ARTICLE IN PRESS

Computational Statistics and Data Analysis ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

The beta generalized half-normal distribution

Rodrigo R. Pescim [a], Clarice G.B. Demétrio [a], Gauss M. Cordeiro [b], Edwin M.M. Ortega [a,*], Mariana R. Urbano [a]

[a] *ESALQ, Universidade de São Paulo-USP, Piracicaba, Brazil*

[b] *Departamento de Estatística e Informática, Universidade Federal Rural de Pernambuco-UFRPE, Recife, Brazil*

ARTICLE INFO

Article history:
Received 23 April 2009
Received in revised form 24 August 2009
Accepted 7 October 2009
Available online xxxx

ABSTRACT

For the first time, we propose the so-called beta generalized half-normal distribution, which contains some important distributions as special cases, such as the half-normal and generalized half-normal (Cooray and Ananda, 2008) distributions. We derive expansions for the cumulative distribution and density functions which do not depend on complicated functions. We obtain formal expressions for the moments of the new distribution. We examine the maximum likelihood estimation of the parameters and provide the expected information matrix. The usefulness of the new distribution is illustrated through a real data set by showing that it is quite flexible in analyzing positive data instead of the generalized half-normal, half-normal, Weibull and beta Weibull distributions.

1. Introduction

Fatigue is a structural damage which occurs when a material is exposed to stress and tension fluctuations. Statistical models allow us to study the random variation of the failure time associated with materials exposed to fatigue as a result of different cyclical patterns and strengths. The most popular models used to describe the lifetime process under fatigue are the half-normal (HN) and Birnbaum–Saunders (BS) distributions. When modeling monotone hazard rates, the HN and BS distributions may be an initial choice because of its negatively and positively skewed density shapes. However, they do not provide a reasonable parametric fit for modeling phenomenon with non-monotone failure rates such as the bathtub shaped and the unimodal failure rates, which are common in reliability and biological studies. Such bathtub hazard curves have nearly flat middle portions and the corresponding densities have a positive anti-mode. According to Nelson (1990, p. 27), the distributions which permit a bathtub fit are sufficiently complex. On the other hand, more flexible distributions usually require five or more parameters. However, more recently, Díaz-García and Leiva (2005) proposed a new family of generalized Birnbaum–Saunders distributions based on contoured elliptical distributions and Cooray and Ananda (2008) proposed the generalized half-normal (GHN) distribution derived from a model for static fatigue.

One major benefit of the beta generalized class of distributions is its ability of fitting skewed data that cannot be properly fitted by existing distributions. Starting from a parent cumulative distribution function (cdf) $G(x)$, Eugene et al. (2002) defined a class of generalized distributions by

$$F(x) = I_{G(x)}(a, b) = \frac{1}{B(a, b)} \int_0^{G(x)} \omega^{a-1}(1-\omega)^{b-1} \mathrm{d}\omega \qquad (1)$$

* Corresponding address: Departamento de Ciências Exatas, USP, Av. Pádua Dias 11 - Caixa Postal 9, 13418-900 Piracicaba - São Paulo - Brazil. Tel.: +55 11 19 34294127; fax: +55 11 19 34294468.

E-mail addresses: rrpescim@esalq.usp.br (R.R. Pescim), clarice@esalq.usp.br (C.G.B. Demétrio), gausscordeiro@uol.com.br (G.M. Cordeiro), edwin@esalq.usp.br (E.M.M. Ortega), mrurbano@esalq.usp.br (M.R. Urbano).

0167-9473/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2009.10.007

for extra parameters $a > 0$ and $b > 0$, where $I_y(a, b) = B_y(a, b)/B(a, b)$ is the incomplete beta function ratio, $B_y(a, b) = \int_0^y w^{a-1}(1-w)^{b-1}\mathrm{d}w$ is the incomplete beta function and $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ is the beta function. This class of generalized distributions has been receiving considerable attention over the last years, in particular after the works of Eugene et al. (2002) and Jones (2004).

We can express (1) in terms of the hypergeometric function (Gradshteyn and Ryzhik, 2000, Section 9.1), since the properties of the hypergeometric function are well established in the literature. We have

$$F(x) = \frac{G(x)}{a\,B(a, b)}\,{}_2F_1(a, 1-b, a+1; G(x)). \tag{2}$$

The probability density function (pdf) corresponding to (1) can be written as

$$f(x) = \frac{g(x)}{B(a, b)} G(x)^{a-1}\{1-G(x)\}^{b-1}, \tag{3}$$

where $g(x) = \mathrm{d}G(x)/\mathrm{d}x$ is the parent density function. The pdf $f(x)$ will be most tractable when both functions $G(x)$ and $g(x)$ have simple analytic expressions. Except for some special choices of these functions, the density $f(x)$ will be difficult to deal with some generality.

Eugene et al. (2002) defined the beta normal (BN) distribution by taking $G(x)$ to be the cdf of the normal distribution and derived some of its first moments. Nadarajah and Kotz (2004) introduced the beta Gumbel (BG) distribution by taking $G(x)$ to be the cdf of the Gumbel distribution and provided expressions for the moments, the asymptotic distribution of the extreme order statistics, and examined the maximum likelihood estimation procedure. Nadarajah and Gupta (2004) introduced the beta Fréchet (BF) distribution by taking $G(x)$ to be the Fréchet distribution, derived the analytical shapes of the density and hazard rate functions and the asymptotic distribution of the extreme order statistics. Further, Nadarajah and Kotz (2005) examined the beta exponential (BE) distribution and obtained the moment generating function, the first four moments, the asymptotic distribution of the extreme order statistics and discussed maximum likelihood estimation. Lee et al. (2007) defined the beta Weibull (BW) distribution and showed that it can have bathtub, unimodal, increasing and decreasing hazard rate functions.

Other authors have introduced new distributions for modeling bathtub shaped failure rate. For example, Rajarshi and Rajarshi (1988) presented a review of these distributions and Haupt and Schabe (1992) considered a lifetime model with bathtub failure rates. However, these models do not present much practicability to be used. Alternative new classes of distributions were proposed in the last years based on modifications of the Weibull distribution to cope with bathtub shaped failure rate. Pham and Lai (2007) provided a good review for some of these models. They presented the exponentiated Weibull (EW) distribution (Mudholkar et al., 1995, 1996), the additive Weibull distribution (Xie and Lai, 1995), the extended Weibull distribution (Xie et al., 2002), the modified Weibull (MW) distribution (Lai et al., 2003), the beta exponential (BE) distribution (Nadarajah and Kotz, 2005), the extended flexible Weibull distribution (Bebbington et al., 2007) and the generalized modified Weibull (GMW) (Carrasco et al., 2008).

In this note, we introduce a new four-parameter distribution, so-called the beta generalized half-normal (BGHN) distribution, which contains as sub-models the half-normal (HN) and generalized half-normal (GHN) (Cooray and Ananda, 2008). The new distribution due to its flexibility in accommodating all the forms of the risk function is an important model to be used in a variety of problems in survival analysis. The BGHN distribution is not only convenient for modeling comfortable bathtub shaped failure rates but it is also suitable for testing goodness of fit of its special sub-models.

The paper is organized as follows. In Section 2, we define the BGHN distribution, present some special cases and provide expansions for its distribution and density functions. Section 3 gives a general expansion for the moments and the moment generating function. In Section 4, we obtain expansions for the moments of the order statistics. Maximum likelihood estimation is discussed in Section 5. Section 6 illustrates the importance of the BGHN distribution applied to a real data set. Finally, concluding remarks are given in Section 7.

2. Beta generalized half-normal distribution

The GHN density with shape parameter $\alpha > 0$ and scale parameter $\theta > 0$ has the form (Cooray and Ananda, 2008)

$$g(x) = \sqrt{\frac{2}{\pi}}\left(\frac{\alpha}{x}\right)\left(\frac{x}{\theta}\right)^{\alpha}\exp\left[-\frac{1}{2}\left(\frac{x}{\theta}\right)^{2\alpha}\right], \quad x > 0. \tag{4}$$

Its cumulative function depends on the error function

$$G(x) = 2\Phi\left[\left(\frac{x}{\theta}\right)^{\alpha}\right] - 1 = \mathrm{erf}\left(\frac{\left(\frac{x}{\theta}\right)^{\alpha}}{\sqrt{2}}\right), \tag{5}$$

where

$$\Phi(x) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)\right] \quad \text{and} \quad \mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x \exp(-t^2)\mathrm{d}t.$$

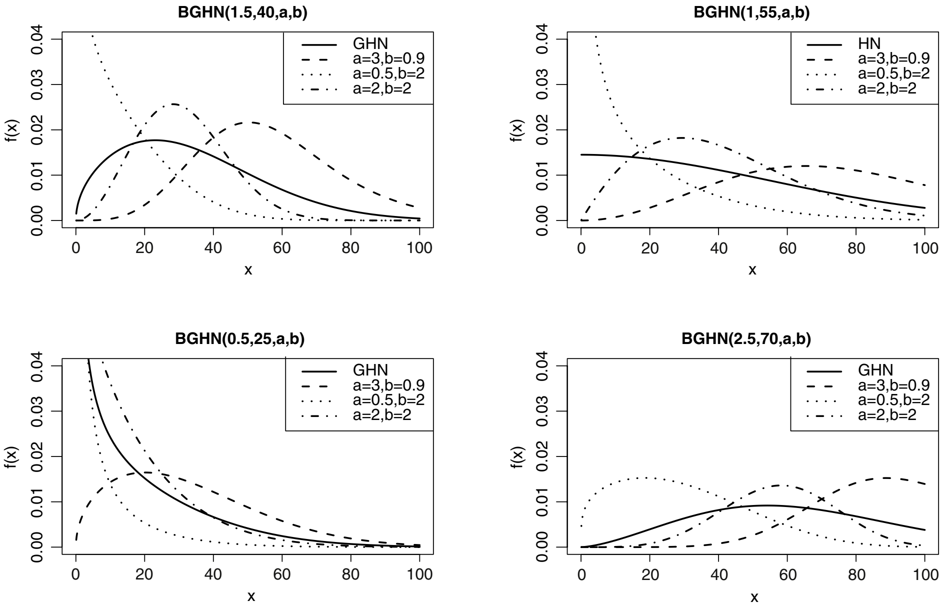

Fig. 1. Plots of the density function (7) for some parameter values.

