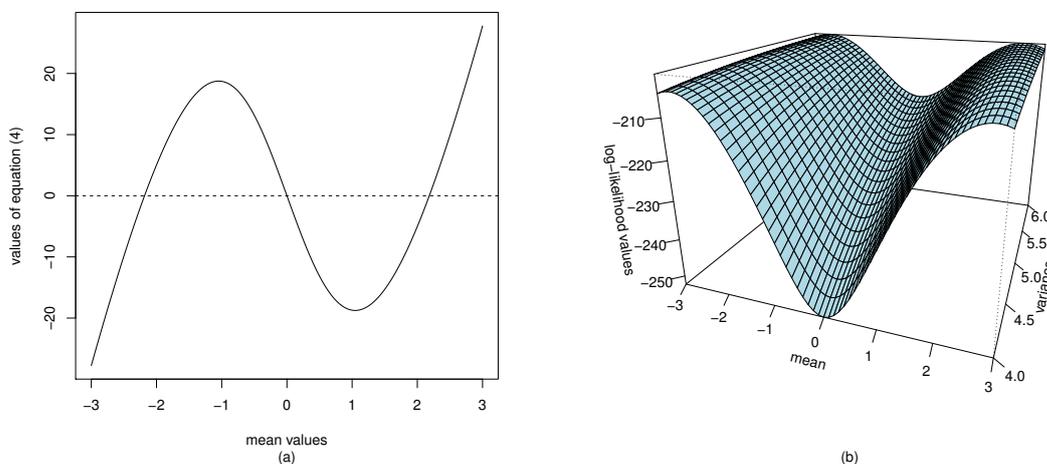
4.2. Simulation Studies

Simulation studies were implemented to examine the accuracy of the estimates using numerical optimization based on the simplex method [10]. Numerical optimization was performed in [15], using the *optim* function. The term accuracy refers to interval estimation rather than point estimation, since the interest was on constructing confidence intervals for the parameters. The number of simulations was set equal to R = 1,000. The sample sizes ranged from 20 to 100 for a range of values of the parameter vector. The R-package *VGAM*[16] offers algorithms for obtaining maximum likelihood estimates of the folded normal, but we have not used it here.

For every simulation, we calculated 95% confidence intervals using the normal approximation, where the variance was estimated from the inverse of the observed information matrix. The maximum likelihood estimates are asymptotically normal with variance equal to the inverse of the Fisher's information. The sample estimate of this information is given by the second derivative (Hessian matrix) of the log-likelihood with respect to the parameter. This is an asymptotic confidence interval.

Bootstrap confidence intervals were also calculated using the percentile method [17]. For every simulation, we produced the bootstrap distribution of the data with $B = 1000$ bootstrap repetitions. Thus, we calculated the 2.5% lower and upper quantiles for each of the parameters. In addition, we calculated the correlations for every pair of the parameters.

Tables 1 to 4 present the coverage of the 95% confidence intervals for the two parameters at different pairs of sample size and mean. The rows correspond to the sample size, whereas the columns correspond to the ratio $\theta = \frac{\mu}{\sigma}$, with $\sigma = 5$ fixed.

Table 1. Estimated coverage probability of the 95% confidence intervals for the mean parameter, μ, using the observed information matrix.

	Values	**of**	**θ**					
Sample size	**0.5**	**1**	**1.5**	**2**	**2.5**	**3**	**3.5**	**4**
20	0.689	0.930	0.955	0.931	0.926	0.940	0.930	0.948
30	0.679	0.921	0.949	0.943	0.925	0.926	0.941	0.915
40	0.690	0.916	0.936	0.933	0.941	0.948	0.944	0.928
50	0.718	0.944	0.955	0.938	0.933	0.948	0.946	0.946
60	0.699	0.950	0.968	0.948	0.949	0.941	0.942	0.946
70	0.721	0.931	0.956	0.939	0.939	0.939	0.949	0.945
80	0.691	0.930	0.950	0.940	0.946	0.936	0.945	0.939
90	0.720	0.932	0.960	0.949	0.949	0.939	0.954	0.944
100	0.738	0.945	0.949	0.938	0.943	0.926	0.946	0.952

What can be seen from Tables 1 and 2 is that whist the sample size is important, the value of θ, the mean to standard deviation ratio, is more important. As this ration increase the coverage probability increases, as well, and reaches the desired nominal 95%. This is also true for the bootstrap confidence intervals, but the coverage is in general higher and increases faster as the sample size increases in contrast to the asymptotic confidence interval. What is more is that when the value of θ is less than one, the

bootstrap confidence interval is to be preferred. When the value of θ becomes equal to or more than one, then both the bootstrap and the asymptotic confidence intervals produce similar coverages.

