Journal of Statistical Computation and Simulation

ISSN: 0094-9655 (Print) 1563-5163 (Online) Journal homepage: http://www.tandfonline.com/loi/gscs20

A simple algorithm for calculating values for folded normal distribution

Moutushi Chatterjee & Ashis Kumar Chakraborty

To cite this article: Moutushi Chatterjee & Ashis Kumar Chakraborty (2016) A simple algorithm for calculating values for folded normal distribution, Journal of Statistical Computation and Simulation, 86:2, 293-305, DOI: 10.1080/00949655.2015.1007984

To link to this article: http://dx.doi.org/10.1080/00949655.2015.1007984

Published online: 03 Feb 2015.

Journal of Statistical Computation and Simulation, 2016
Vol. 86, No. 2, 293–305, http://dx.doi.org/10.1080/00949655.2015.1007984

Taylor & Francis
Taylor & Francis Group

A simple algorithm for calculating values for folded normal distribution

Moutushi Chatterjee* and Ashis Kumar Chakraborty

SQC & OR Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata 700108, India

(*Received 23 June 2014; accepted 12 January 2015*)

Folded normal distribution originates from the modulus of normal distribution. In the present article, we have formulated the cumulative distribution function (cdf) of a folded normal distribution in terms of standard normal cdf and the parameters of the mother normal distribution. Although cdf values of folded normal distribution were earlier tabulated in the literature, we have shown that those values are valid for very particular situations. We have also provided a simple approach to obtain values of the parameters of the mother normal distribution from those of the folded normal distribution. These results find ample application in practice, for example, in obtaining the so-called upper and lower α-points of folded normal distribution, which, in turn, is useful in testing of the hypothesis relating to folded normal distribution and in designing process capability control chart of some process capability indices. A thorough study has been made to compare the performance of the newly developed theory to the existing ones. Some simulated as well as real-life examples have been discussed to supplement the theory developed in this article. Codes (generated by *R* software) for the theory developed in this article are also presented for the ease of application.

Keywords: cumulative distribution function; folded normal distribution; normal distribution; parameters; standard normal cdf; statistical table

2010 Mathematics Subject Classifications: 62Q05; 65C60; 46N30

1. Introduction

Folded normal distribution originates when, for example, the actual algebraic signs of the observations, coming from a normal distribution, are irretrievably lost. Leone et al. [1] first studied the properties of this distribution. Suppose Z is a random variable such that $Z \sim N(\mu, \sigma^2)$. Then, $X = |Z|$ follows folded normal distribution whose probability density function (pdf) is given by

$$\begin{aligned} f_X(z) &= h_Z(z) + h_Z(-z) \\ &= \frac{1}{\sigma\sqrt{2\pi}}\left[\exp\left\{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2\right\} + \exp\left\{-\frac{1}{2}\left(\frac{z+\mu}{\sigma}\right)^2\right\}\right], \quad z > 0, \end{aligned} \tag{1}$$

where h and f denote the pdfs of the normal and the folded normal distributions, respectively.

*Corresponding author. Email: tushi.stats@gmail.com

Leone et al. [1] have derived the expressions for the mean (μ_f) and variance (σ_f^2) of a folded normal distribution as

$$\mu_\mathrm{f} = \sigma\sqrt{\frac{2}{\pi}}\exp\left(-\frac{\mu^2}{2\sigma^2}\right) + \mu\left[1 - 2\Phi\left(-\frac{\mu}{\sigma}\right)\right] \tag{2}$$

$$\text{and} \quad \sigma_\mathrm{f}^2 = \mu^2 + \sigma^2 - \left\{\sigma\sqrt{\frac{2}{\pi}}\exp\left(-\frac{\mu^2}{2\sigma^2}\right) + \mu\left[1 - 2\Phi\left(-\frac{\mu}{\sigma}\right)\right]\right\}^2 \tag{3}$$

with $\Phi(.)$ being the cdf of univariate standard normal distribution. The subscript 'f' is used here to distinguish the mean and variance of a folded normal distribution from that of normal distribution.

A general expression for the rth moment of folded normal distribution has been formulated by Elandt.[2] For this, the author has proposed two methods of estimating the parameters μ and σ of the *parent* normal distribution, namely (i) based on the first and second raw and central moments of folded normal distribution and (ii) based on its third and fourth raw and central moments.

For bivariate folded normal (BVFN) distribution, suppose $\mathbf{Z} = (Z_1, Z_2)' \sim N_2(\boldsymbol{\mu}^{(2)}, \Sigma^{(2)})$ for $\boldsymbol{\mu}^{(2)} = (\mu_1, \mu_2)'$ and $\Sigma^{(2)} = (\begin{smallmatrix}\sigma_1^2 \sigma_{12}\\ \sigma_{21}\sigma_2^2\end{smallmatrix})$. Then, $(|Z_1|, |Z_2|)$ follows BVFN distribution with mean vector $\boldsymbol{\mu}_\mathrm{f}{}^{(2)}$ and dispersion matrix $\Sigma_\mathrm{f}^{(2)}$, where the superscript '(2)' denotes the dimension of $\mathbf{Z}$. Following Psarakis and Panaretos,[3] the pdf of BVFN distribution can be defined as

$$f_{|Z_1|,|Z_2|}(z_1, z_2) = \sum_{\substack{u=z_1,-z_1\\ v=z_2,-z_2}} h_{Z_1,Z_2}(u, v) \quad \text{for } z_1, z_2 > 0, \tag{4}$$

where $h_{Z_1,Z_2}(\boldsymbol{\mu}^{(2)}, \Sigma^{(2)})$ denotes the pdf of bivariate normal (BVN) distribution with mean vector $\boldsymbol{\mu}^{(2)}$ and variance–covariance matrix $\Sigma^{(2)}$.