The HN distribution is obtained when $\alpha = 1$ and its nth moment is $\mu'_n = \sqrt{\frac{2^{\frac{n}{\alpha}}}{\pi}}\,\Gamma\left(\frac{n+\alpha}{2\alpha}\right)\theta^n$, where $\Gamma(.)$ is the gamma function (see Cooray and Ananda, 2008).

We now introduce the BGHN distribution with four parameters $\alpha > 0$, $\theta > 0$, $a > 0$ and $b > 0$ by taking $G(x)$ in Eq. (1) to be the cdf (5) of the GHN distribution. The general form for the BGHN cumulative function is given by

$$F(x) = \frac{1}{B(a,b)} \int_0^{2\Phi\left[\left(\frac{x}{\theta}\right)^\alpha\right]-1} \omega^{a-1}(1-\omega)^{b-1}\mathrm{d}\omega. \tag{6}$$

Inserting (4) and (5) in Eq. (3) yields the BGHN density function (for $x > 0$)

$$f(x) = \frac{\sqrt{\frac{2}{\pi}}\left(\frac{\alpha}{x}\right)\left(\frac{x}{\theta}\right)^\alpha \exp\left[-\frac{1}{2}\left(\frac{x}{\theta}\right)^{2\alpha}\right]}{B(a,b)} 2^{b-1}\left\{2\Phi\left[\left(\frac{x}{\theta}\right)^\alpha\right]-1\right\}^{a-1}\left\{1-\Phi\left[\left(\frac{x}{\theta}\right)^\alpha\right]\right\}^{b-1}. \tag{7}$$

The BGHN distribution contains as special sub-models well-known distributions. It simplifies to the GHN distribution when $a = b = 1$. If $\alpha = 1$, it gives the beta half-normal (BHN) distribution. If $b = 1$, it leads to the exponentiated generalized half-normal (EGHN) distribution. Further, if $a = b = 1$, in addition to $\alpha = 1$, it reduces to the HN distribution. The hazard rate function corresponding to (7) becomes

$$h(x) = \frac{\sqrt{\frac{2}{\pi}}\left(\frac{\alpha}{x}\right)\left(\frac{x}{\theta}\right)^\alpha e^{-\frac{1}{2}\left(\frac{x}{\theta}\right)^{2\alpha}} 2^{b-1}\left\{2\Phi\left[\left(\frac{x}{\theta}\right)^\alpha\right]-1\right\}^{a-1}\left\{1-\Phi\left[\left(\frac{x}{\theta}\right)^\alpha\right]\right\}^{b-1}}{B(a,b)\left\{1-I_{2\Phi\left[\left(\frac{x}{\theta}\right)^\alpha\right]-1}(a,b)\right\}}. \tag{8}$$

If X is a random variable with density (7), we write $X \sim \mathrm{BGHN}(\alpha, \theta, a, b)$. The BGHN distribution is easily simulated from $F(x)$ as follows: if V has a beta distribution with parameters a and b, then the solution of the nonlinear equation $(\frac{X}{\theta})^\alpha = \Phi^{-1}(\frac{V+1}{2})$ has the BGHN (α, θ, a, b) distribution. To simulate data from this nonlinear equation, we can use the matrix programming language Ox through *SolveNLE* subroutine (see Doornik, 2007).

Plots of the density and hazard rate functions of the BGHN distribution are presented in Figs. 1 and 2, respectively, for selected parameter values, including the special cases of the GHN and HN distributions.

We provide two simple formulae for the cdf of the BGHN distribution depending if the parameter $b > 0$ is real non-integer or integer. First, if $|z| < 1$ and $b > 0$ is real non-integer, it follows the series (Nadarajah and Kotz, 2004, Eq. (1.7))

$$(1-z)^{b-1} = \sum_{j=0}^{\infty} \frac{(-1)^j \Gamma(b)}{\Gamma(b-j)j!} z^j. \tag{9}$$

Using the representation (9), the BGHN cumulative function (1) for $b > 0$ real non-integer can be expanded as

$$F(x) = \frac{\Gamma(b)}{B(a,b)} \sum_{j=0}^{\infty} \frac{(-1)^j \{2\Phi\left[\left(\frac{x}{\theta}\right)^\alpha\right]-1\}^{a+j}}{\Gamma(b-j)j!(a+j)}. \tag{10}$$

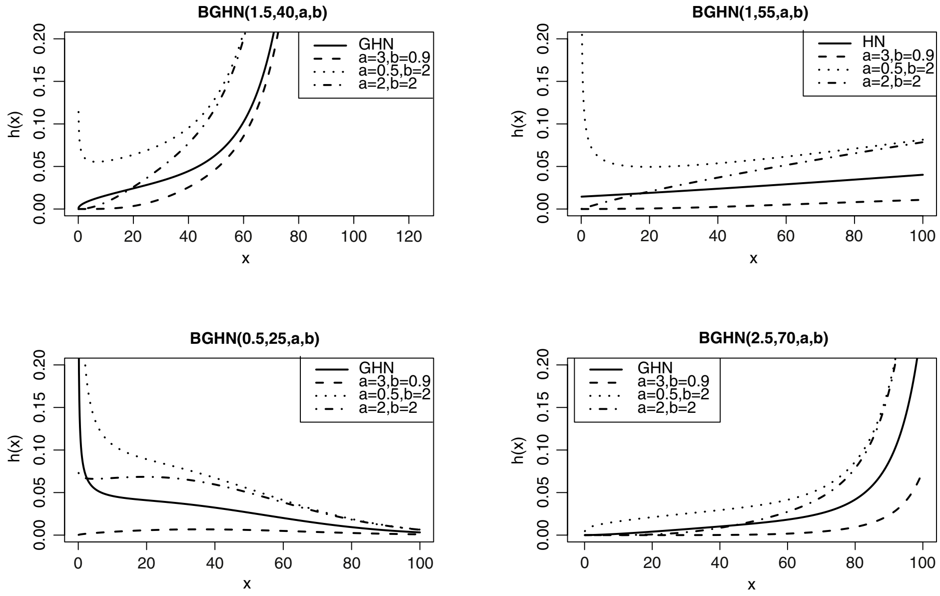

Fig. 2. Plots of the hazard rate function (8) for some parameter values.

If $a > 0$ is an integer, Eq. (10) gives the cdf of the BGHN distribution in terms of an infinite weighted power series of the GHN cumulative function. Otherwise, if $a > 0$ is real non-integer, the binomial expansion in Eq. (10) yields

$$F(x) = \frac{\Gamma(b)}{B(a,b)} \sum_{j,r=0}^{\infty} \sum_{s=0}^{r} \frac{(-1)^{j+r+s} 2^r \Gamma(a+j+1)}{\Gamma(b-j) j! (a+j)} \binom{r}{s} \left\{ \Phi\left[\left(\frac{x}{\theta}\right)^{\alpha}\right] \right\}^{s}. \tag{11}$$

Eq. (11) shows that for both b and a real non-integers, the cdf of the BGHN distribution can be expressed as an infinite weighted sum of powers of the cdf of the normal $N(0, 1)$ distribution.

By application of the binomial expansion in Eq. (1), when $b > 0$ is an integer, we obtain

$$F(x) = \frac{1}{B(a,b)} \sum_{j=0}^{b-1} \binom{b-1}{j} \frac{(-1)^j}{a+j} \left\{ 2\Phi\left[\left(\frac{x}{\theta}\right)^{\alpha}\right] - 1 \right\}^{a+j}. \tag{12}$$

For $a > 0$ integer, using the binomial expansion in Eq. (12), gives

$$F(x) = \frac{1}{B(a,b)} \sum_{j=0}^{b-1} \sum_{r=0}^{a+j} \frac{2^{a+j-r}(-1)^{j+r}}{(a+j)} \binom{b-1}{j} \binom{a+j}{r} \left\{ \Phi\left[\left(\frac{x}{\theta}\right)^{\alpha}\right] \right\}^{a+j-r}. \tag{13}$$

For $a > 0$ real non-integer, using (9) and the binomial expansion, Eq. (12) can be rewritten as

$$F(x) = \frac{1}{B(a,b)} \sum_{j=0}^{b-1} \sum_{r=0}^{\infty} \sum_{s=0}^{r} \frac{(-1)^{j+r+s} 2^r \Gamma(a+j+1)}{(a+j)\Gamma(a+j+1-r) r!} \binom{b-1}{j} \binom{r}{s} \left\{ \Phi\left[\left(\frac{x}{\theta}\right)^{\alpha}\right] \right\}^{s}. \tag{14}$$

The cdf (5) of the GHN distribution comes from Eq. (12) when $a = b = 1$. Further, if $a = b = \alpha = 1$, Eq. (12) yields the cdf of the HN distribution.

