Brazilian Journal of Probability and Statistics
2016, Vol. 30, No. 3, 366–384
DOI: 10.1214/15-BJPS284
© Brazilian Statistical Association, 2016

The extended generalized half-normal distribution

Jeniffer J. Duarte Sanchez, Wanessa W. da Luz Freitas and Gauss M. Cordeiro

Universidade Federal de Pernambuco

Abstract. Fatigue is a structural damage which occurs when a material is exposed to stress and tension fluctuations. We propose and study the extended generalized half-normal distribution for modeling skewed fatigue life data. The new model contains as special cases the half-normal and generalized half-normal (*Comm. Statist. Theory Methods* **37** (2008) 1323–1337) distributions. Various of its structural properties are derived, including the density function, moments, quantile and generating functions, mean deviations and order statistics. We investigate maximum likelihood estimation of the model parameters. An application illustrates the potentiality of the new distribution.

1 Introduction

Broadly speaking, there has been an increased interest in defining new generators for univariate continuous families of distributions by introducing one or more additional shape parameter(s) to a baseline distribution. This induction of parameter(s) has been proved useful in exploring tail properties and also for improving the goodness-of-fit of the family under study. One of these generators is the *exponentiated generalized* (EG) class of distributions pioneered by Cordeiro, Ortega and da Cunha (2013). For an arbitrary baseline cumulative distribution function (c.d.f.) $G(x)$, they defined the probability density function (p.d.f.) $f(x)$ and the c.d.f. $F(x)$ by

$$f(x) = ab\left[1 - G(x)\right]^{a-1}\left[1 - \{1 - G(x)\}^{a}\right]^{b-1} g(x) \tag{1.1}$$

and

$$F(x) = \left[1 - \{1 - G(x)\}^{a}\right]^{b}, \tag{1.2}$$

respectively, where $g(x) = dG(x)/dx$ and $a > 0$ and $b > 0$ are two extra shape parameters to those of the G distribution. If X is a random variable with density (1.1), we write $X \sim$ EG-G(a, b). Except for special choices of the functions $g(x)$ and $G(x)$, the density function $f(x)$ could be very difficult to deal in generality. One major benefit of the EG class is that it extends the exponentiated type distributions. In fact, this class generalizes both exponentiated Lehmann

Key words and phrases. Exponentiated generalized class, generalized half-normal distribution, Kwmaraswamy model, maximum likelihood estimation, survival function.
Received March 2014; accepted February 2015.

types I and II distributions (Cordeiro, Ortega and da Cunha, 2013). Several of its structural properties can be obtained from the exponentiated-G ("exp-G" for short) distribution.

A physical interpretation of (1.1) whenever a and b are positive integers can be given as follows. Consider a device made of b independent components in a parallel system. Further, each component is made of a independent subcomponents identically distributed according to $G(x)$ in a series system. The device fails if all b components fail and each component fails if any subcomponent fails. Then the time to failure of the device follows the distribution (1.1).

Generalizations of continuous distributions have been studied in recent decades. One reason for generalizing a well-known model is because its generalized form can accommodate non-monotone forms of the hazard rate function (h.r.f.). The fatigue process begins with an imperceptible fissure, the initiation, growth and propagation of which produces a dominant crack in the specimen due to cyclic patterns of stress, whose ultimate extension causes the rupture or failure of this specimen. The failure occurs when the total extension of the crack exceeds a critical threshold for the first time. 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 initial choices because of their negatively and positively skewed density shapes. However, they do not provide a reasonable parametric fit for modeling phenomenon with non-monotone failure rates. Cooray and Ananda (2008) defined the generalized half-normal (GHN) distribution derived from a model for static fatigue. The distributions for lifetime under fatigue which allow bathtub shaped failure rates are rather complex and usually require five or more parameters. It is important to describe the lifetime process under fatigue using simple generated distributions. Following this idea, we propose a simple distribution which extends the GHN and HN models.

