Sinopfbd, 2022; 7(2): 116-131 https://doi.org/10.33484/sinopfbd.1164454

Sinop Üniversitesi Fen Bilimleri Dergisi
Sinop Uni J Nat Sci

ISSN: 2536-4383
E-ISSN: 2564-4383

https://dergipark.org.tr/tr/pub/sinopfbd

The Proposal of Gamma Folded-Normal Distribution

Yasin ALTINIŞIK[1], Tahir CEYLAN[2] and Emel ÇANKAYA[1]

How to cite: Altınışık, Y., Ceylan, T., & Çankaya, E. (2022). The proposal of gamma folded-normal distribution. *Sinop Üniversitesi Fen Bilimleri Dergisi, 7*(2), 116-131. https://doi.org/10.33484/sinopfbd.1164454

Research Article

Corresponding Author
Yasin ALTINISIK
yaltinisik@sinop.edu.tr

ORCID of the Authors
Y.A: 0000-0001-9375-2276
T.C: 0000-0002-3187-2800
E.Ç: 0000-0002-2892-2520

Received: 19.08.2022
Accepted: 11.10.2022

Abstract

Development of new flexible distributions for modeling non-negative measurements occuring in lifetime or reliability studies is a prominent research area in Statistics. As being the most favoured positive definite form, the Gamma distribution poses a basis for such improvements. It is well known that transforming a Gamma variable with another continuous random variable (X) creates Gamma-X family of distributions. Following this principle, we here attempted to define the X variable as Folded-Normal distributed which is also positive definite so as to propose a new family of distributions. Named as Gamma Folded-Normal distribution (GFN), our proposal is a generalization of Gamma Half-Normal distribution and contains more freely estimated parameters. This study evaluates some mathematical properties of GFN distribution such as moments and illustrates the estimation procedure for unknown parameters through a simulation study. A separate simulation is also conducted to compare the performance of this new distribution with the Folded-Normal, Half-Normal and Gamma Half-Normal distributions. Besides, the practical importance of our new proposal is illustrated by analyzing a real world data set.

Keywords: Gamma-X Family, Gamma Half-Normal, Folded-Normal, Lifetime Data, Non-negative Modeling

Gamma Katlanmış-Normal Dağılım Önerisi

[1]Sinop University, Faculty of Arts and Sciences, Statistics Department, Sinop, Türkiye

[2]Sinop University, Faculty of Arts and Sciences, Mathematics Department, Sinop, Türkiye

Öz

Yaşam ve güvenilirlik çalışmalarında ortaya çıkan negatif olmayan ölçümleri modellemek için yeni esnek dağılımların geliştirilmesi, İstatistikte dikkat çeken bir araştırma alanıdır. Pozitif tanımlı ve en çok tercih edilen bir biçim olan Gamma dağılımı, bu tür geliştirmeler için bir temel oluşturur. Bir Gamma değişkenini başka bir sürekli rastgele değişken (X) ile dönüştürmenin, Gamma-X dağılım ailesini oluşturduğu iyi bilinmektedir. Bu prensibi takip ederek, X değişkenini de yine pozitif tanımlı olan Katlanmış-Normal dağılımlı tanımlayarak, burada yeni bir dağılım ailesi önermeye çalıştık. Gamma Katlanmış-Normal dağılım (GFN) olarak adlandırılan önerimiz, Gamma Yarı-Normal dağılımının bir genellemesidir ve daha fazla sayıda serbest tahminli parametreler içerir. Bu çalışma, GFN dağılımının momentler gibi bazı matematiksel özelliklerini geliştirmekte ve bir simülasyon çalışması aracılığıyla bilinmeyen parametreler için tahmin prosedürünü göstermektedir. Bu yeni dağılımın performansını, Katlanmış-Normal, Yarı-Normal ve Gamma Yarı-Normal dağılımları ile karşılaştırmak için ise ayrı bir simülasyon çalışması da gerçekleştirilmiştir. Ayrıca, bir gerçek hayat veri setinin analizi yapılarak yeni önerimizin uygulamada önemi gösterilmiştir.

Anahtar Kelimeler: Gamma-X Ailesi, Gamma Yarı-Normal, Katlanmış-Normal, Yaşam Verisi, Negatif Olmayan Modelleme

Introduction

Attempts to define new Gamma related probability distributions serve for non-negative modeling particularly for lifetime processes. The need for such attempts stems from the fact that many well-known distributions fail to provide precise evaluation of real world phenomena. In this respect, Alzaatreh, Famoye, and Lee [1] suggested a new method to generate a new class of distributions. It is based on transforming a random variable (T) by another random variable (X), resulting in T-X family of distributions. If T is chosen as a Gamma distributed random variable then Gamma-X family of distributions are produced [1]. Particularly suitable to the non-negative nature of lifetime processes, X is further chosen as a positive definite continuous distribution. In the literature, Half-Normal (HN) distribution appears one of the most preferable transformer (X) for this purpose, which creates the Gamma Half-Normal (GHN) distribution [2]. This new form is stated as more flexible due to the additional shape (α) and scale (β) parameters of Gamma to the only parameter of HN (σ) that characterises the shape [3]. The Folded-Normal distribution (FN) is also suitable for modeling positive values and known to be a special case of the Gaussian distribution [4, 5]. It resembles the HN distribution where the location parameter μ is not necessarily zero. FN distribution can be considered as a good candidate for the derivation of a new flexible distribution due to its additional location parameter. The main objective of this study is therefore to offer a new combined form of Gamma using the Folded-Normal distribution with the claim of that the resulting distribution would provide a more accurate assessment of the data. The proposal of the Gamma Folded-Normal (GFN) distribution would be a member of Gamma-X family of distributions and could be considered as a generalization of the GHN distribution as it turns out to be the GHN distribution when its location parameter μ is equal to zero. The aim of this study is to provide various properties of this new distribution including maximum likelihood estimation of its four parameters. Two simulation studies are conducted to investigate the distributional properties of GFN in comparison to the HN, FN, and GHN distributions. A real data example is also provided to support the conclusions drawn from the simulation experiments. The outline of the paper is as follows. Next section gives a general definition of Gamma-X family and Folded-Normal distributions. This is followed by the generation of the GFN distribution based on this definition as well as the derivation of the maximum likelihood estimates of the parameters and moments for this new form. As a crucial step of parameter estimation process, the choice of initials for the model parameters is introduced as a requirement of an optimization process in a separate section. Simulation section provides a performance evaluation for each distribution with respect to bias, precision, and accuracy measures as well as the log likelihood values. Following section discusses the applicability of the GFN distribution through a real world data analysis. The paper is then finalised with a brief discussion.