It can be seen in the Wolfram Functions Site[1] that for integer a and for integer b, we have, respectively,

$$I_y(a,b) = 1 - \frac{(1-y)^b}{\Gamma(b)} \sum_{j=0}^{a-1} \frac{\Gamma(b+j)}{j!} y^j, \qquad I_y(a,b) = \frac{y^a}{\Gamma(a)} \sum_{j=0}^{b-1} \frac{\Gamma(a+j)}{j!} (1-y)^j.$$

Hence, for $a > 0$ integer, we can obtain

$$F(x) = 1 - \frac{2^b \left\{1 - \Phi\left[\left(\frac{x}{\theta}\right)^{\alpha}\right]\right\}^b}{\Gamma(b)} \sum_{j=0}^{a-1} \frac{\Gamma(b+j)}{j!} \left\{ 2\Phi\left[\left(\frac{x}{\theta}\right)^{\alpha}\right] - 1 \right\}^{j},$$

[1] http://functions.wolfram.com/

Please cite this article in press as: Pescim, R.R., et al., The beta generalized half-normal distribution. Computational Statistics and Data Analysis (2009), doi:10.1016/j.csda.2009.10.007

whereas for $b > 0$ integer, we have an alternative form for Eq. (12) given by

$$F(x) = \frac{\left\{2\Phi\left[\left(\frac{x}{\theta}\right)^{\alpha}\right] - 1\right\}^{a}}{\Gamma(a)} \sum_{j=0}^{b-1} \frac{\Gamma(a+j)}{j!} 2^{j} \left\{1 - \Phi\left[\left(\frac{x}{\theta}\right)^{\alpha}\right]\right\}^{j}.$$

Eqs. (10)–(14) are the main expansions for the cdf of the BGHN distribution. They (and other expansions in the paper) can be evaluated in symbolic computation software such as Mathematica and Maple. These symbolic software have currently the ability to deal with analytic expressions of formidable size and complexity.

Alternatively to Eq. (7), we can obtain the BGHN density function for b real non-integer by differentiating (10) and using the representation (9). Thus,

$$f(x) = \frac{\sqrt{\frac{2}{\pi}} \left(\frac{\alpha}{x}\right) \left(\frac{x}{\theta}\right)^{\alpha} \mathrm{e}^{-\frac{1}{2}\left(\frac{x}{\theta}\right)^{2\alpha}}}{B(a,b)} \sum_{j,k=0}^{\infty} \sum_{l=0}^{k} w_{j,k,l}(a,b) \left\{\Phi\left[\left(\frac{x}{\theta}\right)^{\alpha}\right]\right\}^{l}, \tag{15}$$

where the coefficients $w_{j,k,l}(a,b)$ are given by

$$w_{j,k,l}(a,b) = \frac{\Gamma(b)\Gamma(a+j)(-1)^{j+k+l}}{\Gamma(b-j)\Gamma(a+j-k)k!j!} 2^{k} \binom{k}{l}.$$

Replacing $\Phi(x)$ by the error function in Eq. (15), implies

$$f(x) = \sqrt{\frac{2}{\pi}} \left(\frac{\alpha}{x}\right) \left(\frac{x}{\theta}\right)^{\alpha} \mathrm{e}^{-\frac{1}{2}\left(\frac{x}{\theta}\right)^{2\alpha}} \sum_{j,k=0}^{\infty} \sum_{l=0}^{k} \sum_{p=0}^{l} t_{j,k,l,p}(a,b) \left\{\mathrm{erf}\left[\frac{\left(\frac{x}{\theta}\right)^{\alpha}}{\sqrt{2}}\right]\right\}^{p}, \tag{16}$$

where

$$t_{j,k,l,p}(a,b) = \frac{w_{j,k,l}(a,b)(-1)^{p}\binom{l}{p}}{B(a,b)2^{l}}.$$

Eqs. (15)-(16) are the main expansions for the BGHN density function.

3. Properties of the BGHN distribution

We hardly need to emphasize the necessity and importance of moments and moment generating function in any statistical analysis especially in applied work. Some of the most important features and characteristics of a distribution can be studied through moments (e.g., tendency, dispersion, skewness and kurtosis).

3.1. Moments

Theorem 1. *If $X \sim BGHN(\alpha, \theta, a, b)$, then the nth moment is given by*

$$\mu_n' = \theta^{n} \sqrt{\frac{2}{\pi}} \sum_{j,k=0}^{\infty} \sum_{l=0}^{k} \sum_{p=0}^{l} t_{j,k,l,p}(a,b) I\left(\frac{n}{\alpha}, p\right), \tag{17}$$

where $p + \frac{n}{\alpha}$ is a real number, $t_{j,k,l,p}(a,b)$ is just defined after Eq. (16) and

$$I\left(\frac{n}{\alpha}, p\right) = \pi^{-\frac{p}{2}} 2^{p+\frac{n}{2\alpha}-\frac{1}{2}} \sum_{m_1=0}^{\infty} \cdots \sum_{m_p=0}^{\infty} \frac{(-1)^{m_1+\cdots+m_p}}{(m_1+1/2)\cdots(m_p+1/2)m_1!\cdots m_p!} \Gamma\left(m_1 + \cdots + m_p + \frac{p+\frac{n}{\alpha}+1}{2}\right).$$

Proof. The nth moment of the BGHN distribution is $\mu_n' = \int_0^{\infty} x^n f(x)\,\mathrm{d}x$. Hence, if $b > 0$ is real non-integer, we obtain from Eq. (16)

$$\mu_n' = \alpha \sqrt{\frac{2}{\pi}} \sum_{j,k=0}^{\infty} \sum_{l=0}^{k} \sum_{p=0}^{l} t_{j,k,l,p}(a,b) \int_0^{\infty} x^{n-1} \left(\frac{x}{\theta}\right)^{\alpha} \mathrm{e}^{-\frac{1}{2}\left(\frac{x}{\theta}\right)^{2\alpha}} \left\{\mathrm{erf}\left[\frac{\left(\frac{x}{\theta}\right)^{\alpha}}{\sqrt{2}}\right]\right\}^{p} \mathrm{d}x. \tag{18}$$

Next, on setting $u = \left(\frac{x}{\theta}\right)^{\alpha}$, μ_n' reduces to

$$\mu_n' = \theta^{n} \sqrt{\frac{2}{\pi}} \sum_{j,k=0}^{\infty} \sum_{l=0}^{k} \sum_{p=0}^{l} t_{j,k,l,p}(a,b) \int_0^{\infty} u^{\frac{n}{\alpha}} \exp\left(-\frac{u^2}{2}\right) \left[\mathrm{erf}\left(\frac{u}{\sqrt{2}}\right)\right]^{p} \mathrm{d}u. \tag{19}$$

Please cite this article in press as: Pescim, R.R., et al., The beta generalized half-normal distribution. Computational Statistics and Data Analysis (2009), doi:10.1016/j.csda.2009.10.007

Using the series expansion for the error function $\mathrm{erf}(x)=\frac{2}{\sqrt{\pi}}\sum_{m=0}^{\infty}\frac{(-1)^m x^{2m+1}}{(2m+1)m!}$ (see, for example Nadarajah, 2008) and after some algebra, we have

$$\mu_n' = \theta^n\sqrt{\frac{2}{\pi}}\sum_{j,k=0}^{\infty}\sum_{l=0}^{k}\sum_{p=0}^{l} t_{j,k,l,p}(a,b)\left(\frac{2}{\sqrt{\pi}}\right)^p \times \sum_{m_1=0}^{\infty}\cdots\sum_{m_p=0}^{\infty}\frac{(-1)^{m_1+\cdots+m_p}}{2^{m_1+\cdots+m_p+\frac{p}{2}}(2m_1+1)\cdots(2m_p+1)m_1!\cdots m_p!}$$
$$\times\int_0^{\infty}u^{2(m_1+\cdots+m_p)+p+\frac{n}{\alpha}}\exp\left(-\frac{u^2}{2}\right)\mathrm{d}u. \quad (20)$$

Calculating the integral in Eq. (20), we obtain

$$\mu_n' = \theta^n\sqrt{\frac{2}{\pi}}\sum_{j,k=0}^{\infty}\sum_{l=0}^{k}\sum_{p=0}^{l} t_{j,k,l,p}(a,b) I\left(\frac{n}{\alpha},p\right), \quad (21)$$

for all $p+\frac{n}{\alpha}$ real number and

$$I\left(\frac{n}{\alpha},p\right) = \pi^{-\frac{p}{2}}2^{p+\frac{n}{2\alpha}-\frac{1}{2}}\sum_{m_1=0}^{\infty}\cdots\sum_{m_p=0}^{\infty}\frac{(-1)^{m_1+\cdots+m_p}}{(m_1+1/2)\cdots(m_p+1/2)m_1!\cdots m_p!}$$
$$\times\Gamma\left(m_1+\cdots+m_p+\frac{p+\frac{n}{\alpha}+1}{2}\right). \quad \square$$

Note that, if $p+\frac{n}{\alpha}$ is even, the integral $I\left(\frac{n}{\alpha},p\right)$ can be written in terms of the Lauricella function of type A (Exton, 1978; Aarts, 2000) defined by

$$F_A^{(n)}(a;b_1,\ldots,b_n;c_1,\ldots,c_n;x_1,\ldots,x_n) = \sum_{m_1=0}^{\infty}\cdots\sum_{m_n=0}^{\infty}\frac{(a)_{m_1+\cdots+m_n}(b_1)_{m_1}\ldots(b_n)_{m_n}}{(c_1)_{m_1}\ldots(c_n)_{m_n}}\frac{x_1^{m_1}\ldots x_n^{m_n}}{m_1!\ldots m_{n!}},$$

where $(a)_k = a(a-1)\ldots(a-k+1)$ is the ascending factorial (with the convention that $(a)_0=1$). Numerical routines for the direct computation of the Lauricella function of type A are available, see Exton (1978) and Mathematica (Trott, 2006).