The GHN p.d.f. (Cooray and Ananda, 2008) with shape parameter $\alpha > 0$ and scale parameter $\theta > 0$ is given by

$$g(z) = \sqrt{\frac{2}{\pi}}\left(\frac{\alpha}{z}\right)\left(\frac{z}{\theta}\right)^{\alpha}\exp\left[-\frac{1}{2}\left(\frac{z}{\theta}\right)^{2\alpha}\right], \qquad z > 0. \tag{1.3}$$

Its c.d.f. depends on the error function

$$G(z) = 2\Phi\left[\left(\frac{z}{\theta}\right)^{\alpha}\right] - 1 = \operatorname{erf}\left(\frac{(z/\theta)^{\alpha}}{\sqrt{2}}\right), \tag{1.4}$$

where

$$\Phi(z) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right)\right] \quad \text{and} \quad \operatorname{erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}\,dt.$$

The nth moment of (1.3) is given by Cooray and Ananda (2008) $E(Z^n) = \sqrt{\frac{2^{n/\alpha}}{\pi}}\Gamma(\frac{n+\alpha}{2\alpha})\theta^n$, where $\Gamma(\cdot)$ is the gamma function.

In this paper, we study a new four-parameter model, named the *extended generalized half-normal* (EGHN) distribution, whose p.d.f. is obtained by inserting the c.d.f. and p.d.f. of the GHN model in equation (1.1). We have

$$\begin{aligned} f(x) = ab\sqrt{\frac{2}{\pi}}\left(\frac{\alpha}{x}\right)\left(\frac{x}{\theta}\right)^{\alpha}\exp\left[-\frac{1}{2}\left(\frac{x}{\theta}\right)^{2\alpha}\right]\left\{2-2\Phi\left[\left(\frac{x}{\theta}\right)^{\alpha}\right]\right\}^{a-1} \\ \times\left[1-\left\{2-2\Phi\left[\left(\frac{x}{\theta}\right)^{\alpha}\right]\right\}^{a}\right]^{b-1}. \end{aligned} \tag{1.5}$$

If X is a random variable with p.d.f. (1.5), we write $X \sim \text{EGHN}(\alpha, \theta, a, b)$. The new model contains some important sub-models. For $a = b = 1$, it becomes the GHN distribution. For $\alpha = 1$, it gives the *extended half-normal* (EHN) distribution. For $b = 1$, it leads to the Lehmann type II generalized half-normal distribution. Further, if $a = b = 1$, in addition to $\alpha = 1$, it reduces to the HN distribution. The major benefit of the EGHN distribution for modeling fatigue lifetimes is that it can have bathtub failure rates and allow for greater flexibility of its tails. The two additional parameters can promote very different levels of asymmetry and kurtosis. Further, the EGHN model is a very competitive distribution to the beta generalized half-normal model (BGHN) pioneered by Pescim et al. (2010). In fact, we prove in Section 10 by means of a real data set that the EGHN model provides a better fit than the BGHN model.

The c.d.f. and h.r.f. corresponding to (1.5) are

$$F(x) = \left[1-\left\{1-\text{erf}\left(\frac{(x/\theta)^{\alpha}}{\sqrt{2}}\right)\right\}^{a}\right]^{b}, \tag{1.6}$$

and

$$\begin{aligned} \tau(x) = \Bigg(ab\sqrt{\frac{2}{\pi}}\left(\frac{\alpha}{x}\right)\left(\frac{x}{\theta}\right)^{\alpha}\exp\left[-\frac{1}{2}\left(\frac{x}{\theta}\right)^{2\alpha}\right]\left\{2-2\Phi\left[\left(\frac{x}{\theta}\right)^{\alpha}\right]\right\}^{a-1} \\ \times\left[1-\left\{2-2\Phi\left[\left(\frac{x}{\theta}\right)^{\alpha}\right]\right\}^{a}\right]^{b-1}\Bigg) \\ \times\left(1-\left[1-\left\{1-\text{erf}\left(\frac{(x/\theta)^{\alpha}}{\sqrt{2}}\right)\right\}^{a}\right]^{b}\right)^{-1}, \end{aligned} \tag{1.7}$$

respectively. Plots of p.d.f. and h.r.f. of X for selected parameter values are displayed in Figures 1 and 2, respectively. The plots in Figure 2 indicate that the GHN distribution can have bathtub failure rates.