The Folded-Normal and Gamma-X Family of Distributions

Let X be a random variable that has the FN distribution with a location parameter μ and a scale parameter σ. The probability density function (PDF) of $X \sim \text{FN}(\mu,\sigma)$ can be defined as

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

where $\mu \geq 0$ and $\sigma > 0$. Note that this functional form reduces to the PDF of the HN distribution when the location parameter μ is equal to zero. Taking the integral of (1) with respect to X gives the cumulative density function (CDF) of the FN distribution:

$$F(x;\mu,\sigma) = \Phi(\frac{x-\mu}{\sigma}) + \Phi(\frac{x+\mu}{\sigma}) - 1 \tag{2}$$

where $\Phi(.)$ is the CDF of the standard normal distribution. Then, (1) and (2) above can be incorporated into Gamma-X family of distributions to produce our new proposal which would contain four parameters (i.e., α and β belonging to the gamma distribution and μ and σ to the FN distribution).

The Gamma-X family of distributions is a member of a more general the T-X family of distributions described by [1]. The PDF and the CDF of the Gamma-X family of distributions are the main tools for characterizing the new proposal of the current paper. In general definition, the PDF of the Gamma-X family of distributions is given by

$$h(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} f(x)(-\log(1-F(x)))^{\alpha-1}(1-F(x))^{\frac{1}{\beta}-1} \tag{3}$$

where α and β are the parameters of the gamma distribution and $f(x)$ and $F(x)$ are the PDF and CDF of the continuous target distribution (e.g., the FN distribution with parameters μ and σ in our case), respectively. Similarly, the CDF of the Gamma-X family of distributions is defined by

$$H(x) = \int_0^{-\log(1-F(x))} r(t)dt \tag{4}$$

where $r(t) = \frac{1}{\beta^\alpha\Gamma(\alpha)} t^{\alpha-1}e^{-\frac{t}{\beta}}$, $t \geq 0$ is the PDF of the Gamma distribution.

Following the principles of the Gamma-X family of distributions introduced here, the next section gives the derivation of our new proposal, namely Gamma Folded-Normal (GFN), as well as the elaboration of the parameter estimation and the evaluation of the moments for this new form.

The GFN Distribution: Definition and Parameter Estimation

Let X be a positively defined random variable from the GFN distribution ($X \sim GFN(\alpha,\beta,\mu,\sigma)$) with four freely estimated parameters, where $\mu \geq 0$ and α, β, and $\sigma > 0$. The PDF of variable X can be obtained by incorporating the PDF and CDF of the FN distribution in (1) and (2) into the PDF of the

Gamma-X family of distributions in (3). The resulting PDF of X is here named as the GFN distribution and can be defined as

$$h(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)}\frac{1}{\sqrt{2\pi\sigma^2}}[e^{-\frac{1}{2\sigma^2}(x-\mu)^2} + e^{-\frac{1}{2\sigma^2}(x+\mu)^2}]$$
$$\times (-\log(2 - \Phi(\tfrac{x-\mu}{\sigma}) - \Phi(\tfrac{x+\mu}{\sigma})))^{\alpha-1} \tag{5}$$
$$\times (2 - \Phi(\frac{x-\mu}{\sigma}) - \Phi(\frac{x+\mu}{\sigma}))^{\frac{1}{\beta}-1}$$

The Gamma Folded-Normal distribution containing four freely estimated parameters is the end point of a sequence of generalizations which started with the Half-Normal distribution. The $h(x)$ function above reduces to the PDF of the GHN distribution for $\mu = 0$; the FN distribution for $\alpha = \beta = 1$; and the HN distribution for $\alpha = \beta = 1$ and $\mu = 0$. Figure 1 displays the distrbutional forms of the GFN for varying values of α, β, μ, and σ parameters.

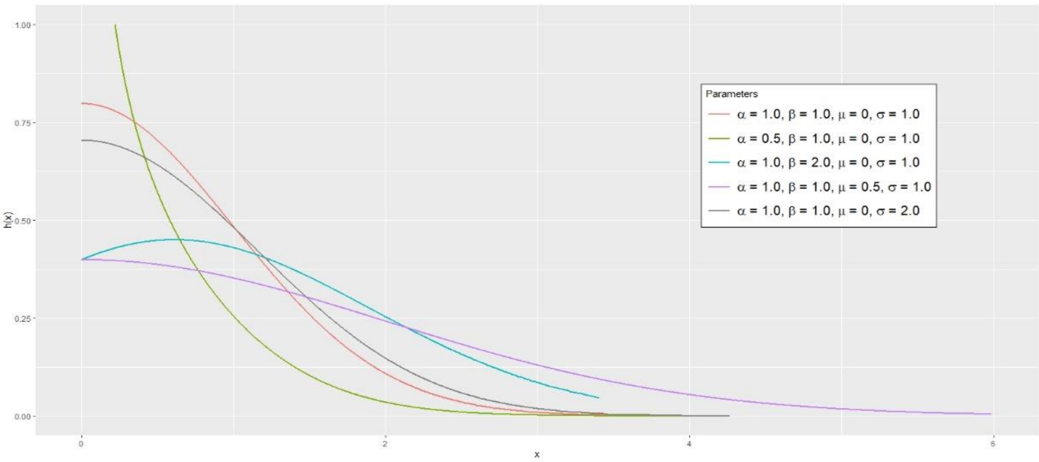

Figure 1. *Illustrations of the GFN distribution with varying values of parameters.*

The CDF of Gamma-X family of distributions in (4) with respect to the FN continuous target distribution is defined as

$$H(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)}\int_0^{-\log(2-\Phi(\frac{x-\mu}{\sigma})-\Phi(\frac{x+\mu}{\sigma}))} t^{\alpha-1}e^{-\frac{t}{\beta}}dt \tag{6}$$