However, Psarakis and Panaretos [3] studied mostly the properties of bivariate folded *standard* normal distribution (i.e. when the mother BVN distribution has mean vector as $\boldsymbol{\mu} = (0, 0)'$ and dispersion matrix $\Sigma = I_2$, I_2 being a 2×2 identity matrix) and their approach is difficult to generalize for higher dimensional scenario. To address these problems, Chakraborty and Chatterjee [4] defined the pdf of multivariate (q-variate, say) folded normal distribution as

$$\begin{aligned} f_q(z_1, z_2, \ldots, z_q) &= \sum_{(s_1,s_2,\ldots,s_q)\in \boldsymbol{S(q)}} h_q(s_1z_1, s_2z_2, \ldots, s_qz_q) \\ &= \sum_{(s_1,s_2,\ldots,s_q)\in \boldsymbol{S(q)}} h_q(\Lambda_s^{(q)}\mathbf{z}^{(q)}) \quad \text{for each } z_i > 0, \end{aligned} \tag{5}$$

where $\boldsymbol{Z} = (Z_1, Z_2, \cdots, Z_q)' \sim N_q(\boldsymbol{\mu} = (\mu_1, \mu_2, \cdot, \mu_q)', \Sigma)$, $\boldsymbol{S(q)} = \{\mathbf{s} : \mathbf{s} = (s_1, s_2, \ldots, s_q)$ with $s_i = \pm 1 \forall 1 \leq i \leq q\}$ and $\Lambda_s^{(q)} = \mathrm{diag}(s_1, s_2, \ldots, s_q)$, such that

$$\begin{bmatrix} s_1z_1 \\ s_2z_2 \\ \vdots \\ s_qz_q \end{bmatrix} = \mathrm{diag}(s_1, s_2, \ldots, s_q) \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_q \end{bmatrix} = \Lambda_s^{(q)}\mathbf{z}^{(q)}.$$

Note that here bold faced letters have been used to denote vectors.

Folded normal distribution arises while computing differences or deviations in the context of normally distributed variables or when the algebraic signs of the observations are irretrievably lost. According to Leone et al.,[1] this is one of the most common type of non-normal distribution encountered in practice. Some typical examples of the contexts, where folded normal distribution can be observed, are measurement of flatness or straightness, the distance between two objects and so on.

Folded normal distribution finds ample applications in practice including the field of statistical quality control. Applications of this distribution, in the context of univariate process capability indices (PCIs) have been made by Lin.[5,6] Similarly, the use of CUSUM control chart, under the folded normal set-up has been discussed by Johnson.[7] An economic tolerance design for folded normal data in manufacturing industries has been discussed by Liao.[8]

Leone et al. [1] in their paper on univariate folded normal distribution, tabulated the values of σ_f and μ, expressed in σ units, corresponding to various values of μ_f/σ_f. Also, the authors have tabulated the areas under the folded normal distribution for various values of μ_f/σ_f. However, the same value of μ_f/σ_f can be obtained for various combinations of (μ_f, σ_f) values. Consequently, the tabulated cdf values of folded normal distribution, as obtained following Leone et al.'s [1] approach, will remain the same for a number of values of (μ_f, σ_f), such that they correspond to the same value of μ_f/σ_f.

Although in the recent years, folded normal distribution has grabbed the attention of many researchers particularly in the field of statistical quality control, some of its very interesting properties are yet to be thoroughly explored. For example, while developing a process capability control chart for the popular PCI C_{pk},[9] similar to the process capability control charts of the PCIs C_{pu} and C_{pl},[10] we were required to compute the so-called upper and lower α-points of univariate folded normal distribution for various values of (μ, σ). Similar problems may also arise in case of testing of hypothesis corresponding to the capability study of a process on the basis of C_{pk}. In both of these two cases, we already have information on (μ, σ) but using Leone et al.'s [1] table requires the corresponding μ_f/σ_f values. Moreover, one can have the same μ_f/σ_f value for different individual values of μ_f and σ_f and consequently, their mother normal distributions will also be different. Since, Leone et al.'s [1] table does not provide the exact values of (μ, σ) corresponding to the values of (μ_f, σ_f), it is further difficult to use those two tables for the said purpose. Although Sundberg [11] discussed about some testing problems related to folded normal distribution, the author considered large sample scenario. As a result, the problem of unavailability of cumulative distribution function (cdf) values of folded normal distribution for various values of (μ, σ) or (μ_f, σ_f) remain unresolved.

In the present article, we have discussed about a very simple step-by-step computer-driven algorithm for having (μ, σ) values from the values of (μ_f, σ_f). We have also expressed the cdf of folded normal distribution as a function of two standard normal cdfs. Based on this algorithm and the expression for cdf, the areas of the folded normal distribution can easily be computed for any plausible combinations of (μ_f, σ_f) or (μ, σ). The statistical package 'R' is used to execute the algorithms.

A simple procedure to compute the values of μ and σ, for the given values of μ_f and σ_f, is discussed in Section 2. Section 3 contains the formulation of the cdf of the folded normal distribution as a linear combination of standard normal cdf and the corresponding table with $x = 0.5(0.5)5$, $\mu = 0(0.5)10$ and $\sigma = 1$. Procedure given in Section 3 can easily be followed to get the cdf values of the folded normal distribution for any other combination of μ and σ values. In Section 4, a comparison has been made between the existing and proposed methods of computing cdf values of a folded normal distribution. Three numerical examples are discussed in Section 5 describing some prospective applications of the theory developed in this article; while Section 6 discusses about some theoretical applications of the discussed theory and future scopes of study. The article concludes in Section 7 with a brief discussion on the theory developed in this article. Finally, two codes, written using *R* software, one for computing (μ, σ^2) from the given values of (μ_f, σ_f^2) and vice versa and the other for computing cdf values of folded normal distribution, for given values of (μ, σ^2), have been provided in the appendix of this article.

2. Relationships between the parameters of folded normal distribution and the parameters of the corresponding mother normal distribution

Having the values of μ and σ^2, one can easily compute the values of the corresponding μ_f and σ_f^2, using Equations (2) and (3) [1]. However, the situation is not that simple when one is required to find out the values of μ and σ^2 given the values of μ_f and σ_f. As has been discussed earlier, Leone et al.'s [1] table does not help either. In the present section, we discuss about a simple approach to deal with this problem.

