5-1-2013

On The Gamma-Half Normal Distribution and Its Applications

Ayman Alzaatreh
Austin Peay State University, Clarksville, TN

Kristen Knight
Austin Peay State University, Clarksville, TN

Follow this and additional works at: http://digitalcommons.wayne.edu/jmasm

Part of the Applied Statistics Commons, Social and Behavioral Sciences Commons, and the Statistical Theory Commons

On The Gamma-Half Normal Distribution and Its Applications

Ayman Alzaatreh   Kristen Knight
Austin Peay State University,
Clarksville, TN

A new distribution, the gamma-half normal distribution, is proposed and studied. Various structural properties of the gamma-half normal distribution are derived. The shape of the distribution may be unimodal or bimodal. Results for moments, limit behavior, mean deviations and Shannon entropy are provided. To estimate the model parameters, the method of maximum likelihood estimation is proposed. Three real-life data sets are used to illustrate the applicability of the gamma-half normal distribution.

Key words: *T-X* families; gamma-*X* family; unimodal; bimodal; Shannon entropy.

Introduction

In recent years, advancements in technology and science have resulted in a wealth of information, which is expanding the level of knowledge across many disciplines. This information is gathered and analyzed by statisticians, who hold the responsibility of accurately assessing the data and making inferences about the population of interest. Without this precise evaluation of data, each field remains limited to its current state of knowledge. In the last decade, it has been discovered that many well-known distributions used to model data sets do not offer enough flexibility to provide an adequate fit. For this reason, new methods are being proposed and used to derive generalizations of well-known distributions. With these distributions, strong applications have been made to real-life scenarios.

Alzaatreh, et al. (2013b) proposed the *T-X* families of distributions. These families of distributions were used to generate a new class of distributions which offer more flexibility in modeling a variety of data sets. Several members of the *T-X* families have been studied

in the literature (e.g., Alzaatreh, et al. (2013a); Alzaatreh, et al. (2013b); Alzaatreh, et al. (2012a); Alzaatreh, et al. (2012b); Lee, et al. (2013)).

One well-known distribution is the half-normal distribution, which has been used in variety of applications. Previous work by Bland and Altman (1999) used the half-normal distribution to study the relationship between measurement error and magnitude. Bland (2005) extended the work of Bland and Altman (1999) by using the distribution to estimate the standard deviation as a function so that measurement error could be controlled. In his work, various exercise tests were analyzed and it was determined that variability of performance does decline with practice (Bland, 2005). Manufacturing industries have utilized the half-normal distribution to model lifetime processes under fatigue. These industries often produce goods with a long lifetime need for customers, making the cost of the resources needed to analyze the product failure times very high. To save time and money the half normal distribution is used in this reliability analysis to study the probabilistic aspects of the product failure times (Castro, et al., 2012).

Due to the fact that the half-normal distribution has only one shape, various generalizations of the distribution have been derived. These generalizations include the generalized half-normal distribution (Cooray, et al., 2008), the beta-generalized half-normal (Pescrim, et al., 2010) and the Kumaraswamy

Ayman Alzaatreh is an Assistant Professor in the Department of Mathematics & Statistics. Email him at: alzaatreha@apsu.edu. Kristen Knight is a student in the Department of Mathematics and Statistics. Email her at: kknight13@my.apsu.edu.

generalized half-normal (Cordeiro, et al., 2012). Several of the corresponding applications include the stress-rupture life of kevlar 49/epoxy strands placed under sustained pressure (Cooray, et al., 2008), failure times of mechanical components and flood data (Cordeiro, et al., 2012). In this article the gamma and half normal distributions are combined to propose a new generalization of the half-normal distribution, namely, the gamma half-normal distribution.

Let $F(x)$ be the cumulative distribution function (CDF) of any random variable X and $r(t)$ be the probability density function (PDF) of a random variable T defined on $[0, \infty)$. The CDF of the *T-X* family of distributions defined by Alzaatreh, et al. (2013b) is given by

$$G(x) = \int_0^{-\log(1-F(x))} r(t)dt = R\{-\log(1-F(x)\}$$

(1.1)

When X is a continuous random variable, the probability density function of the *T-X* family is

$$g(x) = \frac{f(x)}{1-F(x)} r\left(-\log\left(1-F(x)\right)\right)$$
$$= h(x) r\left(H(x)\right),$$

(1.2)

where $h(x)$ and $H(x)$ are the hazard and the cumulative hazard functions of the random variable X associated with $f(x)$.

If a random variable T follows the gamma distribution with parameters α and β,

$$r(t) = \left(\beta^\alpha \Gamma(\alpha)\right)^{-1} t^{\alpha-1} e^{-t/\beta}, \ t \geq 0, \quad \text{then the}$$

definition in (1.2) leads to the gamma-*X* family with the PDF
$$g(x) =$$
$$\frac{1}{\Gamma(\alpha)\beta^\alpha} f(x)\left(-\log\left(1-F(x)\right)\right)^{\alpha-1}\left(1-F(x)\right)^{\frac{1}{\beta}-1}.$$

(1.3)

When $\beta = 1$, the gamma-*X* family in (1.3) reduces to the gamma-generated distribution introduced by Zografos and Balakrishnan

(2009). When $\alpha = 1$ and $1/\beta = n \in \mathbb{N}$, the gamma-*X* family reduces to the distribution of the first order statistics of the random variable X.

If X is the half normal random variable with the density function

$$f(x) = \frac{\sqrt{2}}{\theta\sqrt{\pi}} e^{-x^2/2\theta^2}, x > 0, \text{ then (1.3) gives}$$

$$g(x) =$$

$$\frac{\sqrt{2}}{\sqrt{\pi}\,\theta\,\Gamma(\alpha)\beta^\alpha} e^{\frac{-x^2}{2\theta^2}} (-\log(2\Phi(\frac{-x}{\theta})))^{\alpha-1} (2\Phi(\frac{-x}{\theta}))^{\frac{1}{\beta}-1},$$

$$\alpha, \beta, \theta > 0; x > 0$$

(1.4)

where Φ is the CDF of the standard normal distribution.

A random variable X with the PDF $g(x)$ in (1.4) is said to follow the gamma-half normal distribution with parameters α, β and θ. From (1.1), the CDF of the gamma- half normal distribution is obtained as

$$G(x) = \gamma\{\alpha, -\beta^{-1} \log(2\Phi(-x/\theta))\}/\Gamma(\alpha),$$

(1.5)

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

A series representation of $G(x)$ in (1.5) can be obtained by using the series expansion of the incomplete gamma function from Nadarajah and Pal (2008) as

$$\gamma(\alpha,x) = \sum_{k=0}^\infty \frac{(-1)^k x^{\alpha+k}}{k!(\alpha+k)}.$$

(1.6)

From (1.6), the CDF of the gamma half-normal distribution can be written as

$$G(x) = \frac{1}{\Gamma(\alpha)} \sum_{k=0}^\infty \frac{(-1)^k [-\log(2\Phi(-x/\theta))]^{\alpha+k}}{k!(\alpha+k)\beta^{\alpha+k}}.$$

(1.7)

The hazard function associated with the gamma-half normal distribution is

$$h(x) = \frac{g(x)}{1-G(x)}$$

$$= \frac{\sqrt{2}e^{-x^2/2\theta^2}(-\log(2\Phi(-x/\theta)))^{\alpha-1}(2\Phi(-x/\theta))^{\frac{1}{\beta}-1}}{\sqrt{\pi}\theta\beta^{\alpha}[\Gamma(\alpha)-\gamma\{\alpha,-\beta^{-1}\log(2\Phi(-x/\theta))\}]},$$

$$x > 0.$$

(1.8)

Some Properties of the Gamma-Half Normal Distribution

Lemma 1 gives the relation between the gamma-half normal distribution and the gamma distribution.