By applying $u = \frac{t}{\beta}$ transformation, the $H(x)$ function above becomes

$$H(x) = \gamma(\alpha, -\frac{1}{\beta}\log(2 - \Phi(\frac{x-\mu}{\sigma}) - \Phi(\frac{x+\mu}{\sigma}))/\Gamma(\alpha) \tag{7}$$

where $\gamma(.)$ is the lower incomplete gamma function with $\gamma(\alpha, A) = \int_0^A u^{\alpha-1}e^{-u}du.$

Let $x_1, x_2, \ldots, x_n$ be a sample of size n from the GFN distribution. To facilitate the log likelihood function based on this sample, the $h(x)$ function is analogously expressed by

$$h(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)}\frac{1}{\sqrt{2\pi\sigma^2}}[e^{-\frac{(x-\mu)^2}{2\sigma^2}}(1 + e^{-\frac{(x+\mu)^2}{2\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}})]$$

ISSN: 2536-4383

$$\times (-\log(2-\Phi(\tfrac{x-\mu}{\sigma})-\Phi(\tfrac{x+\mu}{\sigma})))^{\alpha-1} \tag{8}$$
$$\times (2-\Phi(\frac{x-\mu}{\sigma})-\Phi(\frac{x+\mu}{\sigma}))^{\frac{1}{\beta}-1}$$

where $\Phi(.)$ is the CDF of the standard normal distribution.

Then, the log likelihood function for the Gamma Folded-Normal distribution is given by

$$\log L(\alpha,\beta,\mu,\sigma) = -\alpha n\log(\beta) - n\log(\Gamma(\alpha)) - \frac{n}{2}\log(2\pi\sigma^2)$$
$$-\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2} + \sum_{i=1}^{n}\log(1+e^{-\frac{2\mu x_i}{\sigma^2}})$$
$$+(\alpha-1)\textstyle\sum_{i=1}^{n}\log(-\log(2-\Phi(\frac{x_i-\mu}{\sigma})-\Phi(\frac{x_i+\mu}{\sigma}))) \tag{9}$$
$$+(\frac{1}{\beta}-1)\sum_{i=1}^{n}\log(2-\Phi(\frac{x_i-\mu}{\sigma})-\Phi(\frac{x_i+\mu}{\sigma}))$$

The log likelihood function above can be maximized by the estimates of α, β, μ, and σ parameters. This can be ensured by taking the derivative of the log likelihood function with respect to each parameter and setting the resulting expression equal to zero (see Appendix A). However, these expressions do not have closed form solutions, since they involve the functions of the CDF of a Normal distribution and/or the derivative of the Gamma function. Thus, a numerical procedure is utilized to estimate model parameters. The Rv3.2.4 package *maxLik* with the subroutine Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimisation method is used for maximum likelihood estimation. Implementation of the BFGS procedure in R statistical software [6] requires a set of properly chosen initial values for the parameters, which will be elaborated after defining the moments for the GFN distribution in the following section.

Moments

Lemma 1 If $u = -log[2-\Phi(\frac{x-\mu}{\sigma})-\Phi(\frac{x+\mu}{\sigma})]$, the derivative with respect to u can be defined as

$$du = \frac{\frac{1}{\sqrt{2\pi}\sigma}[e^{-\frac{(x-\mu)^2}{2\sigma^2}}+e^{-\frac{(x+\mu)^2}{2\sigma^2}}]}{\Phi(-\frac{x-\mu}{\sigma})+\Phi(-\frac{x+\mu}{\sigma})}$$

Proof $u = -\log[2-\Phi(\frac{x-\mu}{\sigma})-\Phi(\frac{x+\mu}{\sigma})]$

$$= -\log[1-\Phi(\tfrac{x-\mu}{\sigma})+1-\Phi(\tfrac{x+\mu}{\sigma})]$$
$$= -\log[\Phi(-\tfrac{x-\mu}{\sigma})+\Phi(-\tfrac{x+\mu}{\sigma})]$$

where $\Phi(.)$ is the CDF of the standard normal distribution. Thus,

$$du = \frac{\frac{1}{\sigma}[\Phi'(-\frac{x-\mu}{\sigma})+\Phi'(-\frac{x+\mu}{\sigma})]}{\Phi(-\frac{x-\mu}{\sigma})+\Phi(-\frac{x+\mu}{\sigma})]}$$

where $\Phi'(-\frac{x-\mu}{\sigma}) = \frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

and $\Phi'(-\frac{x+\mu}{\sigma}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x+\mu)^2}{2\sigma^2}}$.

Lemma 2 If $u = -log[2 - \Phi(\frac{x-\mu}{\sigma}) - \Phi(\frac{x+\mu}{\sigma})]$, variable X following GFN distribution can be defined with respect to u as

$$x = (-0.5)\sigma\Phi^{-1}(e^{-u})$$

Proof. By following Lemma 1,

$$\Phi(-\frac{x-\mu}{\sigma}) + \Phi(-\frac{x+\mu}{\sigma}) = e^{-u}.$$

If we take the inverse CDF of both sides of the equation above

$$-[\frac{x-\mu}{\sigma} + \frac{x+\mu}{\sigma}] = \Phi^{-1}(e^{-u})$$

and then

$$x = (-0.5)\sigma\Phi^{-1}(e^{-u}).$$

The rth moments of the GFN distribution using the PDF of variable X in (5) can be defined as

$$E(X^r) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^\infty X^r [e^{-\frac{1}{2\sigma^2}(x-\mu)^2} + e^{-\frac{1}{2\sigma^2}(x+\mu)^2}]$$
$$\times (-\log(2 - \Phi(\frac{x-\mu}{\sigma}) - \Phi(\frac{x+\mu}{\sigma})))^{\alpha-1} \quad (10)$$
$$\times (2 - \Phi(\frac{x-\mu}{\sigma}) - \Phi(\frac{x+\mu}{\sigma}))^{\frac{1}{\beta}-1}.$$

By following Lemma 1 and Lemma 2, the equation above can be simplified as

$$E(X^r) = \frac{(-0.5)^r \sigma^r}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty [\Phi^{-1}(e^{-u})]^r u^{\alpha-1} e^{-\frac{u}{\beta}} du. \quad (11)$$

Note that there is no closed form solution available for the integral above, similar to the moments of the GHN distribution (see [2, p. 109]).