From Equations (2) and (3),

$$\mu^2 + \sigma^2 = \mu_\mathrm{f}^2 + \sigma_\mathrm{f}^2. \tag{6}$$

Again, from Equation (2),

$$\mu_\mathrm{f} = \sigma\sqrt{\frac{2}{\pi}} \times \mathrm{e}^{-(\mu_\mathrm{f}^2+\sigma_\mathrm{f}^2-\sigma^2)/2\sigma^2} + \sqrt{\mu_\mathrm{f}^2+\sigma_\mathrm{f}^2-\sigma^2}\left[1 - 2\Phi\left(-\sqrt{\frac{\mu_\mathrm{f}^2+\sigma_\mathrm{f}^2-\sigma^2}{\sigma^2}}\right)\right]. \tag{7}$$

Then, the values of μ and σ^2 can be obtained by solving Equations (6) and (7). In this context, numerical methods such as the Newton–Raphson method or the method of iteration can be used to solve such complicated simultaneous equations with more than one unknown parameters.[12] We have presented, in the appendix, a simple code for computing (μ, σ^2) or $(\mu_\mathrm{f}, \sigma_\mathrm{f}^2)$ values, when the other is available. The code is written using *rootSolve* package of *R* software, whose underlying algorithm is based on the Newton–Raphson method.[13]

3. A simple formulation of the CDF of folded normal distribution

Suppose, Z and X are two random variables, such that $Z \sim N(\mu, \sigma^2)$ and $X = |Z|$. By definition, X follows folded normal distribution. The cdf of X can be obtained as

$$\begin{aligned}\Phi^{(\mathrm{FN})}(x) &= P(X < x)\\ &= \Phi\left(\frac{x-\mu}{\sigma}\right) + \Phi\left(\frac{x+\mu}{\sigma}\right) - 1 \quad \text{for } x \geq 0,\end{aligned} \tag{8}$$

where $\Phi^{(\mathrm{FN})}(.)$ denotes the cdf of folded normal distribution. Interestingly, the value of $\Phi^{(\mathrm{FN})}(x)$ is same for (μ, σ) and $(-\mu, \sigma)$, when $\mu > 0$, and it changes for any change in at least one of the μ and σ values.

This is due to the fact that, for both (μ, σ) and $(-\mu, \sigma)$, with $\mu > 0$, the values of μ_f and σ_f remain the same and hence, the value of $\Phi^{(\mathrm{FN})}(x)$ also remains the same.

When the mother normal distribution corresponding to a folded normal distribution is standard normal, that is, having $\mu = 0$ and $\sigma = 1$, then, the corresponding folded normal distribution will be *folded standard normal distribution* and its cdf can be obtained, from Equation (8), as

$$\Phi^{(\mathrm{FN})}(x) = 2\Phi(x) - 1 \quad \text{for } x \geq 0. \tag{9}$$

The values of $\Phi^{(\mathrm{FN})}(x)$, corresponding to $x = 0.5(0.5)5$, $\mu = 0(0.5)10.0$ and $\sigma = 1$, are given in Table 1.

Following observations can be made from Table 1.

(1) The values of $\Phi^{(\mathrm{FN})}(x)$ can be computed for any $x \geq 0$, as $X = |Z|$ and $Z \sim N(\mu, \sigma^2)$.
However, for $x = 0$, $\Phi^{(\mathrm{FN})}(x) = 0$. Hence, effectively, we need to consider $x > 0$.

Table 1. Cdf values ($\Phi^{(\mathrm{FN})}(x)$) of univariate folded normal distribution.

	x									
μ	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
0	0.38292	0.68269	0.86639	0.95450	0.98758	0.99730	0.99953	0.999937	0.99999	0.99999
0.5	0.34134	0.62465	0.81859	0.92698	0.9759	0.99356	0.99862	0.99976	0.99997	0.99999
1.0	0.24173	0.47725	0.68525	0.83999	0.93296	0.97722	0.99379	0.99865	0.99977	0.99997
1.5	0.13590	0.30233	0.49865	0.69123	0.84131	0.993319	0.97725	0.99379	0.99865	0.99767
2.0	0.06060	0.15730	0.30830	0.49997	0.69146	0.84134	0.93319	0.97725	0.99379	0.99865
2.5	0.02140	0.06657	0.15862	0.30853	0.5	0.69146	0.84135	0.93319	0.97725	0.99379
3.0	0.00598	0.02272	0.06681	0.15865	0.30854	0.5	0.69416	0.84134	0.93319	0.97725
3.5	0.00132	0.00621	0.02275	0.06681	0.15865	0.30854	0.5	0.69416	0.84134	0.93319
4.0	2.2923*E*4	0.00135	0.00621	0.02275	0.06681	0.15865	0.30854	0.5	0.69416	0.84134
4.5	3.1385*E*5	2.3261*E*4	0.00135	0.00621	0.02275	0.06681	0.15865	0.30854	0.5	0.69146
5.0	3.3787*E*6	3.1670*E*5	2.3263*E*4	0.00135	0.0621	0.02275	0.06681	0.15865	0.30854	0.5
5.5	2.8566*E*7	3.3976*E*6	3.1671*E*5	2.3263*E*4	0.00135	0.00621	0.02275	0.06681	0.15865	0.30854
6.0	0	2.8665*E*7	3.3977*E*6	3.1671*E*5	2.3263*E*4	0.00135	0.00621	0.02275	0.06681	0.15865
6.5	0	0	2.8665*E*7	3.3977*E*6	3.1671*E*5	2.3263*E*4	0.00135	0.00621	0.02275	0.06681
7.0	0	0	0	2.8665*E*7	3.3977*E*6	3.1671*E*5	2.3263*E*4	0.00135	0.00621	0.02275
7.5	0	0	0	0	2.8665*E*7	3.3977*E*6	3.1671*E*5	2.3263*E*4	0.00135	0.00621
8.0	0	0	0	0	0	2.8665*E*7	3.3977*E*6	3.1671*E*5	2.3263*E*4	0.00135
8.5	0	0	0	0	0	0	2.8665*E*7	3.3977*E*6	3.1671*E*5	2.3263*E*4
9.0	0	0	0	0	0	0	0	2.8665*E*7	3.3977*E*6	3.1671*E*5
9.5	0	0	0	0	0	0	0	0	2.8665*E*7	3.3977*E*6
10.0	0	0	0	0	0	0	0	0	0	2.8665*E*7

(2) In Table 1, the values of $\Phi^{(\mathrm{FN})}(x)$ are tabulated assuming $\sigma = 1$. For $\sigma \neq 1$, the $\Phi^{(\mathrm{FN})}(.)$ values can be obtained from the table, using the following steps:
  (i) Calculate $x^* = x/\sigma$ and $\mu^* = \mu/\sigma$.
  (ii) Find the value of $\Phi^{(\mathrm{FN})}(.)$ from the table, corresponding to (μ^*, x^*).

(3) Although, the table is given for $\mu = 0(0.5)10$ and $x = 0.5(0.5)5$, it can be extended for any μ and $x > 0$.

(4) From Equation (8), $\Phi^{(\mathrm{FN})}(x)|_{\mu>0} = \Phi^{(\mathrm{FN})}(x)|_{\mu<0}$, for $x > 0$. Hence, the table is also valid for $\mu = -10.0(0.5)0$.