Lemma 1 (Transformation)

If a random variable Y follows the gamma distribution with parameters α and β, then the random variable $X = \theta\Phi^{-1}(1-0.5e^{-Y})$ follows the gamma-half normal distribution with parameters α, β and θ.

Lemma 1 (Transformation) Proof

The results follow by using the transformation technique.

The limiting behaviors of the gamma-half normal PDF and the hazard function are given in Lemma 2.

Lemma 2

The limit of the gamma-half normal density function as $x \to \infty$ is 0 and the limit of the gamma-half normal hazard function as $x \to \infty$ is ∞. Also, the limit of the gamma-half normal and hazard function as $x \to 0^+$ is given by

$$\lim_{x \to 0^+} g(x) = \lim_{x \to 0^+} h(x) = \begin{cases} 0, & \alpha > 1 \\ \dfrac{\sqrt{2}}{\sqrt{\pi}\,\theta\beta}, & \alpha = 1 \\ \infty, & \alpha < 1. \end{cases}$$

(2.1)

Lemma 2 Proof

Since the random variable X is defined on $(0,\infty)$, this implies $\lim_{x\to\infty} g(x) = 0$. Using L'Hôpital's rule it can be shown that $\lim_{x\to\infty} h(x) = \infty$. Now, $h(x)[1-G(x)] = g(x)$ implies that $\lim_{x\to 0^+} g(x) = \lim_{x\to 0^+} h(x)$. Results in (2.1) follow immediately from definition (1.4). The modes of the gamma-half normal distribution can be obtained by taking the derivative of $g(x)$. The derivative with respect to x of (1.4) can be simplified to

$$g'(x) =$$

$$\frac{\sqrt{2}}{\sqrt{\pi}\theta\Gamma(\alpha)\beta^{\alpha}}e^{\frac{-x^2}{2\theta^2}}(-\log(2\Phi(\frac{-x}{\theta})))^{\alpha-2}[2\Phi(\frac{-x}{\theta})]^{\frac{1}{\beta}-2}k(x),$$

(2.2)

where

$$k(x) =$$

$$\begin{bmatrix} \dfrac{x}{\theta}\log(2\Phi(\dfrac{-x}{\theta})) + (\alpha-1)h_z(\dfrac{x}{\theta}) \\ + (\beta^{-1}-1)h_z(\dfrac{x}{\theta})\log(2\Phi(\dfrac{-x}{\theta})) \end{bmatrix}$$

Setting (2.2) to 0, the critical values of $g(x)$ are $x = 0$ and the solution of the equation $k(x) = 0$. The solution of $k(x) = 0$ is equivalent to the equation

$$x = \theta h_z(x/\theta)\left\{1 - \frac{1}{\beta} - \frac{\alpha-1}{\log(2\Phi(-x/\theta))}\right\},$$

(2.3)

where $h_z(x/\theta) = \phi(x/\theta)/(1-\Phi(x/\theta))$.

Corollary 1

If $\alpha \le 1$ and $\beta \le 1$, the gamma-half normal distribution is unimodal and the mode is at $x = 0$.

Corollary 1 Proof

If $\alpha < 1$, then Lemma 2 implies that $x = 0$ is a modal point. When $\alpha < 1$ and $\beta \leq 1$, it follows from (2.3) that $x < 0$, and hence equation (2.3) has no solution, thus, $x = 0$ is a unique modal point. The proof is complete by noting that when $\alpha = 1$ and $\beta \leq 1$, the PDF of gamma-half normal in (1.4) is a strictly decreasing function.

Figures 1-3 show various graphs of $g(x)$ and $h(x)$. These figures indicate that the gamma-half normal PDF may take on a variety of shapes for different values of α, β and θ. The shapes range from reversed-J shape, bimodal, right-skewed and approximately symmetric. As β decreases, the right tail of the gamma-half normal distribution becomes longer. Bimodality appears when α is less than 1. Figure 3 indicates that the gamma-half normal hazard function is either a bathtub shape or increasing failure rate shape.

When $\alpha < 1$, and for certain values of β, the gamma-half normal distribution becomes bimodal. It is difficult to find analytically the region where the distribution is bimodal. However, a numerical solution is obtained to determine the number of roots of the derivative of the gamma-half normal distribution. Figure 4 shows the boundary region of α and β where the gamma-half normal distribution is bimodal.

Lemma 3

If $Q(\lambda)$, $0 < \lambda < 1$ denotes the quantile function for the gamma-half normal distribution, then

$$Q(\lambda) = -\theta \Phi^{-1}\left(0.5 \exp\{-\beta \gamma^{-1}(\alpha, \lambda \Gamma(\alpha))\}\right).$$

(2.4)

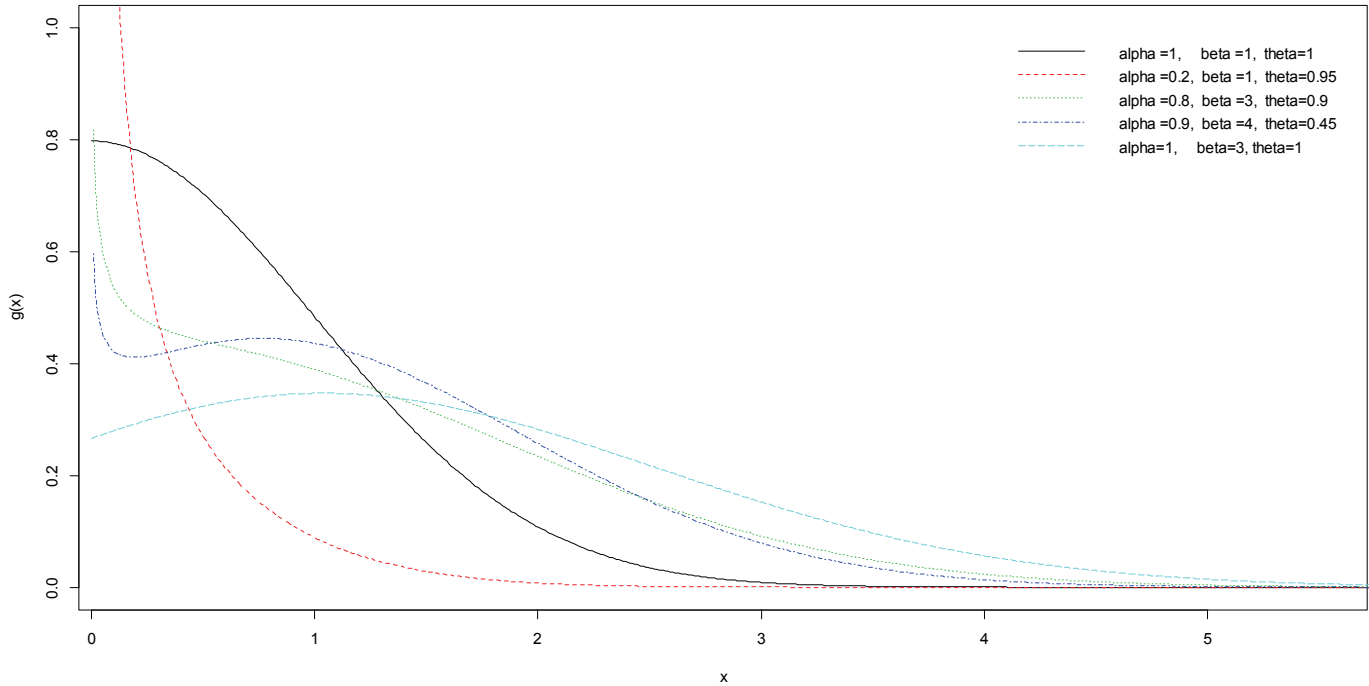Figure 1: The Gamma-Half Normal PDF for Various Values of α, β and θ