Hence, μ_n' can be written in terms of the Lauricella functions of type A

$$\mu_n' = \theta^n\sqrt{\frac{2}{\pi}}\sum_{j,k=0}^{\infty}\sum_{l=0}^{k}\sum_{p=0}^{l} t_{j,k,l,p}(a,b)\pi^{-\frac{p}{2}}2^{2p+\frac{n}{2\alpha}-\frac{1}{2}}\Gamma\left(\frac{p+\frac{n}{\alpha}+1}{2}\right)$$
$$\times F_A^{(p)}\left(\frac{p+\frac{n}{\alpha}+1}{2};\frac{1}{2},\ldots,\frac{1}{2};\frac{3}{2},\ldots,\frac{3}{2};-1,\ldots,-1\right). \quad (22)$$

Eq. (22) is an infinite sum of the Lauricella functions of type A. Note also that the terms of the sums in (22) vanish when $p+\frac{n}{\alpha}$ is odd.

The skewness and kurtosis measures can now be calculated from the ordinary moments using well-known relationships. Plots of the skewness and kurtosis for some choices of the parameter b as function of the parameter a, and for some choices of the parameter a as function of the parameter b, for $\alpha=1$ and $\theta=55$, are shown in Figs. 3 and 4, respectively. These figures immediately reveal that the skewness and kurtosis curves increase (decrease) with b (a) for fixed a (b).

3.2. Moment generating function

Theorem 2. *If $X\sim BGHN(\alpha,\theta,a,b)$, then its moment generating function (mgf) is given by*

$$M(s) = \sqrt{\frac{2}{\pi}}\sum_{j,k=0}^{\infty}\sum_{l=0}^{k}\sum_{p=0}^{l} t_{j,k,l,p}(a,b)\pi^{-\frac{p}{2}}2^{p+\frac{r}{2\alpha}-\frac{1}{2}}\sum_{r=0}^{\infty}\frac{\theta^r s^r}{r!}I\left(\frac{r}{\alpha},p\right)$$

where $I\left(\frac{r}{\alpha},p\right)$ is just defined after Eq. (17).

Proof. The mgf of the BGHN distribution is calculated from the relation

$$M(s) = \int_0^{\infty}\exp(sx)\sqrt{\frac{2}{\pi}}\left(\frac{\alpha}{x}\right)\left(\frac{x}{\theta}\right)^{\alpha}\mathrm{e}^{-\frac{1}{2}\left(\frac{x}{\theta}\right)^{2\alpha}}\sum_{j,k=0}^{\infty}\sum_{l=0}^{k}\sum_{p=0}^{l}t_{j,k,l,p}(a,b)\left\{\mathrm{erf}\left[\frac{\left(\frac{x}{\theta}\right)^{\alpha}}{\sqrt{2}}\right]\right\}^p\mathrm{d}x,$$

and then

$$M(s) = \alpha\sqrt{\frac{2}{\pi}}\sum_{j,k=0}^{\infty}\sum_{l=0}^{k}\sum_{p=0}^{l}t_{j,k,l,p}(a,b)\sum_{r=0}^{\infty}\frac{s^r}{r!}\int_0^{\infty}x^{r-1}\left(\frac{x}{\theta}\right)^{\alpha}\mathrm{e}^{-\frac{1}{2}\left(\frac{x}{\theta}\right)^{2\alpha}}\left\{\mathrm{erf}\left[\frac{\left(\frac{x}{\theta}\right)^{\alpha}}{\sqrt{2}}\right]\right\}^p\mathrm{d}x.$$

Please cite this article in press as: Pescim, R.R., et al., The beta generalized half-normal distribution. Computational Statistics and Data Analysis (2009), doi:10.1016/j.csda.2009.10.007

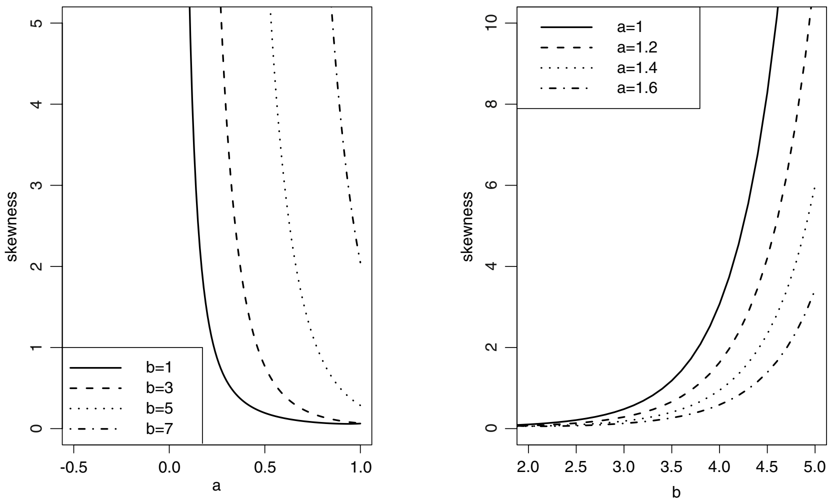

Fig. 3. Skewness of the BGHN distribution as function of *a* for some values of *b* and as function of *b* for some values of *a*.

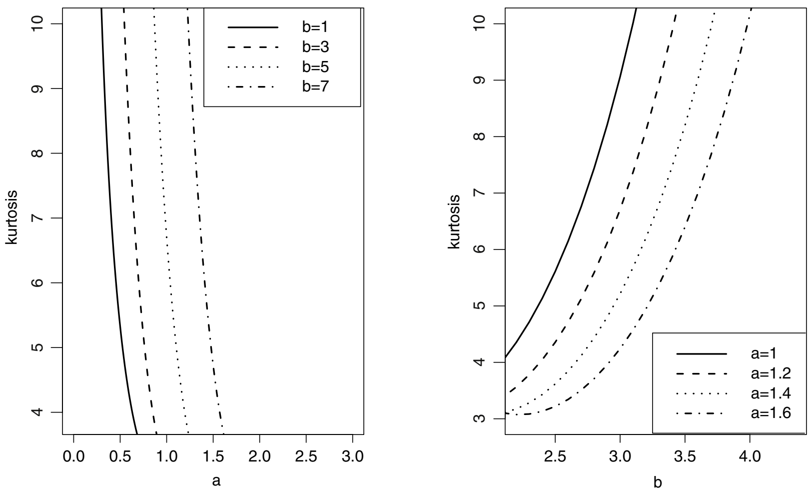

Fig. 4. Kurtosis of the BGHN distribution as function of *a* for some values of *b* and as function of *b* for some values of *a*.

Substituting $u = \left(\frac{x}{\theta}\right)^{\alpha}$, we have

$$M(s) = \sqrt{\frac{2}{\pi}} \sum_{j,k=0}^{\infty} \sum_{l=0}^{k} \sum_{p=0}^{l} t_{j,k,l,p}(a, b) \sum_{r=0}^{\infty} \frac{\theta^r s^r}{r!} \int_0^{\infty} u^{\frac{r}{\alpha}} \exp\left(-\frac{u^2}{2}\right) \left[\operatorname{erf}\left(\frac{u}{\sqrt{2}}\right)\right]^p \mathrm{d}u. \tag{23}$$

Following similar steps of Theorem 1, the mgf of the BGHN distribution takes the form

$$M(s) = \sqrt{\frac{2}{\pi}} \sum_{j,k=0}^{\infty} \sum_{l=0}^{k} \sum_{p=0}^{l} t_{j,k,l,p}(a, b) \pi^{-\frac{p}{2}} 2^{p+\frac{r}{2\alpha}-\frac{1}{2}} \sum_{r=0}^{\infty} \frac{\theta^r s^r}{r!} I\left(\frac{r}{\alpha}, p\right). \quad \square \tag{24}$$

4. Expansions for the order statistics

Moments of order statistics play an important role in quality control testing and reliability, where a practitioner needs to predict the failure of future items based on the times of a few early failures. These predictors are often based on moments of order statistics.