Choice of Initials

This section elaborates on how to choose initial values for the parameter(s) of the HN, FN, GHN, and GFN distributions. Like many optimization techniques, the BFGS method often converges and performs well in estimating model parameters if the initials for the parameters are close enough to the roots of expressions set to zero. The sample mean ($\bar{x}$) is considered in this paper as a reasonable initial for the location parameter μ in the FN distribution, that is, $\mu_{\text{init}} = \bar{x}$, where $X \sim FN(\mu, \sigma)$. Moreover, an association between parameters μ and σ in the FN distribution is given by [6].

$$\sigma^2 = \frac{\sum_{i=1}^n x_i^2}{n} - \mu^2. \quad (12)$$

Based on this relationship, the initial for the scale parameter σ is chosen as

$$\sigma_{\text{init}} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2} \quad (13)$$

where the parameter μ is replaced by the sample mean $\bar{x}$.
In addition, a set of initials for the parameters of the GHN distribution is provided by [2]

$$\sigma_{\text{init}} = \sqrt{\frac{\pi}{2}}\bar{x}, \quad \alpha_{\text{init}} = \frac{\bar{y}^2}{s_y^2}, \quad \text{and} \quad \beta_{\text{init}} = \frac{\bar{s}_y^2}{\bar{y}} \tag{14}$$

where $\bar{y}$ and s_y are the sample mean and standard deviation of $y_i = \log(2\Phi(-\frac{x_i}{\sigma}))$, $i = 1,2,\ldots,n$ that is assumed to be drawn from the Gamma distribution with parameters α and β. The equation $\sigma_{\text{init}} = \sqrt{\frac{\pi}{2}}\bar{x}$ can also be used as the initial for the scale parameter σ in the HN distribution, since this distribution is a special case of the GHN distribution where $\alpha = \beta = 1$. This paper utilizes the sample mean as the initial for the location parameter μ in the GFN distribution. The formulas in (14) are used to obtain the initials for the other parameters, where $y_i = -\log(2 - \Phi(\frac{x_i-\mu}{\sigma}) - \Phi(\frac{x_i+\mu}{\sigma}))$, $i = 1,2,\ldots,n$ come from the Gamma distribution. As will be elaborated in the next section, the conformity of the set of initials above when estimating model parameters for the GFN distribution is empirically tested and validated via a simulation study.

Simulations

This section conducts two separate simulations. In the first, the conformity of the sets of initials presented in the previous section are empirically tested and validated by measuring the bias, precision, and accuracy with varying sample sizes for the HN, FN, GHN, and GFN distributions, respectively. In the second simulation, the performance of the distributions are compared to each other in relation to varying sample sizes based on the log likelihood values. Parameter estimation in the simulations is performed using the method of maximum likelihood with the Conjugate-Gradient (CG) optimisation method.
For the first simulation, the measures used for the estimates of each parameter can be defined as:

$$\text{Bias} = \frac{1}{S}\sum_{s=1}^{S} (\hat{\theta}_s - \theta),$$
$$\text{Precision} = \tfrac{1}{S}\textstyle\sum_{s=1}^{S} (\hat{\theta}_s - \bar{\theta})^2, \tag{15}$$
$$\text{Accuracy} = \frac{1}{S}\sum_{s=1}^{S} (\hat{\theta}_s - \theta)^2,$$

where $\bar{\theta} = \frac{1}{S}\sum_{s=1}^{S} \hat{\theta}_s$ for $s = 1,2,\ldots,1000$. The steps below are executed to obtain these measures under each of the distributions:

Step 1:
(a) *For the HN distribution:* Set $\sigma = 2$ as the population value of the scale parameter, where the location parameter μ is assumed by definition to be set to zero.

ISSN: 2536-4383

(b) *For the FN distribution:* Set $\mu = 4$ and $\sigma = 1$ as the population values of the location and scale parameters, respectively.

(c) *For the GHN distribution:* Set the population values of the parameters as $\alpha = 1$, $\beta = 1$, and $\sigma = 2$.

(d) *For the GFN distribution:* Set the population values of the parameters as $\alpha = 1$, $\beta = 1$, $\mu = 30$, and $\sigma = 0.8$.

Step 2: Choose the sample size as $N = 20, 50, 100$ or 1000.

Step 3: Generate the data from one of the listed distributions with the particular parameter values defined in Step 1 and with the sample size chosen in Step 2.

Step 4: Set the initials for the parameters based on the information given in the previous subsection.

Step 5: Estimate the model parameter(s) for each distribution and calculate the bias, precision, and accuracy measures in (15). Repeat steps from 3 to 5 by 1000 times.

Step 6: Calculate the means of the estimates and measures in Step 5 across 1000 simulation trials.

Table 1 displays the results of the first simulation for the sample sizes of 20, 50, 100 or 1000. The estimates of parameters get closer to the population values of parameters (i.e., smaller bias), thus better precision and accuracy measures are obtained as the sample size increases. This implies that the set of initials provided in the previous section are adequately determined.

For the second simulation, the data sets are generated with varying sample sizes (i.e., 20, 50, 100, and 1000) from the HN, FN, and GHN distributions. The performance of the GFN distribution is inspected along with these distributions with respect to log likelihood values. Figures 2, 3, and 4 display the log likelihood values obtained across the 1000 simulation trials, where $X \sim HN(\sigma = 2)$, $X \sim FN(\mu = 4, \sigma = 1)$, and $X \sim GHN(\alpha = 1, \beta = 1, \sigma = 2)$, respectively. The log likelihoods for the GFN distribution in these figures are often larger than those for other distributions, no matter what the underlying distributions of the simulated data sets are. The superiority of the GFN distribution becomes more apparent for larger sample sizes. It should be noted that, since the HN distribution has only the scale parameter, it is not a flexible distribution, and thus, it does not perform well (i.e., it has small log likelihood value) when compared to other distributions. In the simulations, this happened when the data sets are generated from the HN distribution using a small sample size (i.e., when $n = 20$) and when the data sets are generated from the FN distribution, where the location parameter is $\mu = 4$ and not $\mu = 0$.