Figure 2: The Gamma-Half Normal PDF for Various Values of α, β and θ
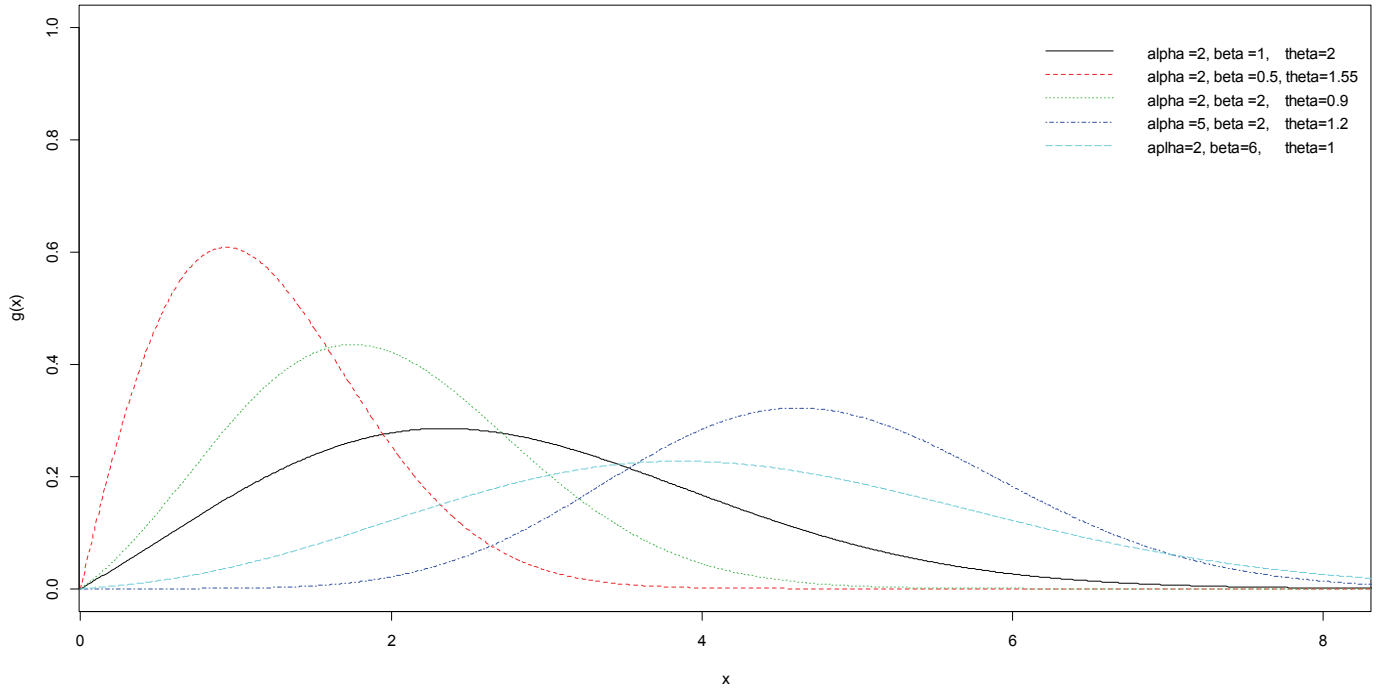
Figure 3: The Gamma-Half Normal Hazard Function for Various Values of α, β and θ
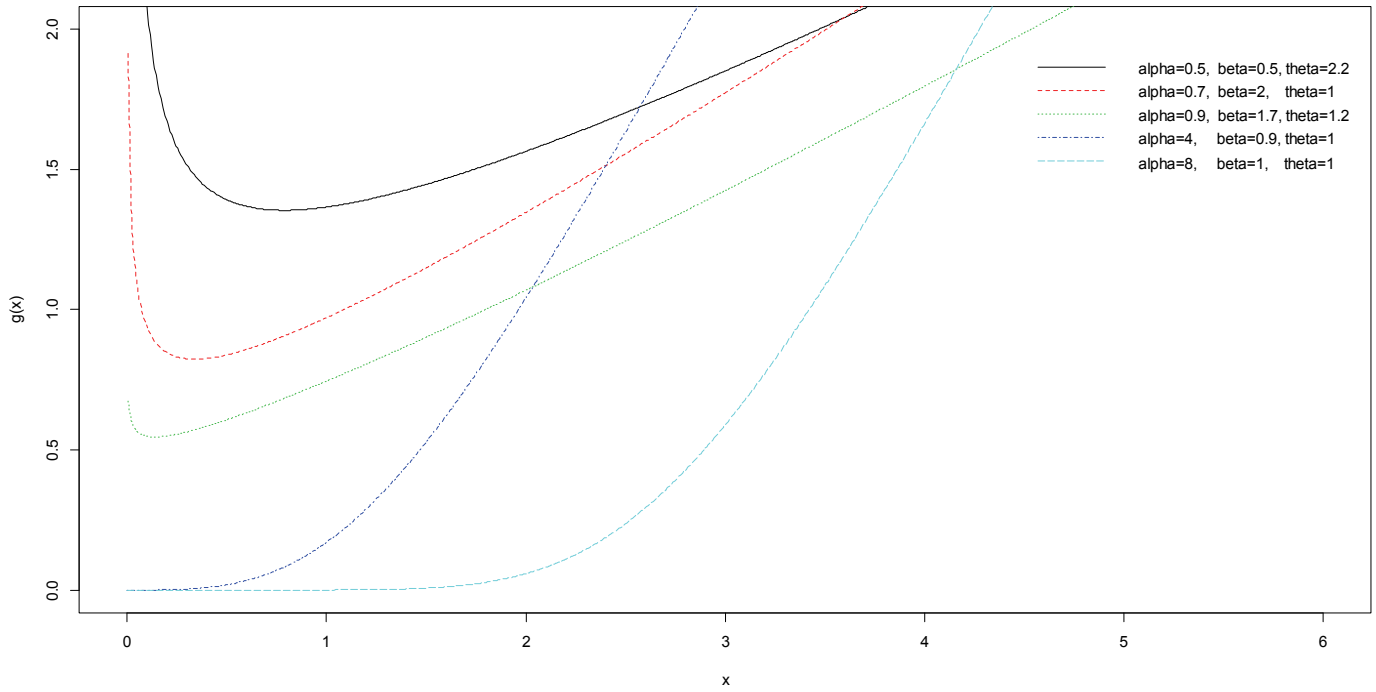
Lemma 3 Proof

The proof follows by taking the inverse function of (1.5).

The entropy of a random variable X is a measure of variation of uncertainty (Rényi, 1961). Shannon's entropy (Shannon, 1948), for a random variable X with PDF $g(x)$ is defined as $E\{-\log(g(X))\}$. Since 1948 many applications have been used with Shannon's entropy in different areas, including engineering, physics, biology, economics and information theory.