The results regarding the variance are presented in Tables 3 and 4. When the value of θ is small, both ways of obtaining confidence intervals for this parameter are rather conservative. The bootstrap intervals tend to perform better, but not up to the expectations. Even when the value of θ is large, if the sample sizes are not large enough, the nominal coverage of 95% is not attained.

Table 2. Estimated coverage probability of the bootstrap 95% confidence intervals for the mean parameter, μ, using the percentile method.

	Values		**of**		θ			
Sample size	**0.5**	**1**	**1.5**	**2**	**2.5**	**3**	**3.5**	**4**
20	0.890	0.925	0.939	0.921	0.918	0.940	0.929	0.942
30	0.894	0.931	0.933	0.943	0.926	0.922	0.942	0.910
40	0.910	0.925	0.927	0.933	0.941	0.947	0.946	0.928
50	0.914	0.943	0.942	0.934	0.934	0.945	0.946	0.943
60	0.904	0.949	0.953	0.950	0.941	0.938	0.943	0.944
70	0.893	0.934	0.943	0.936	0.937	0.938	0.949	0.939
80	0.918	0.940	0.939	0.939	0.944	0.935	0.946	0.938
90	0.920	0.934	0.952	0.948	0.946	0.939	0.951	0.947
100	0.918	0.940	0.936	0.932	0.946	0.925	0.945	0.949

Table 3. Estimated coverage probability of the 95% confidence intervals for the variance parameter, σ^2, using the observed information matrix.

	Values		**of**		θ			
Sample size	**0.5**	**1**	**1.5**	**2**	**2.5**	**3**	**3.5**	**4**
20	0.649	0.765	0.854	0.853	0.876	0.870	0.862	0.885
30	0.697	0.794	0.870	0.898	0.892	0.898	0.894	0.896
40	0.723	0.849	0.893	0.914	0.919	0.913	0.909	0.902
50	0.751	0.867	0.916	0.907	0.911	0.924	0.899	0.912
60	0.745	0.865	0.911	0.913	0.916	0.906	0.920	0.933
70	0.769	0.874	0.928	0.928	0.912	0.930	0.926	0.935
80	0.776	0.883	0.927	0.919	0.934	0.936	0.916	0.924
90	0.795	0.901	0.931	0.932	0.925	0.930	0.940	0.941
100	0.824	0.904	0.927	0.933	0.925	0.936	0.932	0.942

The correlation between the two parameters was also estimated for every simulation from the observed information matrix. The results are displayed in Table 5. The correlation between the two parameters is always negative irrespective of the sample size or the value of θ, except for the case when $\theta = 4$. In this case, the correlation becomes zero as expected. As the value of θ grows larger, the probability of the normal distribution, which lies on the negative axis, becomes smaller until it becomes negligible. In this

case, the distribution equals the classical normal distribution for which the two parameters are known to be orthogonal.

Table 4. Estimated coverage probability of the bootstrap 95% confidence intervals for the variance parameter, σ^2, using the percentile method.

	Values		of		θ			
Sample size	0.5	1	1.5	2	2.5	3	3.5	4
20	0.657	0.814	0.862	0.842	0.840	0.832	0.818	0.824
30	0.701	0.850	0.885	0.891	0.882	0.867	0.869	0.866
40	0.743	0.881	0.896	0.913	0.912	0.886	0.881	0.878
50	0.772	0.895	0.921	0.916	0.897	0.901	0.885	0.892
60	0.797	0.907	0.912	0.910	0.906	0.897	0.907	0.916
70	0.807	0.904	0.925	0.915	0.909	0.918	0.908	0.924
80	0.822	0.895	0.925	0.914	0.925	0.917	0.909	0.909
90	0.869	0.916	0.932	0.922	0.919	0.915	0.934	0.929
100	0.873	0.915	0.918	0.925	0.906	0.931	0.920	0.939

Table 5. Estimated correlations between the two parameters obtained from the observed information matrix.