We now derive an explicit expression for the density of the ith order statistic $X_{i:n}$, say $f_{i:n}(x)$, in a random sample of size n from the BGHN distribution. It is well known that

$$f_{i:n}(x) = \frac{1}{B(i, n-i+1)} f(x) F(x)^{i-1} \{1 - F(x)\}^{n-i},$$

Please cite this article in press as: Pescim, R.R., et al., The beta generalized half-normal distribution. Computational Statistics and Data Analysis (2009), doi:10.1016/j.csda.2009.10.007

for $i = 1, \ldots, n$. Combining (1) and (3), the density $f_{i:n}(x)$ can expressed in terms of the incomplete beta function ratio

$$f_{i:n}(x) = \frac{n!g(x)}{(i-1)!(n-i)!B(a,b)} G(x)^{a-1}[1-G(x)]^{b-1}\left[I_{G(x)}(a,b)\right]^{i-1}\left[1-I_{G(x)}(a,b)\right]^{n-i}. \tag{25}$$

Substituting (7), (9) and (10) in Eq. (25), the density $f_{i:n}(x)$ for $b > 0$ real non-integer becomes

$$f_{i:n}(x) = \sum_{k=0}^{n-i} \frac{(-1)^k \binom{n-i}{k} \Gamma(b)^{i+k-1} \sqrt{\frac{2}{\pi}} \left(\frac{\alpha}{x}\right) \left(\frac{x}{\theta}\right)^{\alpha} \mathrm{e}^{-\frac{1}{2}\left(\frac{x}{\theta}\right)^{2\alpha}} 2^{b-1} \left\{2\Phi\left[\left(\frac{x}{\theta}\right)^{\alpha}\right]-1\right\}^{a(i+k)-1}}{B(a,b)^{i+k} B(i,n-1+i)\left\{1-\Phi\left[\left(\frac{x}{\theta}\right)^{\alpha}\right]\right\}^{-(b-1)}}$$
$$\times \left[\sum_{j=0}^{\infty} \frac{(-1)^j \{2\Phi\left[\left(\frac{x}{\theta}\right)^{\alpha}\right]-1\}^j}{\Gamma(b-j)j!(a+j)}\right]^{i+k-1}. \tag{26}$$

We now consider the expression

$$A = \sum_{j=0}^{\infty} \frac{(-1)^j \{2\Phi\left[\left(\frac{x}{\theta}\right)^{\alpha}\right]-1\}^j}{\Gamma(b-j)j!(a+j)}. \tag{27}$$

Let $u = \left(\frac{x}{\theta}\right)^{\alpha}$. By applying Eq. (5) in (27), we obtain

$$A = \sum_{j=0}^{\infty} \frac{(-1)^j \left[\operatorname{erf}\left(\frac{u}{\sqrt{2}}\right)\right]^j}{\Gamma(b-j)j!(a+j)}. \tag{28}$$

Using the series expansion for the error function erf(.) in Eq. (28), yields

$$A = \sum_{j=0}^{\infty} \left\{ \frac{(-1)^j \left[\frac{2}{\sqrt{\pi}} \sum_{m=0}^{\infty} \frac{(-1)^m u^{2m+1}}{2^{m+\frac{1}{2}}(2m+1)m!}\right]^j}{\Gamma(b-j)j!(a+j)} \right\}.$$

Hence,

$$A = \sum_{j=0}^{\infty} \left\{ \frac{(-1)^j \left(\frac{2}{\sqrt{\pi}}\right)^j \left[\sum_{m_1=0}^{\infty} \cdots \sum_{m_j=0}^{\infty} \frac{(-1)^{m_1+\cdots+m_j} u^{2(m_1+\cdots+m_j)}}{2^{m_1+\cdots+m_j+\frac{j}{2}}(2m_1+1)\cdots(2m_j+1)m_1!\cdots m_j!}\right]}{\Gamma(b-j)j!(a+j)} u^j \right\}. \tag{29}$$

We use the identity $(\sum_{k=0}^{\infty} a_k x^k)^n = \sum_{k=0}^{\infty} c_{k,n} x^k$ (see Gradshteyn and Ryzhik, 2000), where now a_k comes by identifying (29) with the corresponding quantity which is elevated to the power $i+k-1$ in Eq. (26). We have

$$a_k = \frac{(-1)^k \left(\frac{2}{\sqrt{\pi}}\right)^k \left[\sum_{m_1=0}^{\infty} \cdots \sum_{m_k=0}^{\infty} \frac{(-1)^{m_1+\cdots+m_k} u^{2(m_1+\cdots+m_k)}}{2^{m_1+\cdots+m_k+\frac{k}{2}}(2m_1+1)\cdots(2m_k+1)m_1!\cdots m_k!}\right]}{\Gamma(b-k)k!(a+k)},$$

$$c_{0,n} = a_0^n \quad \text{and} \quad c_{k,n} = (ka_0)^{-1} \sum_{l=1}^{k} (nl-k+l) a_l c_{k-l,n}$$

for $k = 1, 2, \ldots$ Thus, after some algebra, we can write

$$f_{i:n}(x) = \sum_{k=0}^{n-i} \sum_{j=0}^{\infty} \frac{(-1)^k \binom{n-i}{k} \Gamma(b)^{i+k-1} B[a(i+k)+j,b] d_{i,j,k}}{B(a,b)^{i+k} B(i,n-1+i)} f_{i,j,k}(x), \tag{30}$$

where

$$f_{i,j,k}(x) = \frac{\sqrt{\frac{2}{\pi}} \left(\frac{\alpha}{x}\right) \left(\frac{x}{\theta}\right)^{\alpha} \mathrm{e}^{-\frac{1}{2}\left(\frac{x}{\theta}\right)^{2\alpha}} 2^{b-1}}{B[a(i+k)+j,b]} \left\{2\Phi\left[\left(\frac{x}{\theta}\right)^{\alpha}\right]-1\right\}^{a(i+k)+j-1} \left\{1-\Phi\left[\left(\frac{x}{\theta}\right)^{\alpha}\right]\right\}^{b-1}$$

denotes the density of a BGHN$(\alpha, \theta, a(i+k)+j, b)$ distribution and the constants $d_{i,j,k}$ can be obtained recursively from the following equations

$$d_{i,0,k} = \left\{\frac{1}{a\Gamma(b)}\right\}^{i+k-1} \quad \text{and} \quad d_{i,j,k} = \frac{a\Gamma(b)}{j}\sum_{l=1}^{j}\frac{(-1)^l\{l(i+k)-j\}}{\Gamma(b-l)(a+l)l!}c_{j-l,i+k-1}, \quad j \geq 1.$$

The density of the BGHN order statistics is then an infinite mixture of BGHN densities. Hence, the ordinary and central moments of the order statistics can be calculated directly from the corresponding quantities of the BGHN distribution given in Section 3. For $b > 0$ integer, expansion (30) holds but the sum in j stops at $(b-1)(k+i-1)$.

An alternative expansion for the density of the order statistics follow from the identity $\left(\sum_{i=1}^{\infty} a_i\right)^k = \sum_{\{m_1,\ldots,m_k\}=0}^{\infty} a_{m_1}\ldots a_{m_k}$ for k a positive integer. Using this identity and Eq. (26), for $b > 0$ real non-integer, it turns out that

$$f_{i:n}(x) = \sum_{k=0}^{n-i}\sum_{m_1=0}^{\infty}\cdots\sum_{m_{i+k-1}=0}^{\infty}\frac{\sqrt{\frac{2}{\pi}}\left(\frac{\alpha}{x}\right)\left(\frac{x}{\theta}\right)^{\alpha} e^{-\frac{1}{2}\left(\frac{x}{\theta}\right)^{2\alpha}} 2^{b-1}\left\{2\Phi\left[\left(\frac{x}{\theta}\right)^{\alpha}\right]-1\right\}^{a-1}\left\{1-\Phi\left[\left(\frac{x}{\theta}\right)^{\alpha}\right]\right\}^{b-1}}{B(a,b)^{i+k}B(i,n-i+1)}$$
$$\times\frac{(-1)^{k+\sum_{j=1}^{i+k-1}m_j}\binom{n-i}{k}\Gamma(b)^{i+k-1}\left\{2\Phi\left[\left(\frac{x}{\theta}\right)^{\alpha}\right]-1\right\}^{a(i+k)+\sum_{j=1}^{i+k-1}m_j}}{\prod_{j=1}^{i+k-1}\Gamma(b-m_j)m_j!(a+m_j)}.$$

Hence,

$$f_{i:n}(x) = \sum_{k=0}^{n-i}\sum_{m_1=0}^{\infty}\cdots\sum_{m_{i+k-1}=0}^{\infty}\delta_{i,k}f_{i,k}(x), \tag{31}$$

where

$$f_{i,k}(x) = \frac{\sqrt{\frac{2}{\pi}}\left(\frac{\alpha}{x}\right)\left(\frac{x}{\theta}\right)^{\alpha} e^{-\frac{1}{2}\left(\frac{x}{\theta}\right)^{2\alpha}} 2^{b-1}}{B\left[a(i+k)+\sum_{j=1}^{i+k-1}m_j, b\right]}\left\{2\Phi\left[\left(\frac{x}{\theta}\right)^{\alpha}\right]-1\right\}^{\left[a(i+k)+\sum_{j=1}^{i+k-1}m_j\right]-1}\left\{1-\Phi\left[\left(\frac{x}{\theta}\right)^{\alpha}\right]\right\}^{b-1}$$

denotes the density of a BGHN $(\alpha, \theta, a(i+k)+\sum_{j=1}^{i+k-1}m_j, b)$ distribution and

$$\delta_{i,k} = \frac{(-1)^{k+\sum_{j=1}^{i+k-1}m_j}\binom{n-i}{k}B\left(a(i+k)+\sum_{j=1}^{i+k-1}m_j, b\right)\Gamma(b)^{i+k-1}}{B(a,b)^{i+k}B(i,n-i+1)\prod_{j=1}^{i+k-1}\Gamma(b-m_j)m_j!(a+m_j)}.$$

The constants $\delta_{i,k}$ are easily obtained given i, n, k and a sequence of indices $m_1, \ldots, m_{i+k-1}$. The sums in (31) extend over all $(i+k)$-tuples $(k, m_1, \ldots, m_{i+k-1})$ of non-negative integers and can be implementable on a computer. If $b > 0$ is an integer, Eq. (31) holds but the indices $m_1, \ldots, m_{i+k-1}$ vary from zero to $b-1$. Expansion (30) is much simpler to be calculated numerically in applications and its CPU times are usually smaller than using (31).