According to Alzaatreh, et al. (2013b), the Shannon entropy of the gamma-X family of distributions is given by

$$\eta_X = -E\left\{\log f\left(F^{-1}\left(1-e^{-T}\right)\right)\right\} + \alpha(1-\beta)$$
$$+ \log \beta + \log \Gamma(\alpha) + (1-\alpha)\psi(\alpha)$$

(2.5)

where ψ is the digamma function and T is the gamma random variable with parameters α and β.
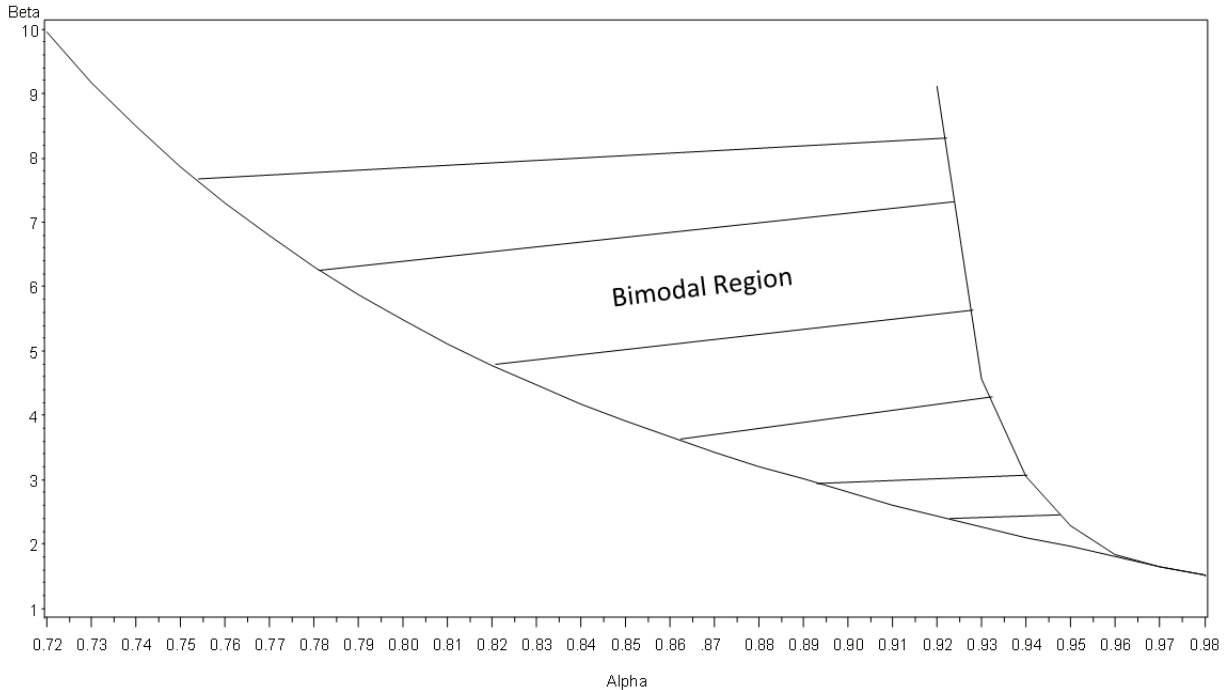
Theorem 3

The Shannon entropy for the gamma-half normal distribution is given by

$$\eta_X = -\log\sqrt{2} + \log\sqrt{\pi} + \log\theta$$
$$+ \frac{1}{2\theta^2}(\sigma^2 + \mu^2) + \alpha(1-\beta) + \log\beta$$
$$+ \log\Gamma(\alpha) + (1-\alpha)\psi(\alpha),$$

(2.6)

where μ and σ are the mean and variance of the gamma-half normal, respectively.

Figure 4: Bimodal Region for the Gamma-Half Normal Density Function where $\theta = 1$

Theorem 3 Proof

First it is necessary to find $-E\left\{\log f\left(F^{-1}\left(1-0.5e^{-T}\right)\right)\right\}$, where $f(x)$ and $F(x)$ are the PDF and CDF of the half-normal distribution, respectively. From the CDF of the half-normal distribution it follows that $F^{-1}(x) = \theta \Phi^{-1}(\frac{x+1}{2})$ and hence,

$$
\begin{aligned}
-E\left\{\log f\left(F^{-1}\left(1-e^{-T}\right)\right)\right\} \\
= -\log\sqrt{2} + \log\sqrt{\pi} + \log\theta \\
+ (\frac{1}{2\theta^2})E(\theta\Phi^{-1}(1-0.5e^{-T}))^2
\end{aligned}
$$

where T follows the gamma distribution. By Lemma 1, $\theta\Phi^{-1}(1-0.5e^{-T})$ follows the gamma-half normal with parameters α, β and θ. Hence $E(\theta\Phi^{-1}(1-0.5e^{-T})) = \sigma^2 + \mu^2$ where σ^2 and μ are the variance and the mean for the gamma-half normal distribution. The result in (2.5) follows from equation (2.4).

Moments and Mean Deviations

The rth moments for the gamma-half normal distribution in (1.4) can be written as

$$
E(X^r) =
$$

$$
\frac{\sqrt{2}}{\sqrt{\pi}\theta\Gamma(\alpha)\beta^\alpha}\int_0^\infty x^r e^{\frac{-x^2}{2\theta^2}}(-\log(2\Phi(\frac{-x}{\theta})))^{\alpha-1}(2\Phi(\frac{-x}{\theta}))^{\frac{1}{\beta}-1}dx.
$$

$$(3.1)$$

Using the substitution $u = -\log(2\Phi(-x/\theta))$, (3.1) reduces to

$$
E(X^r) = \frac{(-1)^r\theta^r}{\Gamma(\alpha)\beta^\alpha}\int_0^\infty (\Phi^{-1}(0.5e^{-u}))^r u^{\alpha-1}e^{-u/\beta}du.
$$

$$(3.2)$$

Because no closed form is found for (3.2), numerical integration can be used to calculate the rth moments.

Table 1 provides the mode, mean, median and variance of the gamma-half normal distribution for various values of α and β when $\theta = 1$. Equations (2.2) and (3.2) are used for these calculations. For fixed α and θ, the mode, mean, median and variance are increasing functions of β. Also, for fixed β and θ, the mode, mean, median and variance are increasing functions of α. Figure 5 displays the skewness and kurtosis graphs of the gamma-half normal distribution for different values of α and β with $\theta = 1$. For fixed α, the skewness and kurtosis are decreasing functions of β; for fixed β, the skewness and kurtosis are decreasing functions of α.

Lemma 5

If the median is denoted by M, then the mean deviation from the mean, $D(\mu)$, and the mean deviation from the median, $D(M)$, for the gamma-X distribution are given by

$$
D(\mu) = 2\mu\gamma\{\alpha, \beta^{-1}\log(2\Phi(-\mu)\}/\Gamma(\alpha) - 2I_\mu
$$

and

$$
D(M) = \mu - 2I_M, \qquad (3.3)
$$

where $\quad I_m = \frac{\sqrt{2}\theta}{\Gamma(\alpha)}\sum_{k=0}^\infty\sum_{i=0}^{2k+1}a(k,i)\gamma(\alpha,\delta_i(m))$,

$$
a(k,i) = \frac{c_k}{2k+1}(\sqrt{\pi}/2)^{2k+1}\binom{2k+1}{i}\frac{(-1)^i}{(1+i\beta)^\alpha}
$$

and $\delta_i(m) = -(i+1/\beta)\log(2\Phi(-m/\theta))$.