(5) The first row (after the header), that is, the row for $\mu = 0$ gives the cdf values of folded *standard* normal distribution, that is, the folded normal distribution for which the mother normal distribution is $N(0, 1)$.

(6) Apart from this table, the $\Phi^{(\mathrm{FN})}(x)$ value can be computed, for any combination of μ, σ^2 and x, using Equation (8).

(7) If (μ, σ^2) is unknown, but (μ_f, σ_f^2) is known, one can obtain the values of μ and σ^2 following the approach discussed in Section 2 and then, $\Phi^{(\mathrm{FN})}(.)$ can be computed using Equation (8).

(8) The upper and lower α points of the folded normal distribution can also be obtained using Equation (8). Suppose, x' and x'' denote the upper and lower α points of a folded normal distribution with (μ, σ) being the parameters of the corresponding mother normal distribution. Then, from Equation (8), $\Phi^{(\mathrm{FN})}(x') = 1 - \alpha$ and $\Phi^{(\mathrm{FN})}(x'') = \alpha$.

(9) Changes in the distribution of folded normal distribution, as reflected by the corresponding cdf values, for changes in the values of σ are depicted in Figures 1–4. Here, μ, x and cdf values are plotted in the X, Y and Z axes, respectively, of Figures 1–4. From these figures, it is evident that, although in general folded normal distribution is positively skewed, its skewness decreases with the increase in σ value. Also, in Figure 5, cdf values of folded normal distribution are plotted against $x = 0.5(0.5)5$ for fixed $\sigma = 1$ and $\mu = 0, 1, 3, 4, 5$. From this figure, it can be observed that, for fixed value of σ, cdf values decrease with the increase in the values of μ.

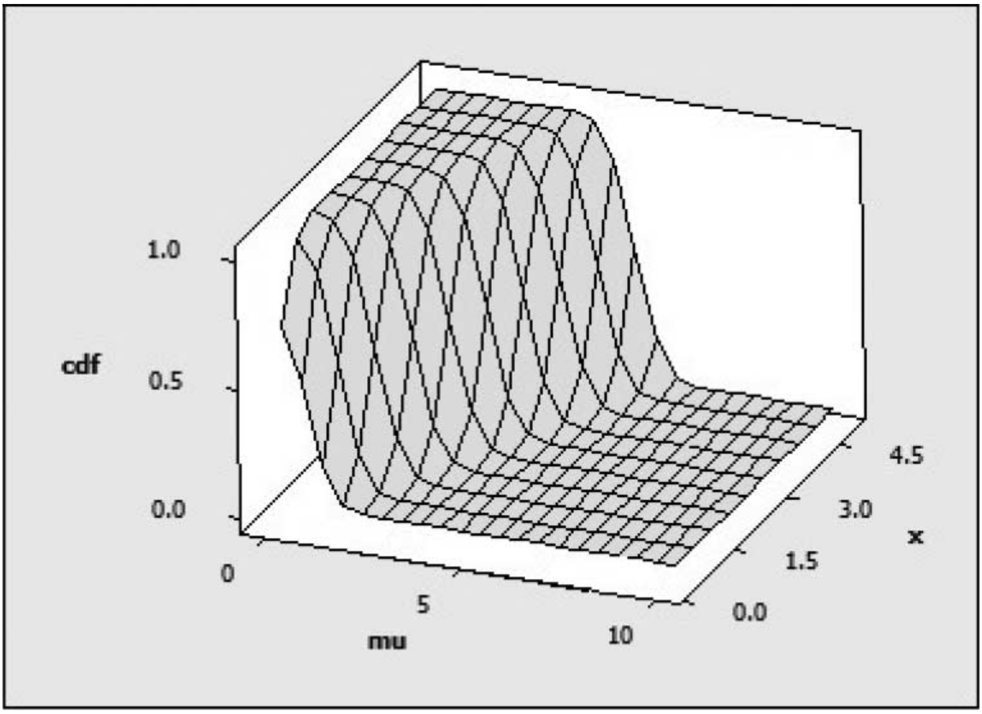

Figure 1. cdf of folded normal distribution with $x = 0.5(0.5)5$, $\mu = 0(0.5)10$ and $\sigma = 0.5$.

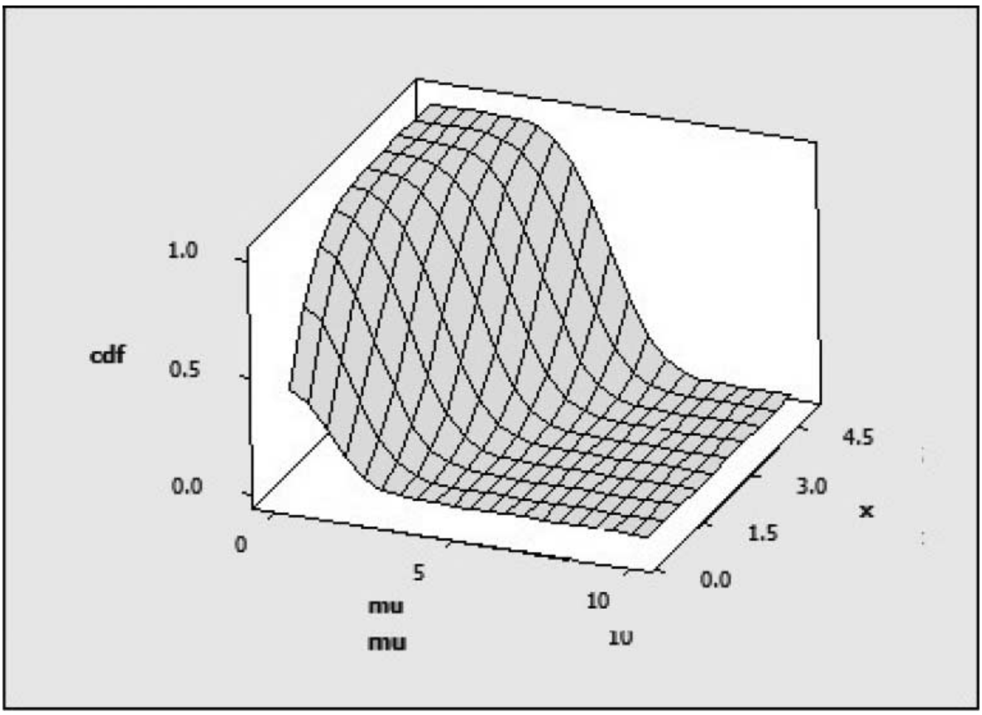

Figure 2. cdf of folded normal distribution with $x = 0.5(0.5)5$, $\mu = 0(0.5)10$ and $\sigma = 1$.