The sth moment of $X_{i:n}$ for $b > 0$ real non-integer comes from (30)

$$\mathrm{E}(X_{i:n}^s) = \sum_{k=0}^{n-i}\sum_{j=0}^{\infty}\frac{(-1)^k\binom{n-i}{k}\Gamma(b)^{i+k-1}B(a(i+k)+j,b)d_{i,j,k}}{B(a,b)^{i+k}B(i,n-i+1)}E(X_{i,j,k}^s), \tag{32}$$

where $X_{i,j,k} \sim$ BGHN $(\alpha, \theta, a(i+k)+j, b)$ and the constants $d_{i,j,k}$ were defined before. If b is an integer, the sum in j stops at $b-1$.

From Eq. (31), we can obtain an alternative expression for the moments of the order statistics valid for $b > 0$ real non-integer

$$\mathrm{E}(X_{i:n}^s) = \sum_{k=0}^{n-i}\sum_{m_1=0}^{\infty}\cdots\sum_{m_{i+k-1}=0}^{\infty}\delta_{i,k}E(X_{i,k}^s), \tag{33}$$

where $X_{i,k} \sim$ BGHN$(\alpha, \theta, a(i+k)+\sum_{j=1}^{i+k-1}m_j, b)$. For $b > 0$ integer, the indices $m_1, \ldots, m_{i+k-1}$ stop at $b-1$.

We therefore offer two Eqs. (32) and (33) to obtain the moments of the BGHN order statistics which are the main results of this section.

Based upon the moments given in Eqs. (32) and (33), we can easily derive expansions for the L-moments of the BGHN distribution as infinite weighted linear combinations of the means of suitable BGHN distributions. The L-moments are analogous to the ordinary moments but can be estimated by linear combinations of order statistics. They have the advantage that they exist whenever the mean of the distribution exists, even though some higher moments may not exist, and are relatively robust to the effects of outliers. They are linear functions of expected order statistics defined by (Hosking, 1990)

$$\lambda_{r+1} = r(r+1)^{-1}\sum_{k=0}^{r}\frac{(-1)^k}{k}\mathrm{E}(X_{r+1-k:r+1}), \quad r = 0, 1, \ldots .$$

The first four L-moments are: $\lambda_1 = \mathrm{E}(X_{1:1})$, $\lambda_2 = \frac{1}{2}\mathrm{E}(X_{2:2} - X_{1:2})$, $\lambda_3 = \frac{1}{3}\mathrm{E}(X_{3:3} - 2X_{2:3} + X_{1:3})$ and $\lambda_4 = \frac{1}{4}\mathrm{E}(X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4})$.

5. Estimation and inference

We consider estimation of the parameters of the BGHN distribution by the method of maximum likelihood. If Y follows a BGHN distribution with vector of parameters $\boldsymbol{\lambda} = (\alpha, \theta, a, b)^{\mathrm{T}}$, the log-likelihood for the model parameters from a single observation y of Y is given by

$$\begin{aligned}\ell(\boldsymbol{\lambda}) = {} & \log\left(\sqrt{\frac{2}{\pi}}\right) + \log(\alpha) - \log(y) + \alpha\log\left(\frac{y}{\theta}\right) - \frac{1}{2}\left(\frac{y}{\theta}\right)^{2\alpha} - \log\big[B(a,b)\big] + (b-1)\log(2)\\ & + (a-1)\log\left\{2\Phi\left[\left(\frac{y}{\theta}\right)^{\alpha}\right] - 1\right\} + (b-1)\log\left\{1 - \Phi\left[\left(\frac{y}{\theta}\right)^{\alpha}\right]\right\}, \quad y > 0.\end{aligned}$$

The components of the unit score vector $\mathbf{U} = \left(\frac{\partial\ell}{\partial\alpha}, \frac{\partial\ell}{\partial\theta}, \frac{\partial\ell}{\partial a}, \frac{\partial\ell}{\partial b}\right)^{\mathrm{T}}$ are obtained by differentiation

$$\frac{\partial\ell}{\partial\alpha} = \frac{1}{\alpha} + \log\left(\frac{y}{\theta}\right) - \log\left(\frac{y}{\theta}\right)\left(\frac{y}{\theta}\right)^{2\alpha} + \frac{2(a-1)}{\sqrt{2\pi}}\left\{\frac{v\log\left(\frac{y}{\theta}\right)}{2\Phi\left[\left(\frac{y}{\theta}\right)^{\alpha}\right] - 1}\right\} + \frac{(1-b)}{\sqrt{2\pi}}\left\{\frac{v\log\left(\frac{y}{\theta}\right)}{1 - \Phi\left[\left(\frac{y}{\theta}\right)^{\alpha}\right]}\right\},$$

$$\frac{\partial\ell}{\partial\theta} = \frac{\alpha}{\theta}\left(\frac{y}{\theta}\right)^{2\alpha} - \left(\frac{\alpha}{\theta}\right) + \frac{2(1-a)}{\sqrt{2\pi}}\left\{\frac{v\left(\frac{\alpha}{\theta}\right)}{2\Phi\left[\left(\frac{y}{\theta}\right)^{\alpha}\right] - 1}\right\} + \frac{(1-b)}{\sqrt{2\pi}}\left\{\frac{v\left(\frac{\alpha}{\theta}\right)}{1 - \Phi\left[\left(\frac{y}{\theta}\right)^{\alpha}\right]}\right\},$$

$$\frac{\partial\ell}{\partial a} = \log\left\{2\Phi\left[\left(\frac{y}{\theta}\right)^{\alpha}\right] - 1\right\} - \psi(a) + \psi(a+b), \qquad \frac{\partial\ell}{\partial b} = \log 2 + \log\left\{1 - \Phi\left[\left(\frac{y}{\theta}\right)^{\alpha}\right]\right\} - \psi(b) + \psi(a+b),$$

where $v = \exp\left[-\frac{1}{2}\left(\frac{y}{\theta}\right)^{2\alpha}\right]\left(\frac{y}{\theta}\right)^{\alpha}$ and $\psi(.)$ is the digamma function.

For a random sample $y = (y_1, \ldots, y_n)^{\mathrm{T}}$ of size n from Y, the total log-likelihood is $\ell_n = \ell_n(\boldsymbol{\lambda}) = \sum_{i=1}^{n}\ell^{(i)}(\boldsymbol{\lambda})$, where $\ell^{(i)}(\boldsymbol{\lambda})$ is the log-likelihood for the ith observation ($i = 1, \ldots, n$). The total score function is $\mathbf{U}_n = \sum_{i=1}^{n}\mathbf{U}^{(i)}$, where $\mathbf{U}^{(i)}$ has the form given before for $i = 1, \ldots, n$. The MLE $\hat{\boldsymbol{\lambda}}$ of $\boldsymbol{\lambda}$ is the solution of the system of nonlinear equations $\mathbf{U}_n = \mathbf{0}$.

For interval estimation and tests of hypotheses on the parameters in $\boldsymbol{\lambda}$, we require the 4×4 unit expected information matrix

$$\mathbf{K} = \mathbf{K}(\boldsymbol{\lambda}) = \begin{pmatrix} \kappa_{\alpha,\alpha} & \kappa_{\alpha,\theta} & \kappa_{\alpha,a} & \kappa_{\alpha,b}\\ \kappa_{\theta,\alpha} & \kappa_{\theta,\theta} & \kappa_{\theta,a} & \kappa_{\theta,b}\\ \kappa_{a,\alpha} & \kappa_{a,\theta} & \kappa_{a,a} & \kappa_{a,b}\\ \kappa_{b,\alpha} & \kappa_{b,\theta} & \kappa_{b,a} & \kappa_{b,b}\end{pmatrix},$$

where the corresponding elements are given in Appendix.

Under conditions that are fulfilled for parameters in the interior of the parameter space but not on the boundary, the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda})$ is $\mathrm{N}_4(0, \mathbf{K}(\boldsymbol{\lambda})^{-1})$. The estimated asymptotic multivariate normal $\mathrm{N}_4(0, n^{-1}\mathbf{K}(\hat{\boldsymbol{\lambda}})^{-1})$ distribution of $\hat{\boldsymbol{\lambda}}$ can be used to construct approximate confidence intervals for the parameters and for the hazard rate and survival functions. An asymptotic confidence interval with significance level γ for each parameter λ_r is given by

$$\mathrm{ACI}(\lambda_r, 100(1-\gamma)\%) = (\hat{\lambda}_r - z_{\gamma/2}\sqrt{\hat{\kappa}^{\lambda_r,\lambda_r}}, \hat{\lambda}_r + z_{\gamma/2}\sqrt{\hat{\kappa}^{\lambda_r,\lambda_r}}),$$

where $\hat{\kappa}^{\lambda_r,\lambda_r}$ is the rth diagonal element of $n^{-1}\mathbf{K}(\boldsymbol{\lambda})^{-1}$ estimated at $\hat{\boldsymbol{\lambda}}$, for $r = 1, \ldots, 4$, and $z_{\gamma/2}$ is the quantile $1 - \gamma/2$ of the standard normal distribution.