Lemma 5 Proof

If $g(x)$ and $G(x)$ are the PDF and the CDF of the gamma-half normal distribution, then the mean deviations from the mean and the median can be written as

$$
D(\mu) = 2\mu G(\mu) - 2\int_0^\mu xg(x)dx \quad (3.4)
$$

and

$$
D(M) = \mu - 2\int_0^M xg(x)dx. \qquad (3.5)
$$

Consider the integral:

$$I_m = \int_0^m x g(x)\,dx$$

$$= \frac{\sqrt{2}}{\sqrt{\pi}\,\theta\,\Gamma(\alpha)\beta^\alpha} \int_0^m x e^{\frac{-x^2}{2\theta^2}} (-\log(2\Phi(\frac{-x}{\theta})))^{\alpha-1} (2\Phi(\frac{-x}{\theta}))^{(1/\beta)-1}\,dx$$

$$(3.6)$$

By substituting $u = -\log(2\Phi(-x/\theta))$ and because $\operatorname{erf}(x/\sqrt{2}\,\theta) = 2\Phi(x/\theta) - 1$, equation (3.6) can be written as

$$I_m = \frac{\sqrt{2}\,\theta}{\Gamma(\alpha)\beta^\alpha} \int_0^{-\log(2\Phi(\frac{-m}{\theta}))} \operatorname{erf}^{-1}(1-e^{-u}) u^{\alpha-1} e^{\frac{-u}{\beta}}\,du$$

$$(3.7)$$

Using the series representation for $\operatorname{erf}^{-1}(1-e^{-u})$ (see Wolfram website), results in

$$\operatorname{erf}^{-1}(x) = \sum_{k=0}^\infty \frac{c_k}{2k+1} (\sqrt{\pi}/2)^{2k+1} x^{2k+1},$$

$$(3.8)$$

where $c_k = \sum_{m=0}^{k-1} \frac{c_m c_{k-1-m}}{(m+1)(2m+1)}$

and $c_0 = 1$.

Using (3.8), equation (3.7) reduces to

$$I_m = \frac{\sqrt{2}\theta}{\Gamma(\alpha)\beta^\alpha} \sum_{k=0}^\infty \frac{c_k}{2k+1} (\frac{\sqrt{\pi}}{2})^{2k+1}$$

$$\int_0^{-\log(2\Phi(\frac{-m}{\theta}))} (1-e^{-u})^{2k+1} u^{\alpha-1} e^{\frac{-u}{\beta}}\,du.$$

$$(3.9)$$

Using the series expansion of $(1-e^{-u})^{2k+1}$, (3.9) reduces to

$$I_m = \frac{\sqrt{2}\theta}{\Gamma(\alpha)} \sum_{k=0}^\infty \sum_{i=0}^{2k+1} a(k,i)\gamma(\alpha,\delta_i),$$

where

$$a(k,i) = \frac{c_k}{2k+1} (\sqrt{\pi}/2)^{2k+1} \binom{2k+1}{i} \frac{(-1)^i}{(1+i\beta)^\alpha}$$

and $\delta_i(m/\theta) = -(i+1/\beta)\log(2\Phi(-m/\theta))$.

The results follow by substituting I_m in $D(\mu)$ and $D(M)$ in (3.3).

Order Statistics

The density function of the rth order statistic, $X_{r:n}$, for a random sample of size n drawn from (1.3), is

$$f(x) = \frac{1}{B(r, n-r+1)} g(x)(G(x))^{r-1}(1-G(x))^{n-r}.$$

$$(4.1)$$

Using the binomial expansion, (4.1) can be written as

$$f_{r:n}(x) = \frac{g(x)}{B(r, n-r+1)}$$

$$\sum_{j=0}^{n-r} (-1)^j \binom{n-r}{j} (G(x))^{r+j-1}.$$

$$(4.2)$$

From (1.7), equation (4.2) can be written as

$$f_{r:n}(x) = \frac{g(x)}{B(r, n-r+1)} \sum_{j=0}^{n-r} \frac{(-1)^j}{\Gamma(\alpha)^{r+j-1}} \binom{n-r}{j}$$

$$\times \left(\sum_{k=0}^\infty \frac{(-1)^k}{\beta^{k+\alpha}} \frac{(-\log(2\Phi(-x/\theta)))^{\alpha+k}}{k!(k+\alpha)} \right)^{r+j-1}$$

$$= \frac{g(x)}{B(r, n-r+1)} \sum_{j=0}^{n-r} \sum_{k_1=0}^\infty \sum_{k_2=0}^\infty \cdots$$

$$\sum_{k_{r+j-1}=0}^\infty \frac{1}{\beta^{s_k+(r+j-1)\alpha} P_k} \frac{(-1)^{s_k+j}}{\Gamma(\alpha)^{r+j-1}} \binom{n-r}{j}$$

$$(-\log 2\Phi(\frac{-x}{\theta}))^{s_k+(r+j-1)\alpha}.$$

Table 1: Mode, Mean, Median and Variance for Some Values of α and β with $\theta = 1$

| α	β	Mode	Mean	Median	Variance
0.5	0.5	0	0.2482	0.1351	0.0887
	1	0	0.4347	0.2578	0.23895
	4	0	1.1785	0.8370	1.3385
	7	0	1.6911	1.2718	2.4929
	9	0	1.9761	1.5177	3.2649
0.9	0.5	0	0.4255	0.3292	0.1355
	1	0	0.7298	0.5969	0.3442
	4	0, 1.0373	1.8914	1.6854	1.7215
	7	0, 1.7998	2.6687	2.4242	3.0935
	9	0, 2.1818	3.0968	2.8303	3.9973
1	0.5	0	0.4674	0.3757	0.1449
	1	0	0.7979	0.6745	0.3634
	4	1.3744	2.0485	1.8627	1.7774
	7	2.1441	2.8811	2.6601	3.1744
	9	2.5481	3.3388	3.0973	4.0932
4	0.5	1.3439	1.4376	1.4069	0.2516
	1	2.1644	2.2701	2.2349	0.5257
	4	4.9410	5.1202	5.0607	2.0628
	7	6.7013	6.9392	6.8604	3.5516
	9	7.6885	7.9315	7.8419	4.5361
7	0.5	2.0642	2.1197	2.1012	0.2677
	1	3.1775	3.2457	3.2229	0.5355
	4	6.9108	7.0418	6.9982	2.0578
	7	8.0469	9.4615	9.4032	3.5509
	9	10.5824	10.7830	10.7164	4.5419

Using (1.4),

$$f_{r:n}(x) = \frac{1}{B(r,n-r+1)} \sum_{j=0}^{n-r} \sum_{k_1}^{\infty} \sum_{k_2}^{\infty} \cdots$$

$$\sum_{k_{r+j-1}}^{\infty} \frac{1}{\beta^{s_k+(r+j)\alpha} P_k} \frac{\Gamma(s_k+(r+j)\alpha)}{\Gamma(\alpha)^{r+j-1}}$$

$$\times \binom{n-r}{j} g(x \mid s_k+(r+j)\alpha, \beta, \theta),$$

(4.3)

where $s_k = \sum_{i=1}^{r-j+1} k_i$ and $P_k = \prod_{i=1}^{r-j+1} k_i(k_i+\alpha)$.

The result in (4.3) shows that the PDF of the rth order statistics of the gamma half-normal distribution can be expressed in terms of infinite sums of the gamma half-normal PDFs. Using the same technique as Ristic and Balakrishnan (2012), the asymptotic distribution of the sample minimum $X_{1:n}$ can be obtained by utilizing Theorem 8.3.6 in Arnold, et al. (1992), which states that if $\lim_{\varepsilon \to G^{-1}(0)} \dfrac{G(\varepsilon x)}{G(\varepsilon)} = x^{\gamma}$, then the asymptotic distribution of $X_{1:n}$ will be of Weibull type with shape parameter γ.