ISSN: 2536-4383

***Table 1**. Estimates, bias, precision, and accuracy measures for the parameters of the HN, FN, GHN, and GFN distribution*

$X \sim HN(\sigma = 2)$				
Sample size	$\hat{\sigma}$	Bias (σ)	Precision (σ)	Accuracy (σ)
$N = 20$	1.969	-0.031	0.097	0.098
$N = 50$	1.993	-0.007	0.041	0.041
$N = 100$	1.995	-0.005	0.018	0.018
$N = 1000$	2.000	-0.0004	0.002	0.002
$X \sim FN(\mu = 4, \sigma = 1)$				
Sample size	$\hat{\mu}$	Bias (μ)	Precision (μ)	Accuracy (μ)
$N = 20$	3.999	-0.001	0.052	0.052
$N = 50$	3.991	-0.009	0.021	0.021
$N = 100$	3.997	-0.003	0.011	0.011
$N = 1000$	4.000	0.0005	0.001	0.001
Sample size	$\hat{\sigma}$	Bias (σ)	Precision (σ)	Accuracy (σ)
$N = 20$	0.958	-0.042	0.024	0.026
$N = 50$	0.986	-0.014	0.010	0.010
$N = 100$	0.992	-0.008	0.005	0.005
$N = 1000$	0.999	-0.001	0.0005	0.0005
$X \sim GHN(\alpha = 1, \beta = 1, \sigma = 2)$				
Sample size	$\hat{\alpha}$	Bias (α)	Precision (α)	Accuracy (α)
$N = 20$	1.150	0.150	0.164	0.187
$N = 50$	1.042	0.042	0.055	0.056
$N = 100$	1.019	0.019	0.028	0.029
$N = 1000$	1.008	0.008	0.004	0.004
Sample size	$\hat{\beta}$	Bias (β)	Precision (β)	Accuracy (β)
$N = 20$	1.533	0.533	1.726	2.011
$N = 50$	1.592	0.592	1.727	2.078
$N = 100$	1.405	0.405	1.016	1.180
$N = 1000$	1.024	0.024	0.107	0.108
Sample size	$\hat{\sigma}$	Bias (σ)	Precision (σ)	Accuracy (σ)
$N = 20$	2.242	0.242	1.683	1.741
$N = 50$	2.179	0.179	1.300	1.332
$N = 100$	2.131	0.131	0.960	0.977

ISSN: 2536-4383

Table 1*. …continued*

$X \sim GFN(\alpha = 1, \beta = 1, \mu = 30, \sigma = 0.8)$				
Sample size	$\hat{\alpha}$	Bias (α)	Precision (α)	Accuracy (α)
$N = 20$	0.811	-0.189	0.279	0.315
$N = 50$	0.903	-0.097	0.229	0.238
$N = 100$	0.970	-0.030	0.127	0.128
$N = 1000$	1.013	0.013	0.017	0.017
Sample size	$\hat{\beta}$	Bias (β)	Precision (β)	Accuracy (β)
$N = 20$	1.904	0.904	3.012	3.830
$N = 50$	1.693	0.693	1.794	2.274
$N = 100$	1.344	0.344	0.824	0.942
$N = 1000$	1.062	0.062	0.118	0.122
Sample size	$\hat{\mu}$	Bias (μ)	Precision (μ)	Accuracy (μ)
$N = 20$	30.213	0.213	1.309	1.355
$N = 50$	30.093	0.093	0.955	0.964
$N = 100$	30.048	0.048	0.647	0.619
$N = 1000$	29.988	-0.012	0.103	0.104
Sample size	$\hat{\sigma}$	Bias (σ)	Precision (σ)	Accuracy (σ)
$N = 20$	0.609	-0.191	0.444	0.481
$N = 50$	0.671	-0.129	0.061	0.078
$N = 100$	0.732	-0.068	0.025	0.030
$N = 1000$	0.795	-0.005	0.004	0.005

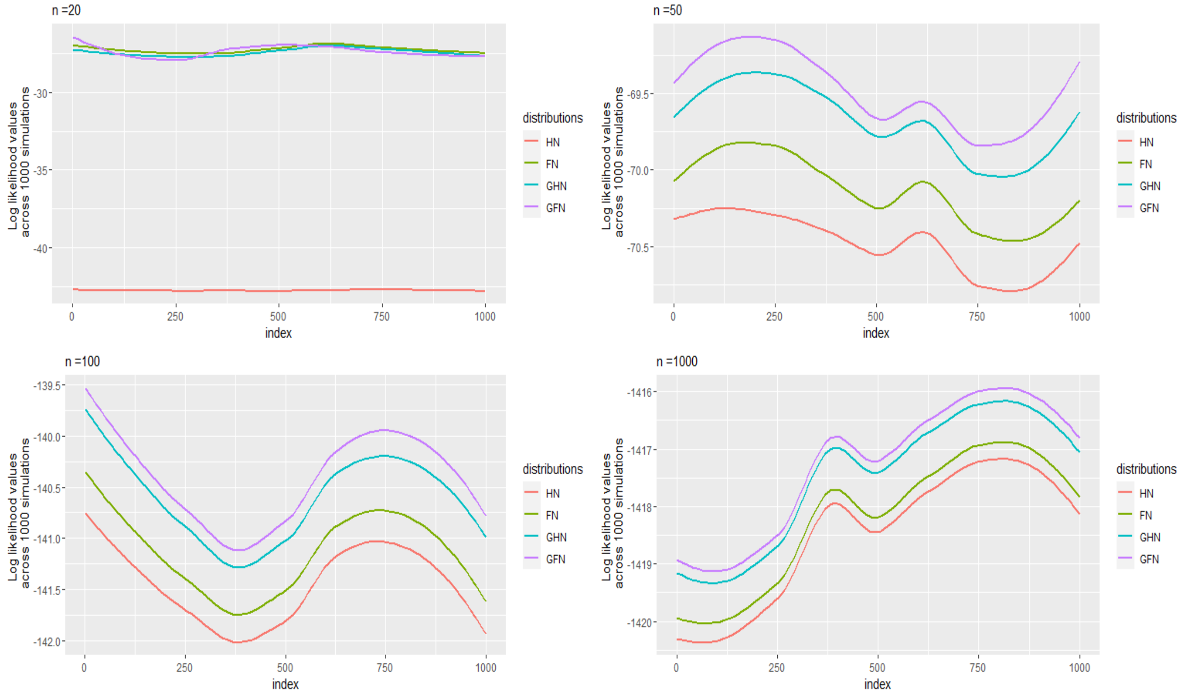

Figure 2. *The log likelihood values with varying sample sizes where* $X \sim HN(\sigma = 2)$.