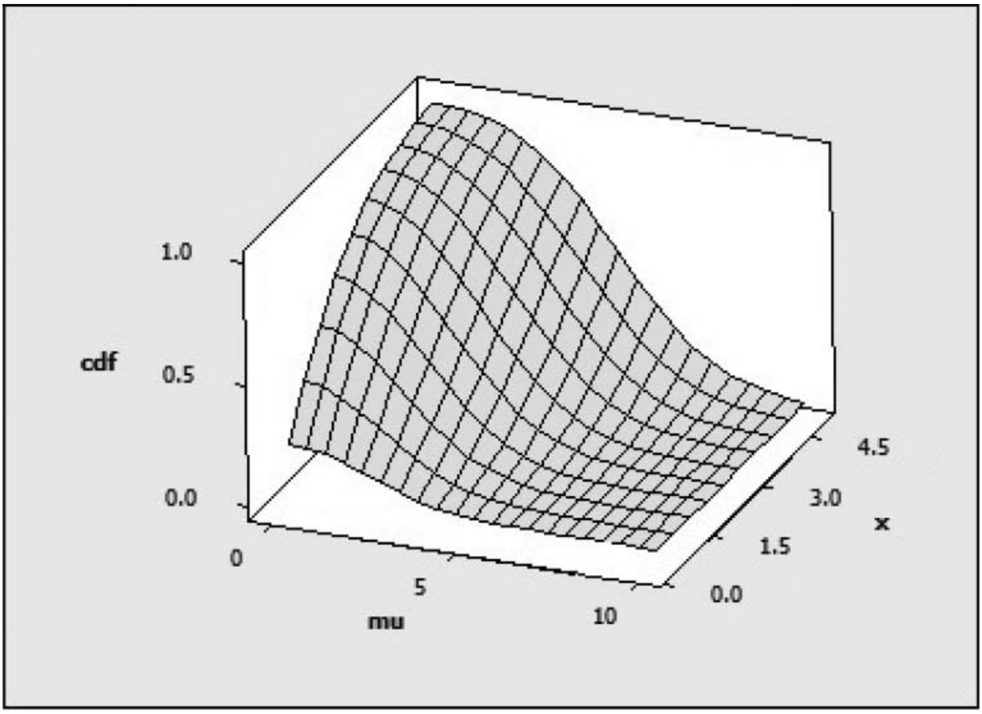

Figure 3. cdf of folded normal distribution with $x = 0.5(0.5)5$, $\mu = 0(0.5)10$ and $\sigma = 2$.

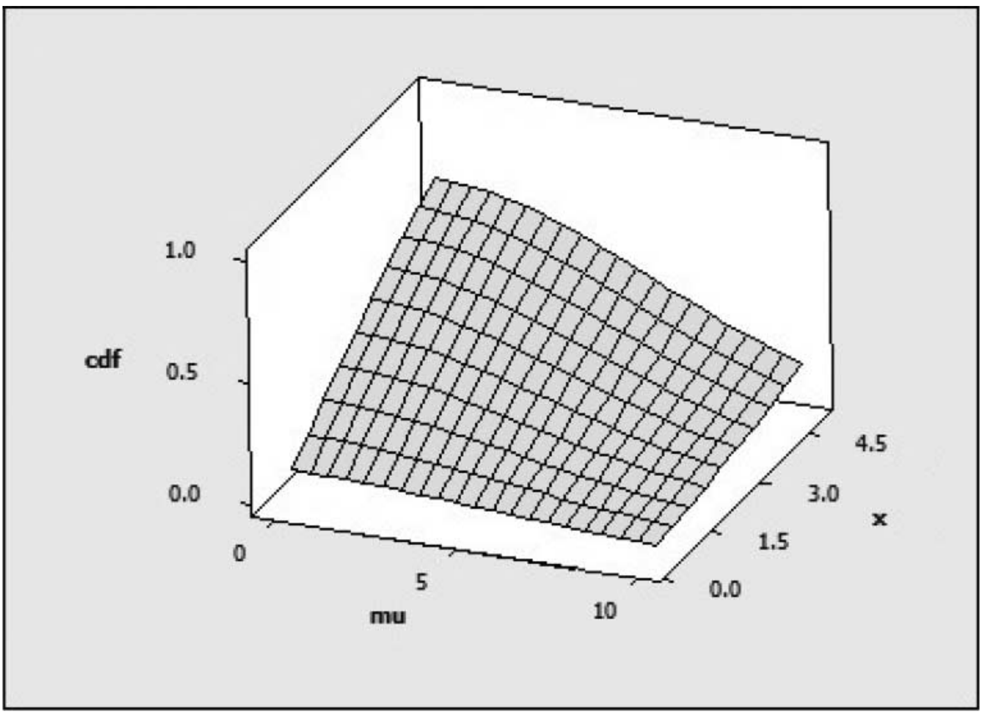

Figure 4. cdf of folded normal distribution with $x = 0.5(0.5)5$, $\mu = 0(0.5)10$ and $\sigma = 5$.

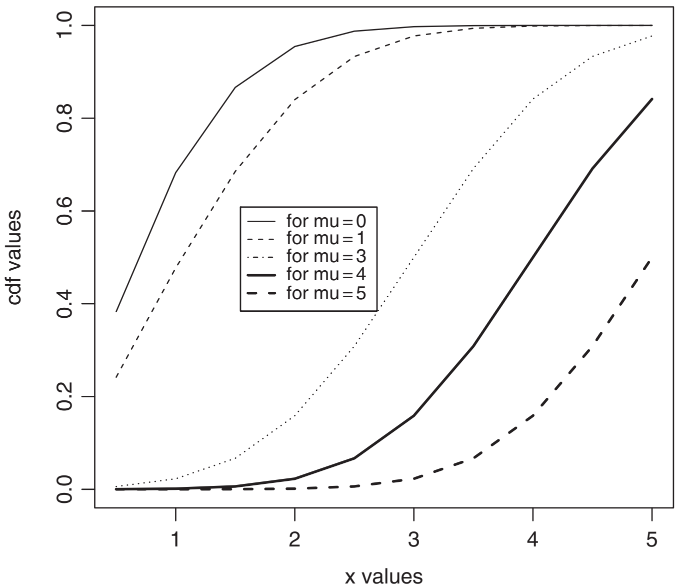

Figure 5. cdf of folded normal distribution with $x = 0.5(0.5)5$, $\mu = 0, 1, 3, 4, 5$ and $\sigma = 1$.