Figure 5: The Gamma-Half Normal Skewness and Kurtosis Graphs for Various Values of α and β when $\theta = 1$
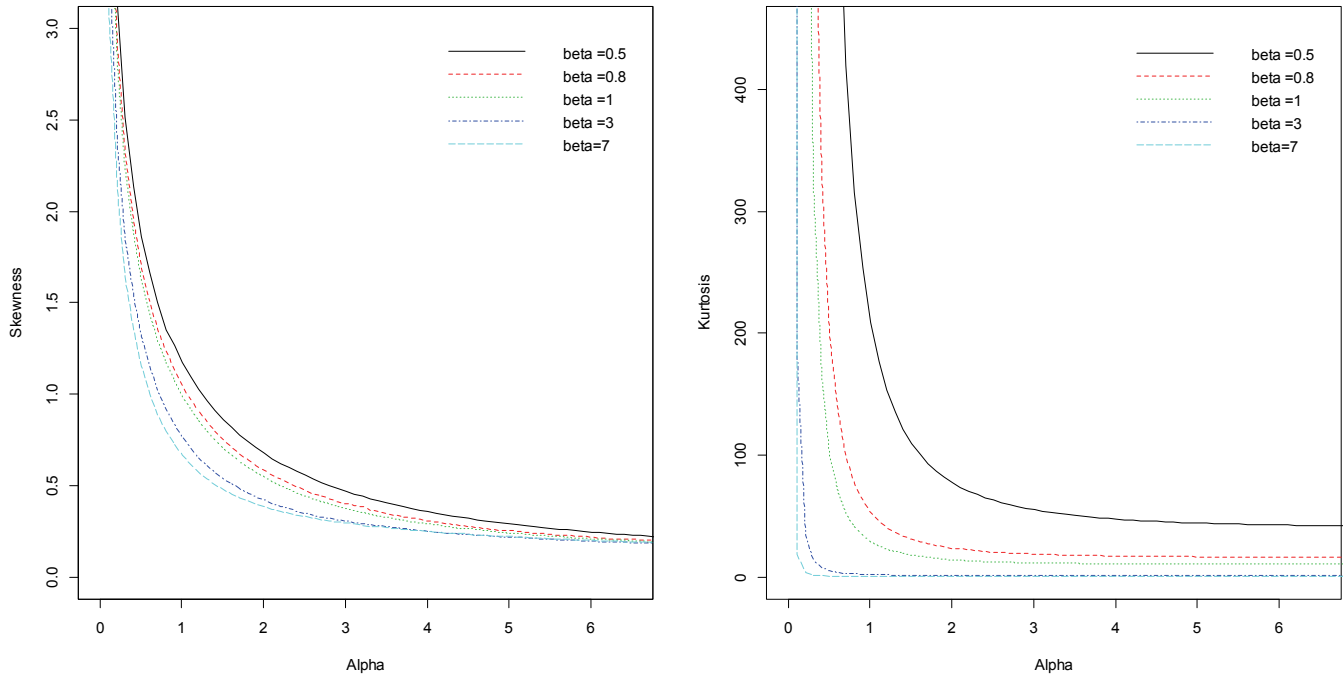
For the gamma half-normal distribution, $G^{-1}(0) = 0$ and

$$\lim_{\varepsilon \to 0^+} \frac{G(\varepsilon x)}{G(\varepsilon)} = x \lim_{\varepsilon \to 0^+} \frac{g(\varepsilon x)}{g(\varepsilon)}$$

$$= x \left[\lim_{\varepsilon \to 0^+} \frac{\log(2\Phi(-\varepsilon x / \theta))}{\log(2\Phi(-\varepsilon / \theta))} \right]^{\alpha-1}$$

$$= x \left[x \lim_{\varepsilon \to 0^+} \left(\frac{\phi(-\varepsilon x / \theta)}{\phi(-\varepsilon / \theta)} \right) \right]^{\alpha-1} = x^\alpha.$$

(4.4)

Hence, the asymptotic distribution $X_{r:n}$ is of Weibull type with shape parameter α. The asymptotic distribution of the sample maximum $X_{n:n}$ can be viewed as $G_n(x)$, where $G_n(x) = 1 - G_1(-x)$ and G_1 is the CDF of $X_{1:n}$.

Parameter Estimation

Let a random sample of size n be taken from the gamma-half normal distribution. The log-likelihood function for the gamma-half normal distribution in (1.4) is given by

$$\log L(\alpha, \beta) =$$

$$\frac{n}{2} \log 2 - n \log \theta - \frac{n}{2} \log \pi$$

$$- n \log \Gamma(\alpha) - \alpha n \log \beta - \frac{1}{2\theta^2} \sum_{i=1}^{n} x_i^2$$

$$+ (\alpha - 1) \sum_{i=1}^{n} \log(-\log(2\Phi(-x_i / \theta)))$$

$$+ (\frac{1}{\beta} - 1) \sum_{i=1}^{n} \log(2\Phi(-x_i / \theta).$$

(5.1)

The derivatives of (5.1) with respect to α, β and θ respectively, are given by

$$\frac{\partial \log L}{\partial \alpha} =$$

$$-n\psi(\alpha) - n \log \beta + \sum_{i=1}^{n} \log(-\log(2\Phi(-x_i / \theta))),$$

(5.2)

$$\frac{\partial \log L}{\partial \beta} = -\frac{\alpha n}{\beta} - \frac{1}{\beta^2} \sum_{i=1}^{n} \log(2\Phi(-x_i / \theta)),$$

(5.3)

$$\frac{\partial \log L}{\partial \theta} =$$

$$-\frac{n}{\theta} + \frac{1}{\theta^3} \sum_{i=1}^{n} x_i^2$$

$$+ (\alpha - 1) \sum_{i=1}^{n} \frac{x_i h_z(x_i / \theta)}{\log(2\Phi(-x_i / \theta))}$$

$$- (\frac{1}{\beta} - 1) \sum_{i=1}^{n} x_i h_z(x_i / \theta)$$

(5.4)

Setting (5.2), (5.3) and (5.4) to zero and solving them simultaneously results in $\hat{\alpha}, \hat{\beta}$ and $\hat{\theta}$.

The initial values for the parameters α, β and θ can be obtained by assuming the random sample, x_i, $i = 1, \ldots, n$ is taken from the half-normal distribution with parameter θ. By equating the population mean to the sample mean of x_i, $i = 1, \ldots, n$ and solving for θ, the initial value $\hat{\theta} = \sqrt{\pi / 2}\, \bar{x}$. Assuming $y_i = -\log(2\Phi(-x_i / \hat{\theta}))$, $i = 1, \ldots, n$ are taken from the gamma distribution with parameters α and β (see Lemma 1). By equating the population mean and the population variance of gamma distribution (with parameters α and β) to the corresponding sample mean and sample variance of y_i, $i = 1, \ldots, n$ and solving for α and β, the initial values are $\alpha_0 = \bar{y}^2 / s_y^2$ and $\beta_0 = s_y^2 / \bar{y}$, where $\bar{y}$ and s_y^2 are the sample mean and the sample variance for y_1, $y_2, \ldots, y_n$.

When $\alpha = \beta = 1$, the gamma-half normal distribution reduces to the half-normal distribution; thus, the likelihood ratio test can be used to determine whether the gamma-half normal distribution or the half-normal distribution is the best model for fitting a given data set. The likelihood ratio test can be used for testing the hypothesis $H_0 : \alpha = \beta = 1$ against $H_a : \alpha \neq 1$ or $\beta \neq 1$, which is based on $\lambda = L_0(\tilde{\theta}) / L_a(\hat{\alpha}, \hat{\beta}, \hat{\theta})$, where L_0 and L_a are the likelihood functions for the half-normal and the gamma-half normal distributions, respectively. The quantity $-2 \log \lambda$ follows the Chi-square distribution with 2 degrees of freedom asymptotically.

Application

Three data sets were applied to the gamma-half normal distribution, and compared with the half-normal, generalized half-normal, beta generalized half-normal and inverse Gaussian distributions. The first two data sets (see Tables 2 and 4), were analyzed by Raqab, et al. (2008). This data represents the tensile strength data measured in GPa for single-carbon fibers that were tested at gauge lengths of 20 mm and 10 mm. The third data set (see Table 6) was analyzed by Cheng, et al. (1981) and represents the flood level for the Susquehanna River at Harrisburg, PA. The maximum likelihood estimates, KS (Kolmogorov-Smirnov) test-statistics and *p*-values for the fitted distributions are reported in Tables 3, 5 and 7.