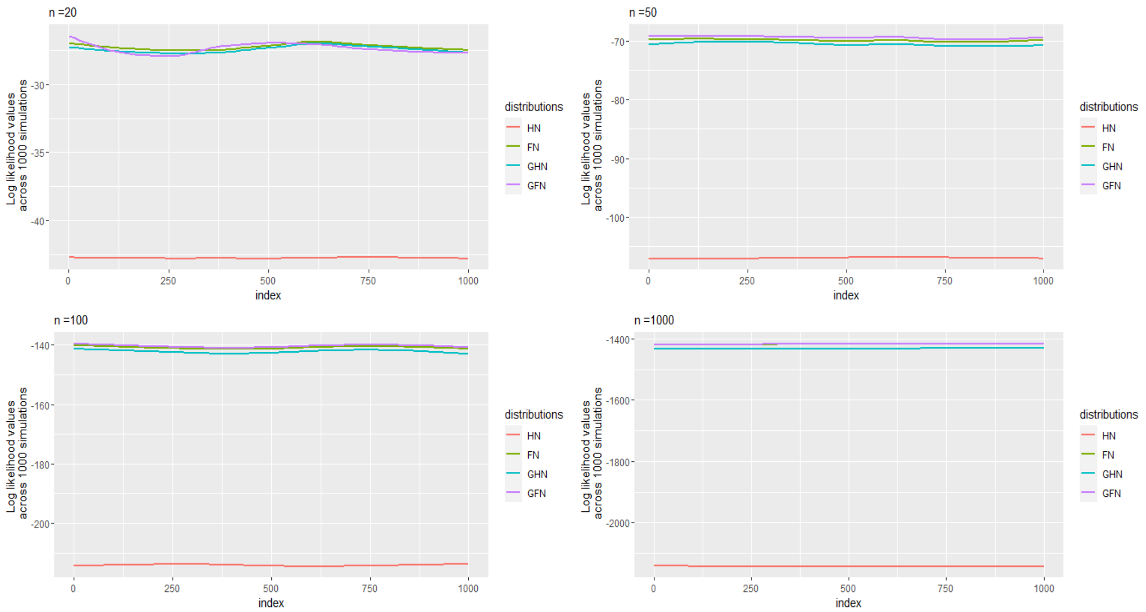

Figure 3. *The log likelihood values with varying sample sizes where* $X \sim FN(\mu = 4, \sigma = 1)$.

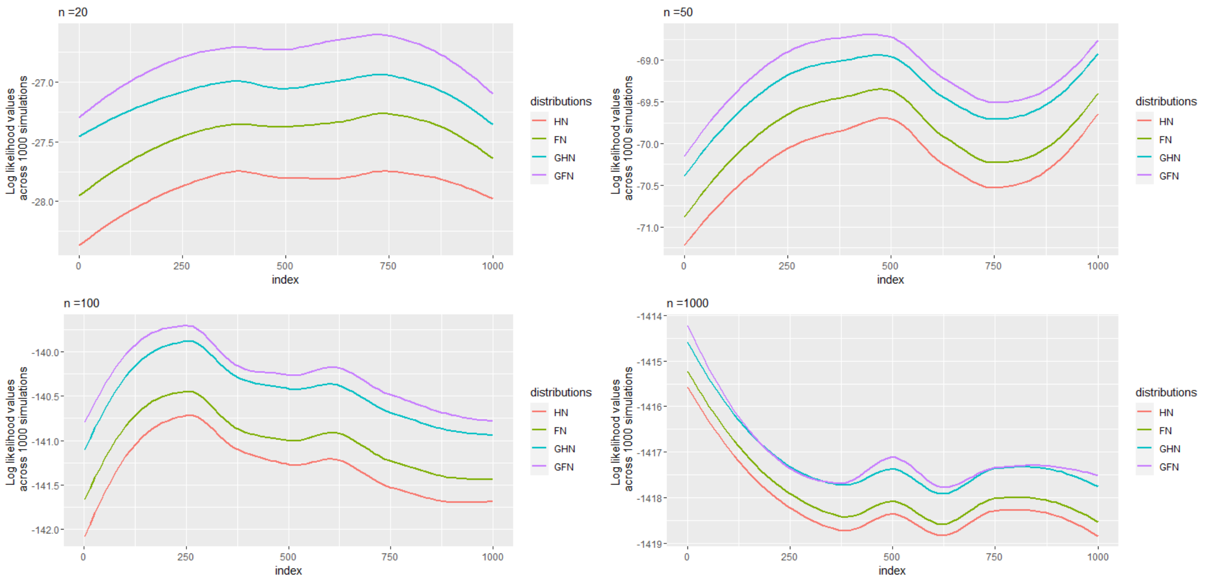

Figure 4. *The log likelihood values with varying sample sizes where* $X \sim GHN(\alpha = 1, \beta = 1, \sigma = 2)$.