	Values		of		θ			
Sample size	0.5	1	1.5	2	2.5	3	3.5	4
20	-0.600	-0.495	-0.272	-0.086	-0.025	-0.006	-0.001	0.000
30	-0.638	-0.537	-0.262	-0.089	-0.022	-0.005	-0.001	0.000
40	-0.695	-0.548	-0.251	-0.081	-0.021	-0.005	-0.001	0.000
50	-0.723	-0.580	-0.259	-0.076	-0.020	-0.005	-0.001	0.000
60	-0.750	-0.597	-0.251	-0.075	-0.019	-0.004	-0.001	0.000
70	-0.771	-0.588	-0.256	-0.073	-0.019	-0.004	-0.001	0.000
80	-0.774	-0.604	-0.253	-0.074	-0.019	-0.004	-0.001	0.000
90	-0.796	-0.599	-0.245	-0.073	-0.018	-0.004	-0.001	0.000
100	-0.804	-0.611	-0.252	-0.072	-0.019	-0.004	-0.001	0.000

Table 6 shows the probability of a normal random variable being less than zero when $\sigma = 5$ and the same values of θ as in the simulation studies.

**Table 6.** Probability of a normal variable having negative values.

Values		of		θ			
0.5	1	1.5	2	2.5	3	3.5	4
0.309	0.159	0.067	0.023	0.006	0.001	0.000	0.000

When the ratio of mean to standard deviation is small, the area of the normal distribution in the negative side is large, and as the value of this ratio increases, the probability decreases until it becomes zero. In this case, the folded normal is the normal distribution, since there are no negative values to fold on to the positive side. This of course is in accordance with all the previous observations and results we saw.

5. Application to Body Mass Index Data

We fitted the folded normal distribution on real data. These are observations of the the body mass index of 700 New Zealand adults, accessible via the R package *VGAM* [16]. These measurements are a random sample from the Fletcher Challenge/Auckland Heart and Health survey conducted in the early 1990s [18]. Figure 6 contains a histogram of the data along with the parametric (folded normal) and the non-parametric (kernel) density estimation. It should be noted that the fitted folded normal here converges in distribution to the normal.

Figure 6. The histogram on the left shows the body mass indices of 700 New Zealand adults. The green line is the fitted folded normal and the blue line is the kernel density. The perspective plot on the right shows the log-likelihood of the body mass index data as a function of the mean and the variance.
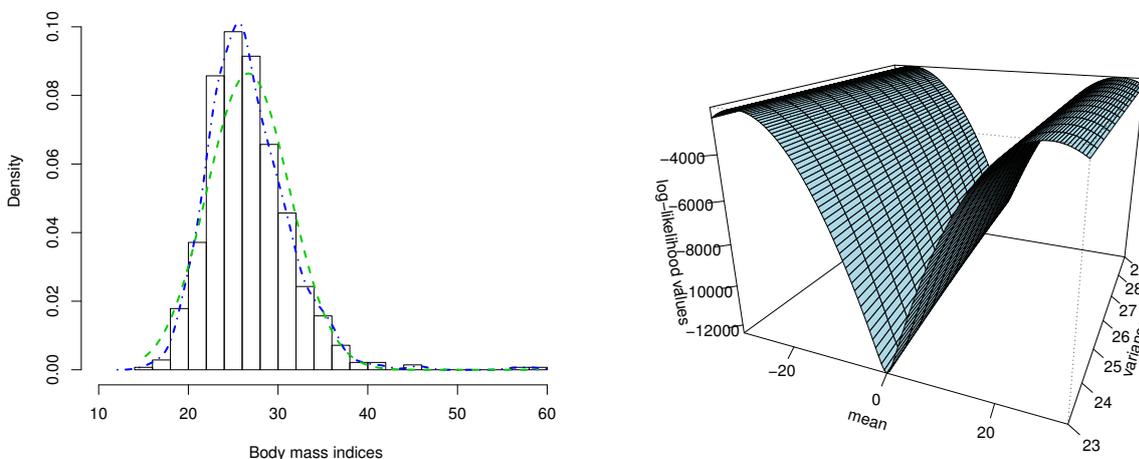
The estimated parameters (using the *optim* command in R) were $\hat{\mu} = 26.685(0.175)$ and $\hat{\sigma}^2 = 21.324(1.140)$, with their standard error appearing inside the parentheses. Since the sample size is very large, there is no need to estimate their standard errors and, consequently, 95% confidence intervals, even though their ratio is only 1.251. Their estimated correlation coefficient was very close to zero (2×10^{-4}), and the estimated probability of the folded normal with these parameters below zero is equal to zero.