4. Comparison between Leone et al.'s [1] approach and the proposed method to compute cdf values of folded normal distribution

We are now in a position to compare Leone et al.'s [1] approach to that of the newly proposed one, in the context of computing cdf values of folded normal distribution.

(1) Leone et al. [1] defined the cdf of folded normal distribution in terms of μ_f/σ_f. However, the same value of μ_f/σ_f can be obtained for a number of individual values of μ_f and σ_f. Consequently, same cdf value can be obtained for a number of μ_f and σ_f values. In order to have a better insight about the problem, let us consider the following example.

From Leone et al.'s [1] table for cdf values of folded normal distribution and for $\mu_f/\sigma_f = 1.5$ and $x = 0.1$, the corresponding tabulated cdf value is 0.036. Now, let us consider the following situations:

Situation 1: Suppose $\mu_f = 3$ and $\sigma_f = 2$. Then, solving for Equations (6) and (7), $\mu = 2.664894$ and $\sigma = 2.42865$. Hence, from Equation (8) and for $x = 0.1$, $\Phi^{(FN)}(0.1) = 0.01799504 \neq 0.036$.

Situation 2: Suppose $\mu_f = 0.75$ and $\sigma_f = 0.5$. Then, $\mu = 0.66624$, $\sigma = 0.607142$ and for $x = 0.1$, $\Phi^{(FN)}(0.1) = 0.07203825 \neq 0.036$.

Situation 3: Suppose $\mu_f = 1.5$ and $\sigma_f = 1$. Then, $\mu = 1.332447$, $\sigma = 1.214325$ and for $x = 0.1$, $\Phi^{(FN)}(0.1) = 0.036$.

On the contrary, in the proposed approach, $\Phi^{(N)}(x)$ is defined as a function of μ and σ, rather than μ_f/σ_f and hence eliminates the chances of encountering similar situations.

(2) It can be easily observed that, only for $\sigma_f = 1$, Leone et al.'s [1] table of cdf values for folded normal distribution gives correct result. In other words, Leone et al.'s [1] table is valid for a very particular case of $\sigma_f = 1$. For $\sigma_f \neq 1$, it provides incorrect result.

In the proposed approach, no such assumption is required. In fact, as has been discussed in Section 3, although, Table 1 is constructed assuming $\sigma = 1$, its entries can be used to compute cdf values of folded normal distribution for any other value of $\sigma > 0$ as well.

(3) For particular values of x and μ_f/σ_f, Leone et al.'s [1] tabulated cdf values of folded normal distribution
 (i) underestimates the actual cdf value for $\sigma_f < 1$,
 (ii) overestimates the actual cdf value for $\sigma_f > 1$.
 Such problem of over or under estimation of cdf values of folded normal distribution is not encountered in the proposed approach, as it is based on individual parameter values, rather than their ratios.

(4) In the proposed formulation, μ_f and σ_f are uniquely expressed in terms of μ and σ. Hence, the cdf values change for even the slightest change in the values of at least one of μ or σ (and consequently, for changes in the values of μ_f or σ_f). However, this is not the case for Leone et al.'s [1] tabulated cdf values, where same cdf value may be attained for various combinations of values of μ_f and σ_f, provided the value of μ_f/σ_f is the same.

(5) By definition, cdf is necessarily an increasing function. However, Leone et al.'s [1] tabulated cdf values of folded normal distribution violates this important property of cdf. The authors themselves have pointed out that the cdf value of folded normal distribution for $\mu_f/\sigma_f = k$ is not necessarily greater than the cdf value of folded normal distribution for $\mu_f/\sigma_f = k'$, where $k' > k$. According to them, this happens due to the different rates of change in μ_f and σ_f values as the position of the fold changes.
 This problem does not exist for our proposed approach. From Table 1, for $\sigma = 1$, $\Phi^{(FN)}(x)$ decreases for increase in μ. Also, when $\sigma \neq 1$, following the discussion in Section 3, the cdf values can be obtained by proper scaling of x and μ. Since, $\sigma > 0$, similar relations will be retained in such cases as well. Therefore, the proposed approach is statistically more consistent than the existing one.

5. Numerical examples

We shall now apply the newly developed theory to three numerical data sets, consisting of both the simulated and real-life data, and observe their performances.

Example 1 (*Based on simulated data*) We have simulated a sample of size 20,000, from normal distribution with mean $\mu = 15$ and standard deviation (sd) $\sigma = 5$. From the simulated data, the estimated values of the mean and sd are found to be $\hat{\mu} = 10.21586$ and $\hat{\sigma} = 4.6659$. Also, from Equations (2) and (6), $\hat{\mu}_f = 10.2628$ and $\hat{\sigma}_f = 4.5615$ and hence, $\hat{\mu}_f/\hat{\sigma}_f = 2.2499 \simeq 2.2$. Using Equation (8), for $x = 0.1$, the cdf value of the said folded normal distribution is found to be 0.0016, while for $x = 2.1$, the cdf value is 0.0368.

However, according to Leone et al.,[1] the cdf value should be 0.0081 and 0.4659, respectively, for $x = 0.1$ and 2.1 and $\hat{\mu}_f/\hat{\sigma}_f \simeq 2.2$. Clearly, here for both the values of x, Leone et al.'s [1] approach overestimates the true cdf value of the folded normal distribution and following the discussion in Section 4, such overestimation has occurred as here $\hat{\sigma}_f > 1$.

Example 2 (*Based on Leone et al.'s [1] data*) Let us now consider the numerical example given by Leone et al. [1] to compare the performance of our newly proposed approach of computing (μ, σ) from (μ_f, σ_f) and computing the cdf value of folded normal distribution. These data pertain to the manufacturing of miniature radio tubes. Straightness of lead wires (expressed in terms of chamber) is the concerned quality characteristic. From the data provided by Leone et al.,[1] the observed values of μ_f and σ_f are 14.014 and 7.7868, respectively. Thus, $\mu_f/\sigma_f = 1.799774 \simeq 1.80$. Following Leone et al.,[1] $\mu = 13.6424$ and $\sigma = 8.4254$. On the other hand, following our proposed approach and solving Equations (6) and (7) with $\mu_f = 14.014$ and $\sigma_f = 7.7868$, we have $\mu = 13.6395$ and $\sigma = 8.4255$. Moreover, the proposed approach is much simpler compared to that of Leone et al.[1]

Again, for $x = 0.1$, Leone et al.'s [1] formulation gives the cdf value as 0.0199; while using Equation (8), the actual cdf value is computed as $\Phi^{(FN)}(0.1) = 0.0025 < 0.0199$. Note that, since here $\sigma_f = 7.7868 > 1$, the cdf value tabulated by Leone et al.,[1] overestimates the actual cdf value.