Real Data Example

In order to illustrate the fitting performance of the new Gamma Folded-Normal distribution, the strength data for single carbon fibers first reported by Bader and Priest [7] were used. Single fibers were measured in GPA and tested under tension at gauge lengths of 1, 10, 20 and 50 mm. For the current study, we considered the transformed form of the tensile strength data for the single fibers of 20mm (N=69) which have also been the subject of some other studies for modeling purposes [8, 2]. The performance assessment of GFN was achieved by comparing its probability density function and cumulative density function with those of the Half-Normal, Folded-Normal, and Gamma Half-Normal distributions. Figure 5 presents the histogram of the data and the estimated HN, FN, GHN, and GFN distributions. This figure shows that the fits of the GFN and FN distributions to the data are better than that of the HN and GHN distributions. Figure 6 displays the cumulative density functions for all distributions. It should be noted that the cumulative density function for the GFN distribution (along with the CDFs of other distributions except the HN distribution) fluctuates around the empirical cumulative distribution function, which implies that the GFN distribution performs well in explaining the underlying distribution of the data.

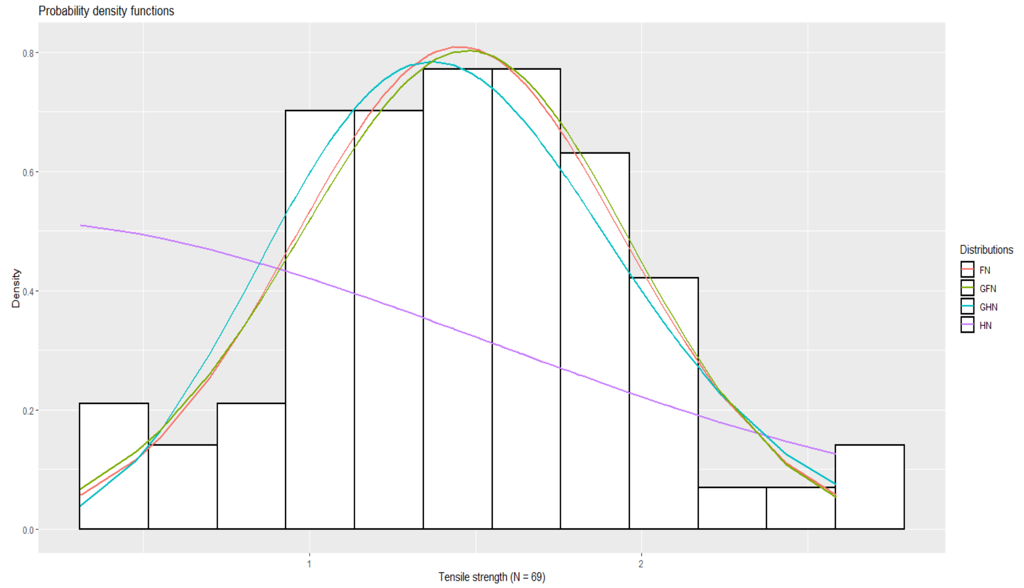

Figure 5. *Histogram of the data and probability density functions with the fit of HN, FN, GHN, and GFN distributions*.

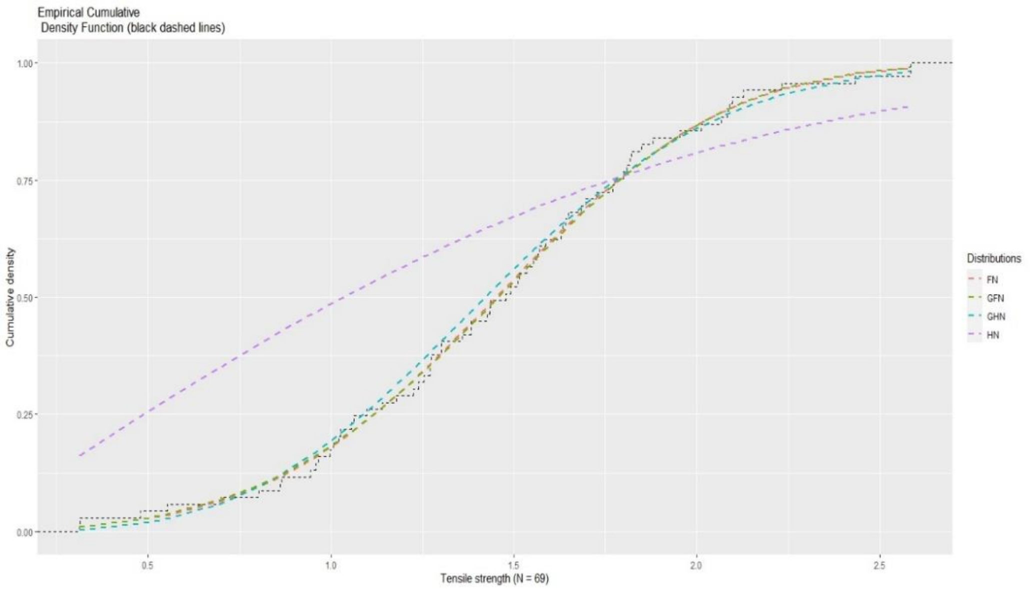

Figure 6. *Cumulative density functions for the HN, FN, GHN, and GFN distributions.*

Table 2 displays the maximum likelihood estimates (MLEs) of model parameters for the HN, FN, GHN, and GFN distributions. Table 3 shows the log likelihood values for the distributions. The results presented in this table indicate that the GFN distribution has the best fit among the four distributions. In comparison to the GHN distribution, the fitting performance seems to be improved by the additional location parameter.

Table 2. *Parameter estimates of the HN, FN, GHN, and GFN distributions for the tensile strength data*

Distributions	Parameters	Estimates
HN	σ_{HN}	$\hat{\sigma}_{HN} = 1.53$
FN	μ_{FN}	$\hat{\mu}_{FN} = 1.45$
	σ_{FN}	$\hat{\sigma}_{FN} = 0.49$
GHN	α_{GHN}	$\hat{\alpha}_{GHN} = 2.89$
	β_{GHN}	$\hat{\beta}_{GHN} = 3.11$
	σ_{GHN}	$\hat{\sigma}_{GHN} = 0.40$
GFN	α_{GFN}	$\hat{\alpha}_{GFN} = 2.09$
	β_{GFN}	$\hat{\beta}_{GFN} = 0.20$
	μ_{GFN}	$\hat{\mu}_{GFN} = 1.95$
	σ_{GFN}	$\hat{\sigma}_{GFN} = 0.94$

Table 3. *The log likelihood values of the four distributions for the tensile strength data*

Distributions	Log likelihoods
HN	-79.527
FN	-48.851
GHN	-49.665
GFN	48.631

Discussion

This paper proposes a new distribution named as the Gamma Folded-Normal distribution. This distribution is flexible in the sense that it contains α and β parameters of the Gamma distribution and μ and σ parameters of the Folded-Normal distribution. First, an appropriate set of initials for these parameters are determined. Then, the adequacy of the set of initials are empirically tested and validated by means of a simulation study. It has been shown by another simulation study that the GFN distribution often has larger log likelihood values than the HN, FN, and GHN distributions especially when the sample size is large. A real life data example is used to illustrate the applicability of the GFN distribution. Both the emprical and real world data indicate that the GFN distribution provide an adequate fit to the data. It can be concluded that our proposal GFN distribution enriches the Gamma-X family of distributions and provides a valuable distributional form for modeling the non-negative measurements.