The data in Tables 2 and 4 are fitted to the gamma-half normal, half-normal, generalized half-normal and beta generalized-half normal distributions. The half-normal distribution did not produce an adequate model for the data. However, the generalized half-normal, beta generalized and gamma-half normal each provide a good fit for the two data sets. Among the three generalizations of the half-normal distribution, the gamma-half normal provides the best fit for the first data set, and generalized half-normal provides the best fit for the second. When graphing the first data set, an approximately symmetric distribution is obtained. The distribution of the second data set, however, is a right-skewed shape. This suggests

that the gamma-half normal distribution is able to model data of both approximately symmetric and right-skewed shapes. Figures 6 and 7 display the empirical and fitted cumulative distribution functions; these figures support the results in Tables 3 and 5, respectively.

Table 2: Single Carbon Fibers at 20 mm

0.312	0.314	0.479	0.552
0.700	0.803	0.861	0.865
0.944	0.958	0.966	0.997
1.006	1.021	1.027	1.055
1.063	1.098	1.140	1.179
1.224	1.240	1.253	1.270
1.272	1.274	1.301	1.301
1.359	1.382	1.382	1.426
1.434	1.435	1.478	1.490
1.511	1.514	1.535	1.554
1.566	1.570	1.586	1.629
1.633	1.642	1.648	1.684
1.697	1.726	1.770	1.773
1.800	1.809	1.818	1.821
1.848	1.880	1.954	2.012
2.067	2.084	2.090	2.096
2.128	2.233	2.433	2.585
2.585			

The third data set (see Table 6) was analyzed by Cheng, et al. (1981) and fitted to the inverse Gaussian distribution. These results, as well as the comparisons made to the half-normal, beta generalized half-normal and gamma-half normal distributions, are reported in Table 7. The generalized half-normal distribution was divergent for the third data set. In view of these results, the gamma half-normal and inverse Gaussian distributions give a moderate fit to the data. The half-normal distribution does not give an adequate fit to the data, while the generalized half-normal provided the best fit. In viewing the distribution of the third data set, another right-skewed distribution is observed. This confirms the fact that the gamma-half normal distribution can be used to fit data of a right-skewed shape. Figure 8 displays the empirical and fitted cumulative distribution functions.

Table 3: Parameter Estimates for Single Carbon Fibers at 20 mm

Distribution	Half-Normal	Generalized Half-Normal	Beta Generalized Half-Normal	Gamma-Half Normal
Parameter Estimates	$\hat{\theta} = 1.5323$	$\hat{\alpha} = 2.2823$ $\hat{\theta} = 1.5879$ $\hat{\mu} = 0.1220$	$\hat{a} = 1.3742$ $\hat{b} = 0.2369$ $\hat{\alpha} = 1.8610$ $\hat{\theta} = 0.9766$	$\hat{\alpha} = 2.8794$ $\hat{\beta} = 3.1725$ $\hat{\theta} = 0.3934$
KS	0.3317	0.0548	0.0863	0.0425
P-value	0.0000	0.9857	0.9985	0.9996

Conclusion

The gamma-half normal distribution, a new generalization of the half-normal distribution, was derived using the method proposed by Alzaatreh, et al. (2013b). Various properties of the distribution were studied including the moments, mean deviations from the mean and median, hazard function, modality and Shannon entropy. The maximum likelihood method was proposed for the estimation of the gamma-half normal parameters. In order to demonstrate the applicability of the gamma-half normal distribution it was fitted to three real data sets and compared with the half-normal, generalized-half normal, inverse Gaussian and beta generalized half-normal distributions.

Results show that the gamma-half normal distribution provides an adequate fit for each data set. Because the distribution was fitted to data sets with right-skewed and approximately symmetric shapes, this indicates that the gamma-half normal distribution offers flexibility that extends beyond the half-normal distribution. Although the gamma-half normal distribution can be bimodal, it was difficult to find data in the literature with the specific form of bimodality. The maximum likelihood functions may be further studied under different types of censoring for future applications of the gamma-half normal distribution.

References

Alzaatreh, A., Famoye, F., & Lee, C. (2013a).Weibull-Pareto distribution and its applications. *Communications in Statistics: Theory & Methods*, 42(9), 1673-1691.

Alzaatreh, A., Lee, C., & Famoye, F. (2013b). A new method for generating families of continuous distributions. To Appear *Metron: International Journal of Statistics*.

Alzaatreh, A., Famoye, F., & Lee, C. (2012a). Gamma-Pareto distribution and its applications. *Journal of Modern Applied Statistical Methods*, *11*(*1*), 78-94.

Alzaatreh, A., Lee, C., & Famoye, F. (2012b). On the discrete analogues of continuous distributions. *Statistical Methodology*, *9*, 589-603.

Arnold, B. C., Balakrish, N., & Nagaraja, H. N. (1992). *A first course in order statistics*. New York, NY: Wiley.

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, *8*, 135-160.

Bland, J. M. (2005). *The half-normal distribution method for measurement error: two case studies*. Unpublished lecture available on http://wwwusers.york.ac.uk/~mb55/talks/halfnor .pdf.

Castro, L., Gomez, H., & Valenzuela, M. (2012). Epsilon half-normal model: properties and inference. *Computational Statistics & Data Analysis*, *56*(*12*), 4338-4347.

Cheng, R. C. H., & Amin, N. A. K., (1981). Maximum likelihood estimation of parameters in the inverse Gaussian distribution, with unknown origin. *Technometrics*, *23*, 257-263.

Cordeiro, G., Pescim, R., & Ortega, E. (2012). The Kumaraswamy generalized half-normal distribution for skewed positive data. *Journal of Data Science*, *10*, 195-224.

Cooray, K., & Anada, M. (2008). A Generalization of the Half-Normal distribution with applications to lifetime data. *Communications in Statistics-Theory and Methods*, *37*, 1323-1337.

Lee, C., Famoye, F. & Alzaatreh, A. (2013). Methods for generating families of continuous distribution. To Appear in WIREs Computational Statistics.

Nadarajah, S., & Pal, M. (2008). Explicit expressions for moments of gamma order statistics. *Bulletin of the Brazilian Mathematical Society*, *39*(*1*), 45-60.

Pescrim, R., et al. (2010). The beta generalized half- normal distribution. *Computational Statistics and Data Analysis*, *54*, 945-957.

Raqab, M., Madi, M., & Debasis, K. (2008). Estimation of $P(Y<X)$ for the 3-parameter generalized exponential distribution. *Communications in Statistics-Theory and Methods*, *37*(*18*), 2854-2864.

Renyi, A. (1961). On measures of entropy and information, *In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, *I*, 547-561. Berkeley, CA: University of California Press.

Ristic, M., & Balakrishnan, N. (2012). The gamma-exponentiated exponential distribution. *Journal of Statistical Computation and Simulation*, *82*(*8*), 1191-1206.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379-432.

Wolfram. http://functions.wolfram.com/ GammaBetaErf/InverseErf/06/01/02/0004/, retrieved on December 04, 2012.

Zografos, K., & Balakrishnan, N. (2009). On families of beta and generalized gamma-generated distributions and associated inference. *Statistical Methodology*, *6*, 344-362.