6. Discussion

We derived the characteristic function of this distribution and, thus, its moment function. The cumulant generating function is simply the logarithm of the moment generating function, and

therefore, it is easy to calculate. The importance of these two functions is that they allow us to calculate all the moments of the distribution. In addition, we calculated the Laplace and Fourier transformations and the mean residual life.

The entropy of the folded normal distribution and the Kullback–Leibler divergence of this distribution from the normal and half normal distributions were approximated using the Taylor series. The results were numerically evaluated against the true values and were as expected.

We reviewed the maximum likelihood estimates and simplified their calculation and saw some properties of them. Confidence intervals for the parameters were obtained using the asymptotic theory and the bootstrap methodology under the umbrella of simulation studies.

The coverage of the confidence intervals for the two parameters was lower than the desired nominal in the small sample cases and when the mean to standard deviation ratio was lower than one. An alternative way to correct the under-coverage of the mean parameter is to use an alternative parametrization. The parameters $\theta = \frac{\mu}{\sigma}$ and σ are calculated in [5]. If we use θ and μ, then the coverage of the interval estimation of μ is corrected, but the corresponding coverage of the confidence interval for σ^2 is still low.

The correlation between the two parameters was always negative and decreasing as the value of θ was increasing, as expected, until the two parameters become independent.

An application of the folded normal distribution to real data was exhibited, providing evidence that it can be used to model non-negative data adequately.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Leone, F.C.; Nelson, L.S.; Nottingham, R.B. The folded normal distribution. *Technometrics* **1961**, *3*, 543–550.
2. Lin, H.C. The measurement of a process capability for folded normal process data. *Int. J. Adv. Manuf. Technol.* **2004**, *24*, 223–228.
3. Chakraborty, A.K.; Chatterjee, M. On multivariate folded normal distribution. *Sankhya* **2013**, *75*, 1–15.
4. Elandt, R.C. The folded normal distribution: Two methods of estimating parameters from moments. *Technometrics* **1961**, *3*, 551–562.
5. Johnson, N.L. The folded normal distribution: Accuracy of estimation by maximum likelihood. *Technometrics* **1962**, *4*, 249–256.
6. Johnson, N.L. Cumulative sum control charts for the folded normal distribution. *Technometrics* **1963**, *5*, 451–458.
7. Sundberg, R. On estimation and testing for the folded normal distribution. *Commun. Stat.-Theory Methods* **1974**, *3*, 55–72.
8. Kim, H.J. On the ratio of two folded normal distributions. *Commun. Stat.-Theory Methods* **2006**, *35*, 965–977.

9. Liao, M.Y. Economic tolerance design for folded normal data. *Int. J. Prod. Res.* **2010**, *48*, 4123–4137.

10. Nelder, J.A.; Mead, R. A simplex method for function minimization. *Comput. J.* **1965**, *7*, 308–313.

11. Johnson, N.L; Kotz, S.; Balakrishnan, N. *Continuous Univariate Distributions*; John Wiley & Sons, Inc.: New York, NY, USA, 1994.

12. Psarakis, S.; Panaretos, J. The folded t distribution. *Commun. Stat.-Theory Methods* **1990**, *19*, 2717–2734.

13. Psarakis, S.; Panaretos, J. On some bivariate extensions of the folded normal and the folded t distributions. *J. Appl. Stat. Sci.* **2000**, *10*, 119–136.

14. Kullback, S. *Information Theory and Statistics*; Dover Publications: New York, NY, USA, 1977.

15. R Development Core Team. R: A Language and Environment for Statistical Computing, 2012. Available online: http://www.R-project.org/ (accessed on 1 December 2013).

16. Yee, T.W. The VGAM package for categorical data analysis. *J. Stat. Softw.* **2010**, *32*, 1–34.

17. Efron, B.; Tibshirani, R. *An Introduction to the Bootstrap*; Chapman and Hall/CRC: New York, NY, USA, 1993.

18. MacMahon, S.; Norton, R.; Jackson, R.; Mackie, M.J.; Cheng, A.; Vander Hoorn, S.; Milne, A.; McCulloch, A. Fletcher challenge-university of Auckland heart and health study: Design and baseline findings. *N. Zeal. Med. J.* **1995**, *108*, 499–502.