Appendix A.

The derivative of the log likelihood function in (3) with respect to each of the four parameters for the GFN distribution are given by

$$\frac{\partial \log L(\alpha,\beta,\mu,\sigma)}{\partial \alpha} = -nlog(\beta) - n\Psi(\alpha), \tag{16}$$

where $\Psi = \frac{d}{dx} log(\Gamma(\alpha)) = \frac{\Gamma\prime(\alpha)}{\Gamma(\alpha)}$,

$$\frac{\partial \log L(\alpha,\beta,\mu,\sigma)}{\partial \beta} = -\frac{1}{\beta^2}\sum_{i=1}^{n} \log(2 - \Phi(\frac{x_i-\mu}{\sigma}) - \Phi(\frac{x_i+\mu}{\sigma})), \tag{17}$$

$$\frac{\partial \log L(\alpha,\beta,\mu,\sigma)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) - \frac{2}{\sigma^2}\sum_{i=1}^{n}\frac{x_i e^{-\frac{2\mu x_i}{\sigma^2}}}{1+e^{-\frac{2\mu x_i}{\sigma^2}}}$$
$$-\frac{\alpha-1}{\sigma}\sum_{i=1}^{n}\frac{h_z(\frac{x_i+\mu}{\sigma}) - h_z(\frac{x_i-\mu}{\sigma})}{\eta\log(\eta)} \tag{18}$$
$$+\frac{(1-\beta)}{\sigma\beta}\sum_{i=1}^{n}\frac{h_z(\frac{x_i-\mu}{\sigma}) - h_z(\frac{x_i+\mu}{\sigma})}{\eta},$$

and

$$\frac{\partial \log L(\alpha,\beta,\mu,\sigma)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)^2 + \frac{4\mu}{\sigma^3}\sum_{i=1}^{n}\frac{x_i e^{-\frac{2\mu x_i}{\sigma^2}}}{1+e^{-\frac{2\mu x_i}{\sigma^2}}}$$
$$+(\alpha-1)\sum_{i=1}^{n}\frac{h_z(\frac{x_i-\mu}{\sigma})(\frac{x_i-\mu}{\sigma^2}) + h_z(\frac{x_i+\mu}{\sigma})(\frac{x_i+\mu}{\sigma^2})}{\eta\log(\eta)} \tag{19}$$
$$+(1-\frac{1}{\beta})\sum_{i=1}^{n}\frac{h_z(\frac{x_i-\mu}{\sigma})(\frac{x_i-\mu}{\sigma^2}) + h_z(\frac{x_i+\mu}{\sigma})(\frac{x_i+\mu}{\sigma^2})}{\eta},$$

where $\eta = 2 - \Phi(\frac{x_i-\mu}{\sigma}) - \Phi(\frac{x_i+\mu}{\sigma})$ and $h_z(.)$ is the hazard rate function with $h_z(\frac{x_i-\mu}{\sigma}) = \frac{\delta(\frac{x_i-\mu}{\sigma})}{1-\Phi(\frac{x_i-\mu}{\sigma})}$ and $h_z(\frac{x_i+\mu}{\sigma}) = \frac{\delta(\frac{x_i+\mu}{\sigma})}{1-\Phi(\frac{x_i+\mu}{\sigma})}$ and $\delta(.)$ is the PDF of the Standard Normal distribution.

Acknowledgments -

Funding/Financial Disclosure The authors have no received any financial support for the research, authorship, or publication of this study.

Ethics Committee Approval and Permissions The study does not require ethics committee permission or any special permission.

Conflict of Interests The authors state that there is no conflict of interest related to this article.

Authors Contribution All authors read and approved the final manuscript.

References

[1] Alzaatreh, A., Lee, C., & Famoye, F. (2013). A new method for generating families of continuous distributions. *Metron*, *71*(1), 63–79. https://doi.org/10.1007/s40300-013-0007-y

[2] Alzaatreh, A. & Knight, K. (2013). On the gamma-half normal distribution and its applications. *Journal of Modern Applied Statistical Methods*, *12*(1), 103-119. https://doi.org/10.22237/jmasm/1367381640

[3] Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, *8*, 135–160. https://doi.org/10.1177/096228029900800204

[4] Leone, F. C., Nelson, L. S., & Nottingham, R. B. (1961). The folded normal distribution. *Technometrics 3*, 543–550. https://doi.org/10.2307/1266560

[5] Tsagris, M., Beneki, C., & Hassani, H. (2014). On the folded normal distribution. *Mathematics*, *2*(1), 12–28. https://doi.org/10.48550/arXiv.1402.3559

[6] R Core Team (2016). R: A language and environment for statistical computing. Vienna, Austria: *R Foundation for Statistical Computing*. https://www.R-project.org/

[7] Bader, M., & Priest, A. (1982). Statistical Aspects of Fiber and Bundle Strength in Hybrid Composites. *In: Hayashi, T., Kawata, S. and Umekawa, S., Eds., Progress in Science and Engineering Composites*, ICCM-IV, Tokyo, 1129-1136.

[8] Raqab, M, Z., & Kundu, D. (2005). Comparison of different estimators of P (Y < X) for a scaled Burr Type X distribution, *Communications in Statistics - Simulation and Computation*, 34(2), 465-483. https://doi.org/10.1080/03610920802162664