The likelihood ratio (LR) statistic is useful for testing goodness of fit of the BGHN distribution and for comparing this distribution with some of its special sub-models. We can compute the maximum values of the unrestricted and restricted log-likelihoods to construct the LR statistics for testing some sub-models of the BGHN distribution. For example, we may use the LR statistic to check if the fit using the BGHN distribution is statistically "superior" to a fit using the GHN distribution for a given data set. In any case, considering the partition $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1^{\mathrm{T}}, \boldsymbol{\lambda}_2^{\mathrm{T}})^{\mathrm{T}}$, tests of hypotheses of the type $H_0 : \boldsymbol{\lambda}_1 = \boldsymbol{\lambda}_1^{(0)}$

versus $H_A : \boldsymbol{\lambda}_1 \neq \boldsymbol{\lambda}_1^{(0)}$ can be performed via the LR statistic $w = 2\{\ell(\hat{\boldsymbol{\lambda}}) - \ell(\tilde{\boldsymbol{\lambda}})\}$, where $\hat{\boldsymbol{\lambda}}$ and $\tilde{\boldsymbol{\lambda}}$ are the MLEs of $\boldsymbol{\lambda}$ under H_A and H_0, respectively. Under the null hypothesis H_0, $w \xrightarrow{d} \chi_q^2$, where q is the dimension of the vector $\boldsymbol{\lambda}_1$ of interest. The LR test rejects H_0 if $w > \xi_\gamma$, where ξ_γ denotes the upper $100\gamma\%$ point of the χ_q^2 distribution. From the score vector and the information matrix given before, we can also construct the score and Wald statistics which are asymptotically equivalent to the LR statistic.

6. Application — Myelogenous leukemia data

The data set is given in Feigl and Zelen (1965) for two groups of patients who died of acute myelogenous leukemia. The patients were classified into the two groups according to the presence or absence of a morphologic characteristic of white cells. The patients termed AG positive were identified by the presence of Auer rods and/or significant granulative of the leukemic cells in the bone marrow at diagnosis. For the AG negative patients these factors were absent. Here, we consider only patients termed AG positive ($n = 17$). The survival times t_i for $i = 1, \ldots, 17$ are given in weeks from the date of diagnosis.

In many applications, there is a qualitative information about the failure rate function shape, which can help in selecting a particular model. In this context, a device called the total time on test (TTT) plot (Aarset, 1987) is useful. The TTT-plot is obtained by plotting $\mathrm{G}(r/n) = [(\sum_{i=1}^{r} T_{i:n}) + (n - r)T_{r:n}]/(\sum_{i=1}^{n} T_{i:n})$, where $r = 1, \ldots, n$ and $T_{i:n}$, $i = 1, \ldots, n$, are the order statistics of the sample, against r/n (Mudholkar et al., 1996).

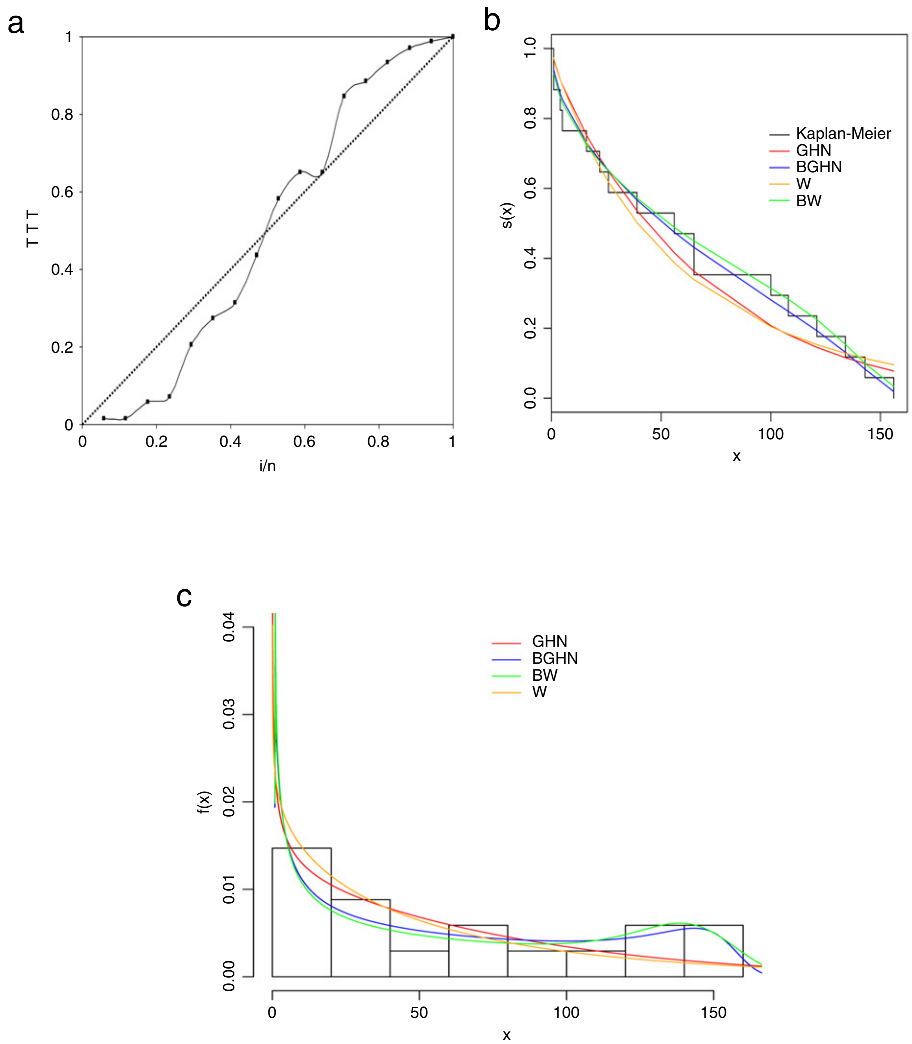

Fig. 5. (a) TTT-plot on myelogenous leukemia data. (b) Estimated survival function from the fitted BGHN, GHN, BW and Weibull distributions and the empirical survival for myelogenous leukemia data. (c) Estimated densities of the BGHN, GHN, BW and Weibull models for myelogenous leukemia data.

Please cite this article in press as: Pescim, R.R., et al., The beta generalized half-normal distribution. Computational Statistics and Data Analysis (2009), doi:10.1016/j.csda.2009.10.007

Table 1
MLEs of the model parameters for the myelogenous leukemia data, the corresponding SE (given in parentheses) and the measures AIC, BIC and CAIC.

Model	α	θ	a	b	AIC	BIC	CAIC
BGHN	7.2688	132.50	0.0733	0.3319	165.8	169.1	169.0
	(0.2117)	(13.7559)	(0.0212)	(0.3240)			
GHN	0.7557	73.6234	1	1	168.9	170.6	169.8
	(0.1651)	(17.8232)	–	–			
	α	γ	a	b			
BW	110.90	7.8003	0.0631	0.1514	174.9	178.3	178.1
	(12.2136)	(0.2754)	(0.0182)	(0.0958)			
Weibull	59.4592	0.8856	1	1	178.2	179.9	179.1
	(17.0307)	(0.1834)	–	–			

Fig. 5a shows that the TTT-plot for these data has first a convex shape and then a concave shape. It indicates a bathtub shaped hazard rate function. Hence, the BGHN distribution could be in principle an appropriate model for fitting such data. Table 1 lists the MLEs (the corresponding standard errors are in parentheses) of the parameters from the fitted BGHN, GHN, BW and Weibull models and the values of the following statistics: AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion) and CAIC (Consistent Akaike Information Criterion). The computations were performed using the subroutine NLMixed in SAS. These results indicate that the BGHN model has the lowest values for the AIC, BIC and CAIC statistics among the fitted models, and therefore it could be chosen as the best model.

In order to assess if the model is appropriate, Fig. 5b plots the empirical and estimated survival functions of the BGHN, GHN, BW and Weibull distributions. Further, Fig. 5c plots the histogram of the data and the fitted BGHN, GHN, BW and Weibull distributions. We conclude that the BGHN distribution provides a good fit for these data.

7. Conclusions

We propose the beta generalized half-normal (BGHN) distribution to extend the half-normal (HN) distribution and the generalized half-normal (GHN) distribution introduced by Cooray and Ananda (2008). We provide a mathematical treatment of the new distribution including expansions for its distribution and density functions. We derive infinite sums for the moments and the moment generating function. We examine the maximum likelihood estimation of the model parameters and derive the expected information matrix. We consider the likelihood ratio (LR) statistic which may be very useful to compare the BGHN model with its sub-models. An application of the BGHN distribution to a real data set is given to demonstrate that this distribution can be used quite effectively to provide better fits than other available models. We hope that this generalization may attract wider applications in survival analysis and biology.

Acknowledgments

We gratefully acknowledge the grants from CNPq. The authors are also grateful to the editor and two referees for their helpful comments and suggestions.