116

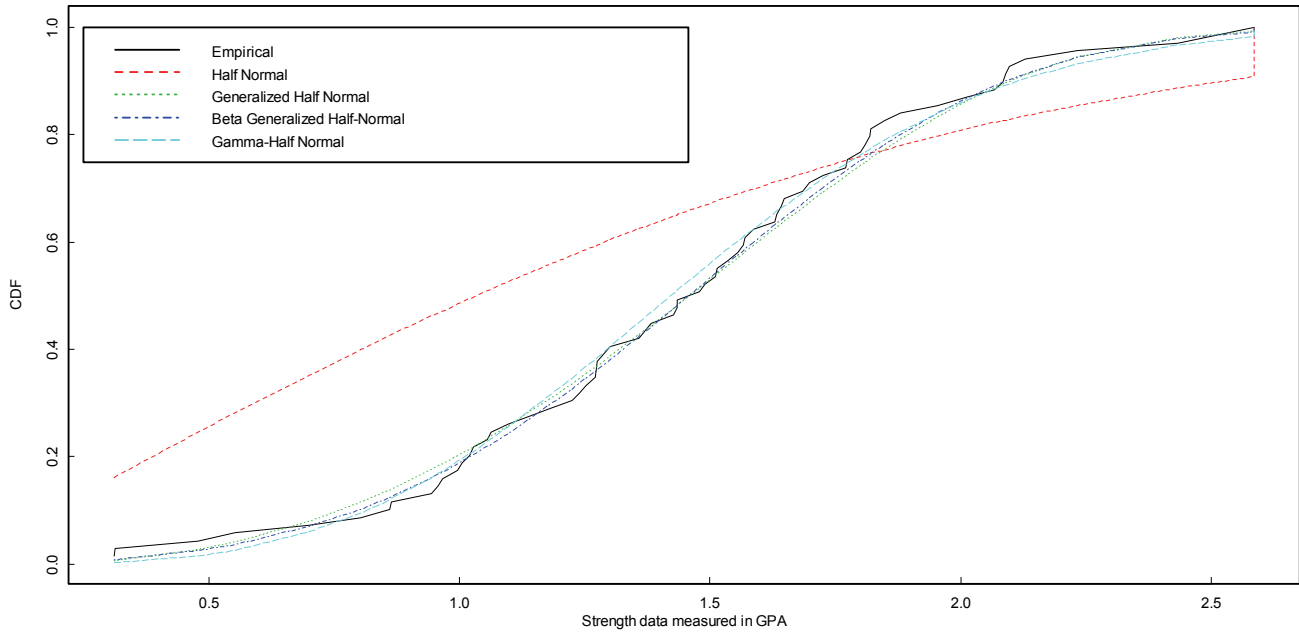Figure 6: CDF for Fitted Distributions for Gauge Length of 20mm Data

Table 4: Single Carbon Fibers at 10 mm

0.101	0.332	0.403	0.428	0.457	0.550	0.561	0.596
0.597	0.645	0.654	0.674	0.718	0.722	0.725	0.732
0.775	0.814	0.816	0.818	0.824	0.859	0.875	0.938
0.940	1.056	1.117	1.128	1.137	1.137	1.177	1.196
1.230	1.325	1.339	1.345	1.420	1.423	1.435	1.443
1.464	1.472	1.494	1.532	1.546	1.577	1.608	1.635
1.693	1.701	1.737	1.754	1.762	1.828	2.052	2.071
2.086	2.171	2.224	2.227	2.425	2.595	3.220	

Table 5: Parameter Estimates for Single Carbon Fibers at 10mm

Distribution	Half-Normal	Generalized Half-Normal	Beta Generalized Half-Normal	Gamma-Half Normal
Parameter Estimates	$\hat{\theta} = 1.4019$	$\hat{\alpha} = 1.5347$ $\hat{\theta} = 1.4798$ $\hat{\mu} = 0.0699$	$\hat{a} = 1.9544$ $\hat{b} = 0.1522$ $\hat{\alpha} = 1.0954$ $\hat{\theta} = 0.4722$	$\hat{\alpha} = 2.4260$ $\hat{\beta} = 0.4216$ $\hat{\theta} = 1.4529$
KS	0.2099	0.0606	0.0863	0.0678
P-value	0.0078	0.9748	0.7361	0.9342

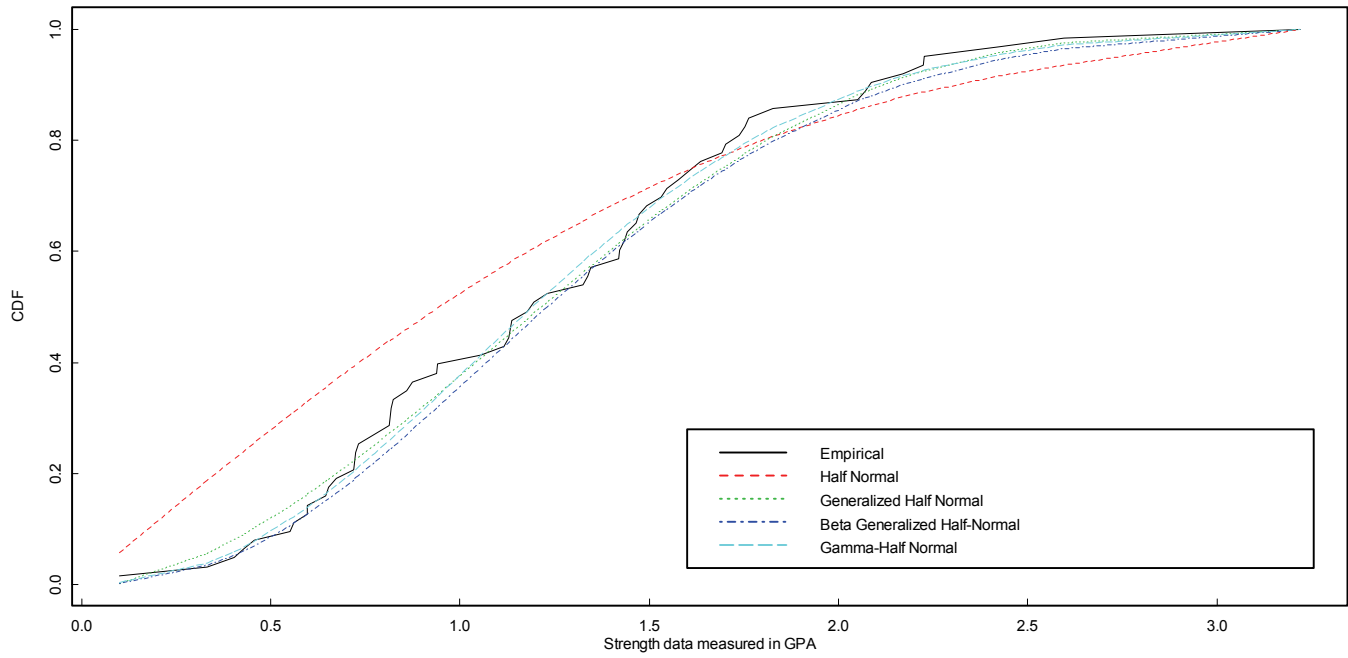Figure 7: CDF for Fitted Distributions for Gauge Lengths of 10mm Data

Table 6: Maximum Flood Level for Susquehanna River

.654	.613	.402	.379
.269	.740	.416	.338
.315	.449	.297	.423
.379	.324	.418	.412
.494	.392	.484	.265

Table 7: Parameter estimates for maximum flood levels

Distribution	Half-Normal	Generalized Half-Normal	Beta Generalized Half-Normal	Gamma-Half Normal
Parameter Estimates	$\hat{\theta} = 0.4404$	Divergent	$\hat{a} = 244.48$ $\hat{b} = 176.11$ $\hat{\alpha} = 0.1901$ $\hat{\theta} = 1.2483$	$\hat{\alpha} = 0.1780$ $\hat{\beta} = 0.245$ $\hat{\lambda} = 0.9140$
KS	0.4026		0.1507	0.1026
P-value	0.0031		0.7540	0.5638

Figure 8: CDF for Fitted Distributions for Maximum Flood Level