Example 3 (*Based on body mass index data*) Our next example is based on body mass index (BMI) data from the 'Fletcher Challenge/ Auckland Heart and Health survey' [14] for which the data are available in R package VGAM.[15] Here we have 700 observations on two variables, namely age and BMI.

Tsagris et al. [16] have shown that the data on BMI follow folded normal distribution. Thus, from the said data, we have $\hat{\mu}_f = 26.6847$ and $\hat{\sigma}_f = 4.6213$. Therefore, following our proposed approach and solving Equations (6) and (7), we have $\hat{\mu} = 26.6847$ and $\hat{\sigma} = 4.6213$.

Interestingly, here $\hat{\mu}_f = \hat{\mu}$ and $\hat{\sigma}_f = \hat{\sigma}$. This validates the observation made by Tsagris et al. [16] that, for the said BMI data, the fitted folded normal converges in distribution to normal distribution.

Finally, using Equation (8), for $x = 0.1$, the cdf value of the said folded normal distribution is found to be 9.9553×10^{-10}, while for $x = 16$, the cdf value is 0.0104.

6. Some applications and future scopes of study

6.1. *Application in process capability analysis using C_{pk}*

PCI, as a measure of statistically assessing the ability of a process to produce items within the pre-assigned specification limits, is grabbing the attention of more and more industrial statisticians as well as shop-floor personnel day by day. Among the PCIs existing in the literature, C_{pk} is the most widely accepted among practitioners and hence the study of the distribution and inferential properties of C_{pk} and its plug-in estimator are of particular interest for several eminent statisticians (refer [9] and the references there-in). Pearn et al. [17] have extensively studied the estimation procedure of C_{pk} from both the classical and Bayesian perspective.

Notationally, under the assumption of normality of the underlying process distribution, C_{pk} can be defined as $C_{pk} = (d - |\mu - M|)/3\sigma$, where $d = (U - L)/2$, $M = (U + L)/2$ and U and L stand for the upper and lower specification limits of the concerned quality characteristic. Thus, the corresponding plug-in estimator, namely $\hat{C}_{pk}$ will be defined as $\hat{C}_{pk} = (d - |\bar{X} - M|)/3S$, where $\bar{X}$ and S stand for sample mean and sd based on available data.

It is easy to observe that the statistical distribution of $\hat{C}_{pk}$ involves folded normal distribution [9] and [17]. Hence, the methodology for obtaining (μ, σ) from (μ_f, σ_f) along with the cdf values of folded normal distribution can be used in hypothesis testing and in obtaining confidence interval (in particular, to obtain the so-called upper and lower α points for certain values of α, say $\alpha = 0.01, 0.05$ and so on) for $\hat{C}_{pk}$ values. These upper and lower α points of folded normal distribution can also be used in designing process capability control chart for C_{pk}.[18]

6.2. *Application in estimating coefficient of variation for folded normal distribution*

Coefficient of variation ($\tau = \sigma/\mu$) of a distribution is one of the most important statistical measures. Mahamoudvand and Hassani [19] have proposed an asymptotically unbiased estimator of τ and constructed the associated confidence interval. Similar study in the context of folded normal distribution would be another interesting topic to study.

In this context, following Mahamoudvand and Hassani's [19] approach, the natural (plug-in) estimator of τ, say $\hat{\tau}$ may be defined, in terms of the plug-in estimators of μ_{f} and σ_{f}, say $\hat{\mu}_{\mathrm{f}}$ and $\hat{\sigma}_{\mathrm{f}}$, respectively, as

$$\begin{aligned}\hat{\tau} &= \frac{\hat{\sigma}_{\mathrm{f}}}{\hat{\mu}_{\mathrm{f}}}\\ &= \hat{\sigma}_{\mathrm{f}} \times \sum_{k=1}^{\infty} \frac{(-1)^{k-1} \times (\hat{\mu}_{\mathrm{f}} - \mu_{\mathrm{f}})^{k-1}}{\mu_{\mathrm{f}}^{k}},\end{aligned} \tag{10}$$

applying the Taylor series expansion of $\hat{\mu}_{\mathrm{f}}$ at $\hat{\mu}_{\mathrm{f}} = \mu_{\mathrm{f}}$.

However, unlike the normal distribution, for which $\hat{\mu}$ and $\hat{\sigma}$ are independently distributed, for folded normal distribution, such conclusion cannot be drawn for $\hat{\mu}_{\mathrm{f}}$ and $\hat{\sigma}_{\mathrm{f}}$, since both of them can be expressed as functions of μ and σ. Hence, derivation of $E(\hat{\tau})$ remains as an interesting open problem for future study.

In this context, following Elandt,[2] the rth raw moment of folded normal distribution can be defined as

$$\mu_{\mathrm{f}}^{\prime(r)} = \sigma^{r} \sum_{j=0}^{r} \binom{r}{j} \theta^{r-j}[I_j(-\theta) + (-1)^{r-j} I_j(\theta)], \tag{11}$$

where $\theta = \mu/\sigma$ and $I_r(a) = (1/\sqrt{2\pi})a^{r-1}\,\mathrm{e}^{-a^2/2} + (r-1)I_{r-2}(a)$, for $r > 0$.

Hence, using the proposed approach to obtain (μ, σ) values from $(\mu_{\mathrm{f}}, \sigma_{\mathrm{f}})$ [vide Equations (6) and (7) and Appendix 1] in Equations (10) and (11) and applying the Newton–Raphson method, $E(\hat{\tau})$ and the corresponding confidence bound can be obtained numerically.

7. Conclusion

In the present article, we have presented a simple approach for finding the mean (μ) and variance (σ^2) of mother normal distribution, when the mean (μ_{f}) and variance (σ_{f}^2) of the corresponding folded normal distribution is known and vice versa. We have also simplified the expression for the cdf of folded normal distribution by expressing it in terms of a linear combination of two standard normal cdfs. Such cdf values and in particular the so-called upper and lower α points of a folded normal distribution, which are nothing but some cdf values calculated at some specific values of x find ample application in testing of hypothesis problems related to folded normal distribution and in designing control chart for PCI such as C_{pk} among others. Necessary R-codes have also been provided in the appendix for future applications.