Appendix

The elements of the 4×4 unit expected information matrix are given by

$$\begin{aligned}
\kappa_{\alpha,\alpha} &= \frac{1}{\alpha^2}\left(1+2T_{0,0,2,2,1,1}\right) + \frac{2(a-1)}{\alpha^2\sqrt{2\pi}}\left[T_{1,0,3,2,2,1} - T_{1,0,1,2,2,1} + \frac{2}{\alpha^2}T_{0,1,2,2,3,1}\right] \\
&\quad + \frac{(1-b)}{\alpha^2\sqrt{2\pi}}\left[T_{1,0,3,2,1,2} - T_{1,0,1,2,1,2} - T_{0,1,2,2,1,3}\right], \\
\kappa_{\alpha,\theta} &= -\frac{1}{\theta}\left[T_{0,0,2,0,1,1} + 2T_{0,0,2,1,1,1} - 1\right] - \frac{2(a-1)}{\theta\sqrt{2\pi}}\left[-T_{1,0,3,1,2,1} + T_{1,0,1,0,2,1}\right] \\
&\quad + T_{1,0,1,1,2,1} - 2T_{0,1,2,1,3,1} - \frac{(1-b)}{\theta\sqrt{2\pi}}\left[-T_{1,0,3,1,1,2} + T_{1,0,1,0,1,2} + T_{1,0,1,1,1,2} + T_{0,1,2,1,1,3}\right],
\end{aligned}$$

$$\kappa_{\alpha,a} = -\frac{2}{\alpha\sqrt{2\pi}}T_{1,0,1,1,2,1}, \quad \kappa_{\alpha,b} = \frac{2}{\alpha\sqrt{2\pi}}T_{1,0,1,1,1,2}, \quad \kappa_{a,b} = -\psi'(a+b),$$

$$\begin{aligned}
\kappa_{\theta,\theta} &= \frac{\alpha}{\theta^2}\left[(1+2\alpha)T_{0,0,2,0,1,1} - 1\right] - \frac{2\alpha(1-a)}{\theta^2\sqrt{2\pi}}\left[\alpha T_{1,0,3,0,2,1} - (1+\alpha)T_{1,0,1,0,2,1} + 2\alpha T_{0,1,2,0,3,1}\right] \\
&\quad - \frac{\alpha(1-b)}{\theta^2\sqrt{2\pi}}\left[\alpha T_{1,0,3,0,1,2} - (1+\alpha)T_{1,0,1,0,1,2} - \alpha T_{0,1,2,0,1,3}\right],
\end{aligned}$$

Please cite this article in press as: Pescim, R.R., et al., The beta generalized half-normal distribution. Computational Statistics and Data Analysis (2009), doi:10.1016/j.csda.2009.10.007

$$\kappa_{\theta,a} = \frac{2\alpha}{\theta\sqrt{2\pi}} T_{1,0,1,0,2,1}, \kappa_{\theta,b} = \frac{\alpha}{\theta\sqrt{2\pi}} T_{1,0,1,0,1,2}, \kappa_{a,a} = \psi'(a) - \psi'(a+b), \kappa_{b,b} = \psi'(b) - \psi'(a+b).$$

Here, we assume that a random variable V has a $Beta(a, b)$ distribution and define the expected value

$$T_{i,j,k,l,m,n} = E\left\{\exp\left\{-\frac{i}{2}\left[\Phi^{-1}\left(\frac{V+1}{2}\right)\right]^2\right\}\exp\left\{-j\left[\Phi^{-1}\left(\frac{V+1}{2}\right)\right]^2\right\}\left[\Phi^{-1}\left(\frac{V+1}{2}\right)\right]^k \right.$$
$$\left. \times \left\{\log\left[\Phi^{-1}\left(\frac{V+1}{2}\right)\right]\right\}^l V^{1-m}\left[\left(\frac{1-V}{2}\right)\right]^{1-n}\right\}.$$

These expected values can be determined numerically using MAPLE and MATHEMATICA for any a and b. For example, for $a = 1.5$ and $b = 2.5$, we easily calculated all T's in the information matrix: $T_{1,0,1,0,1,2} = 0.568733$, $T_{1,0,1,0,2,1} = 1.494227$, $T_{1,0,1,1,1,2} = -1.384679$, $T_{1,0,1,1,2,1} = -1.857818$, $T_{0,0,2,0,1,1} = 0.037946$, $T_{0,0,2,1,1,1} = -0.119946$, $T_{1,0,3,1,2,1} = -0.385587$, $T_{1,0,1,1,2,1} = -1.689706$, $T_{0,1,2,1,3,1} = -0.641735$, $T_{1,0,3,1,1,2} = 2.213943$, $T_{0,1,2,1,1,3} = -3.444849$, $T_{0,0,2,2,1,1} = 0.106341$, $T_{1,0,3,2,2,1} = 0.283879$, $T_{1,0,1,2,2,1} = 1.344295$, $T_{0,1,2,2,3,1} = 0.488817$, $T_{1,0,3,2,1,2} = 0.151188$, $T_{0,1,2,2,1,3} = 1.886413$, $T_{1,0,3,0,2,1} = 0.925717$, $T_{0,1,2,0,3,1} = 1.824153$, $T_{1,0,3,0,1,2} = 0.213493$ and $T_{0,1,2,0,1,3} = 1.381865$.

References

Aarset, M.V., 1987. How to identify bathtub hazard rate. IEEE Transactions Reliability 36, 106–108.

Aarts, R.M., (2000). Lauricella functions. www.mathworld.com/LauricellaFunctions.html. From MathWorld — A Wolfram Web Resouce, created by Eric W. Weisstein.

Bebbington, M., Lai, C.D., Zitikis, R., 2007. A flexible Weibull extension. Reliability Engineering and System Safety 92, 719–726.

Carrasco, J.M.F., Ortega, E.M.M., Cordeiro, G.M., 2008. A generalized modified Weibull distribution for lifetime modeling. Computational Statistics and Data Analysis 53, 450–462.

Cooray, K., Ananda, M.M.A., 2008. A generalization of the half-normal distribution wit applications to lifetime data. Communication in Statistics — Theory and Methods 37, 1323–1337.

Díaz-García, J.A., Leiva, V., 2005. A new family of life distributions based on elliptically contoured distributions. Journal of Statistical Planning and Inference 137, 1512–1513.

Doornik, J., 2007. Ox: An Object-Oriented Matrix Programming Language. International Thomson Business Press.

Eugene, N., Lee, C., Famoye, F., 2002. Beta-normal distribution and its applications. Communication in Statistics - Theory and Methods 31, 497–512.

Exton, H., 1978. Handbook of Hypergeometric Integrals: Theory, Applications, Tables, Computer Programs. Halsted Press, New York.

Feigl, P., Zelen, M., 1965. Estimation of exponential survival probabilities with concomitant information. Biometrics 21, 826–837.

Gradshteyn, I.S., Ryzhik, I.M., 2000. Table of Integrals, Series, and Products. Academic Press, San Diego.

Haupt, E., Schabe, H., 1992. A new model for a lifetime distribution with bathtub shaped failure rate. Microelectronic and Reliability 32, 633–639.

Hosking, J.R.M., 1990. *L*-moments: Analysis and estimation of distributions using linear combinations of order statistics. Journal of the Royal Statistical Society, Series B 52, 105–124.

Jones, M.C., 2004. Families of distribution arising from distribution of order statistics. Test 13, 1–43.

Lai, C.D., Xie, M., Murthy, D.N.P., 2003. A modified Weibull Distribution. IEEE Transactions Reliability 52, 33–37.

Lee, C., Famoye, F., Olumolade, O., 2007. Beta-Weibull distribution: Some properties and applications to censored data. Journal of Modern Applied Statistical Methods 6, 173–186.

Mudholkar, G.S, Srivastava, D.K., Friemer, M., 1995. The exponential Weibull family: A reanalysis of the bus-motor failure data. Technometrics 37, 436–445.

Mudholkar, G.S, Srivastava, D.K., Friemer, M., 1996. A generalization of the Weibull distribution with application to the analysis of survival data. Journal of the American Statistical Association 91, 1575–1583.

Nadarajah, S., 2008. Explicit expressions for moments of order statistics. Statistics and Probability Letters 78, 196–205.

Nadarajah, S., Gupta, A.K., 2004. The beta Fréchet distribution. Far East Journal of Theorical Statistics 15, 15–24.

Nadarajah, S., Kotz, S., 2004. The beta Gumbel distribution. Mathematical Problems in Engineering 10, 323–332.

Nadarajah, S., Kotz, S., 2005. The beta exponential distribution. Reliability Engineering and System Safety 91, 689–697.

Nelson, W.B., 1990. Accelerated Testing: Statistical Models, Test Plans and Data Analysis. Wiley, New York.

Pham, H., Lai, C.D., 2007. On recent generalizations of the Weibull distribution. IEEE Transactions on Reliability 56, 454–458.

Rajarshi, S., Rajarshi, M.B., 1988. Bathtub distributions: A review. Communication in Statistics — Theory and Methods 17, 2597–2621.

Trott, M., 2006. The Mathematica Guidebook for Symbolics. With 1 DVD-ROOM (Windows, Macintosh and UNIX). springer, New York.

Xie, M., Lai, C.D., 1995. Reliability analysis using an additive Weibull model with bathtub-shaped failure rate function. Reliability Engineering and System Safety 52, 87–93.

Xie, M., Tang, Y., Goh, T.N., 2002. A modified Weibull extension with bathtub failure rate function. Reliability Engineering and System Safety 76, 279–285.

Please cite this article in press as: Pescim, R.R., et al., The beta generalized half-normal distribution. Computational Statistics and Data Analysis (2009), doi:10.1016/j.csda.2009.10.007