We have also shown that Leone et al.'s [1] table for cdf values of folded normal distribution is valid for a very particular case of $\sigma_{\mathrm{f}} = 1$; while for $\sigma_{\mathrm{f}} \neq 1$, it over/under estimates the actual cdf value, depending upon the value of σ_{f}. Finally, we have made an extensive comparison of the proposed method to that of the existing one and have observed that, the proposed approach of computing cdf values of a folded normal distribution is better than that of Leone et al.,[1] as here the cdf value changes for the slightest change in at least one of μ and σ (or, μ_{f} and σ_{f}), which is highly desirable. The numerical example provided by Leone et al. [1] has been revisited along with two other numerical examples in the light of the theory developed in this article and it has been found that the proposed approach is easier to compute and is more accurate than the existing one.

Acknowledgements

The authors are extremely thankful to the learned reviewer(s) and the associate editor for their valuable comments and suggestions in improving the quality of the paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

[1] Leone FC, Nelson LS, Nottingham RB. The folded normal distribution. Technometrics. 1961;3(4):543–550.
[2] Elandt RC. The folded normal distribution: two methods of estimating parameters from moments. Technometrics. 1961;3(4):551–562.
[3] Psarakis S, Panaretos J. On some bivariate extensions of the folded normal and the folded t distributions. J Appl Statist Sci. 2001;10(2):119–136.
[4] Chakraborty AK, Chatterjee M. On multivariate folded normal distribution. Sankhya, Series B. 2013;75(1):1–15.
[5] Lin HC. The measurement of a process capability for folded normal process data. Int J Adv Manuf Technol. 2004;24:223–228.
[6] Lin PC. Application of the generalized folded-normal distribution to the process capability measures. Int J Adv Manuf Technol. 2005;26:825–830.
[7] Johnson NL. Cumulative sum control charts for the folded normal distribution. Technometrics. 1963;5:451–458.
[8] Liao MY. Economic tolerance design for folded normal data. Int J Prod Res. 2010;18(14):4123–4137.
[9] Pearn WL, Kotz S. Encyclopedia and handbook of process capability indices: series on quality, reliability and engineering statistics. World Scientific: Singapore; 2007.
[10] Chatterjee M, Chakraborty AK. Distributions and process capability control charts for C_{pu} and C_{pl} using subgroup information. Comm Statist Theory Methods. 2013; Accepted for Publication.
[11] Sundberg R. On estimation and testing for the folded normal distribution. Comm Statist Theory Methods. 1974;3(1):55–72.
[12] Scarborough JB. Numerical mathematical analysis. New Delhi: Oxford and IBH Publishing; 1955.
[13] Soetaert K. rootSolve: nonlinear root finding, equilibrium and steady-state analysis of ordinary differential equations. R Package Version 1.4; 2009.
[14] MacMohan S, Norton R, Jackson R, Mackie MJ, Cheng A, Vander Hoorn S, Milne A, McCulloch A. Fletcher challenge – University of Auckland heart and health study: design and baseline findings. New Zeal Med J. 1995;108:499–502.
[15] Yee TW. The VGAM package for categorical data analysis. J Statist Softw. 2010;32:1–34.
[16] Tsagris M, Beneki C, Hassani H. On the folded normal distribution. Mathematics. 2014;2:12–28.
[17] Pearn WL, Wu CC, Wu CH. Estimating process capability index C_{pk}: classical approach versus Bayesian approach. J Statist Comput Simul. 2014; doi:10.1080/00949655.2014.914211.
[18] Chatterjee M. Exact expression for the expectation of estimated C_{pk} based on control chart information and the corresponding process capability control charts. 2014:17.
[19] Mahamoudvand R, Hassani H. Two new confidence intervals for the coefficient of variation in a normal distribution. J Appl Stat. 2009;36(4):429–442.

Appendix 1. R-Code for computing (μ, σ^2) from the corresponding (μ_f, σ_f^2) and vice versa

```
 library(rootSolve)

######################################################################

# FINDING  mu_f AND sigma_f FROM mu AND sigma

mu_original<-3
mu_original

var_original<-2
var_original

mu_folded_1<-(sqrt(var_original))*(sqrt(2/pi))*(exp(-((mu_original)^2)/
  (2*var_original)))
mu_folded_1

mu_folded_2<-mu_original*(1-2*pnorm(-mu_original/sqrt(var_original)))
mu_folded_2

mu_folded<-mu_folded_1+mu_folded_2
mu_folded
```

```
var_folded<-mu_original^2+var_original-mu_folded^2
var_folded

##########################################################################

## FINDING mu AND sigma FROM mu_f AND sigma_f

x<-mu_folded    # VALUE OF mu_f
x

y_2<-var_folded  # VALUE OF var_f
y_2

eq_2<-x^2+y_2
eq_2

fun<-function(s) sqrt(s)*(sqrt(2/pi))*exp(-((eq_2-s)/(2*s)))
+(sqrt(eq_2-s))*(1-2*pnorm(-(sqrt((eq_2-s)/s))))-x

All <- uniroot.all(fun, c(0, eq_2))
All

sigma_2<-All  # VARIANCE OF MOTHER NORMAL
sigma_2

mu_2<-eq_2-All
mu_2

mu<-sqrt(mu_2)   # MEAN OF MOTHER NORMAL
mu
```

Appendix 2. R-Code for computing cdf of folded normal distribution

```
x<-seq(.5,5,length=10)   # SEQUENCE OF x- VALUES
x

mu<-seq(0,10,length=21)  # SEQUENCE OF mu- VALUES
mu

cdf_folded<-array(0,dim=c(10,21))  # Initializing cdf values
cdf_folded

f<-function(a,b){
z<-pnorm((a-b)/s)+pnorm((a+b)/s)-1
z
}

s<-1    # VALUE OF SIGMA

for(i in 1 : 10)
 for(j in 1:21)
    cdf_folded[i,j]<-f(x[i],mu[j])

cdf_folded   # COMPUTED cdf VALUES OF FOLDED NORMAL DISTRIBUTION

cdf_folded_trans<-as.vector(t(cdf_folded))
cdf_folded_trans

x_rep<-rep(x, each = 21)
x_rep

mu_rep<-rep(mu, 10)
mu_rep

folded_table<-cbind(x_rep,mu_rep,cdf_folded_trans)
folded_table    # TABULATED VALUES OF mu AND THE cdf FOR THE CHOSEN VALUE OF 'x'
```