Let $x = Q(u)$ be the EGHN quantile function (q.f.) derived by inverting (1.6). If $x = Q_N(u) = \Phi^{-1}(u)$ denotes the standard normal q.f., we can obtain

$$Q(u) = \theta\, Q_N\left(\frac{2-[1-u^{1/b}]^{1/a}}{2}\right)^{1/\alpha}. \tag{1.8}$$

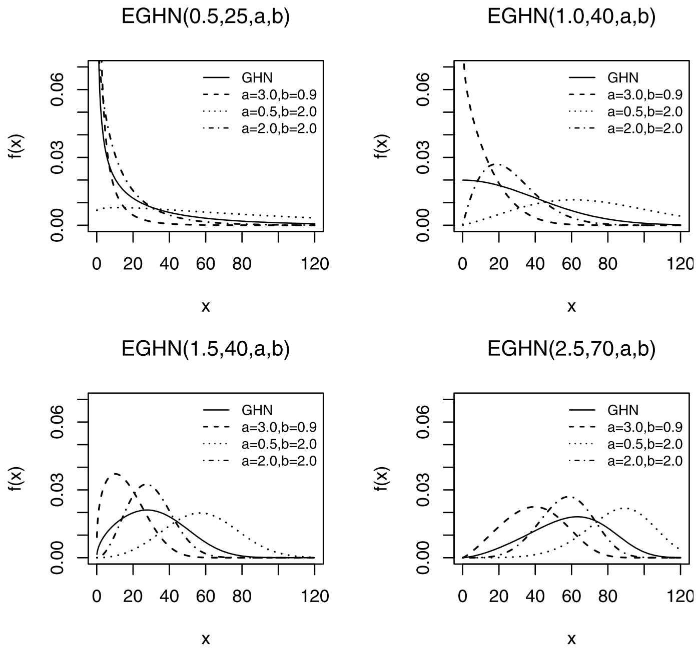

Figure 1 *Plots of the p.d.f.* (1.5) *for some parameter values.*

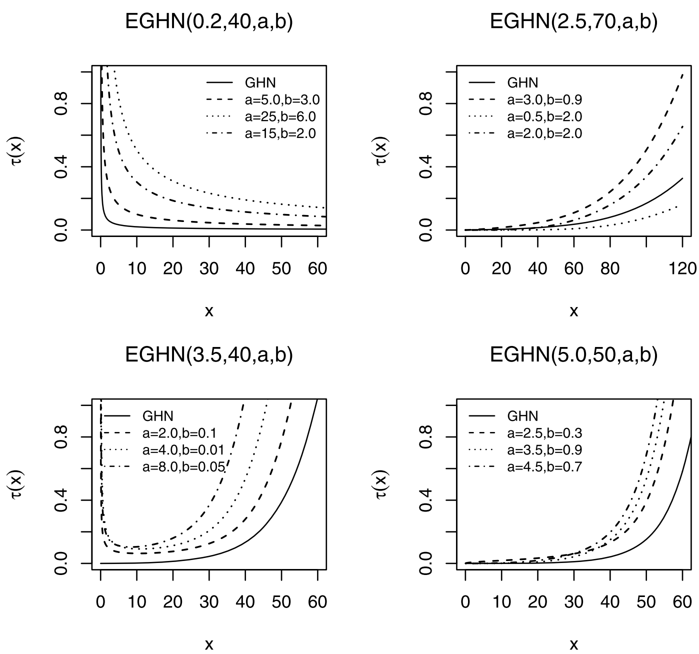

Figure 2 *Plots of the h.r.f.* (1.7) *for some parameter values.*

Clearly, the EGHN distribution is easily simulated by $X = Q(U)$, where U is a uniform random variable on the unit interval (0, 1). The paper is organized as follows. In Section 2, we derive an expansion for the EGHN density function. In Section 3, we study the behavior of the Bowley skewness and Moors kurtosis. In Section 4, we obtain the moments of X. We derive power series expansions for the quantile and generating functions in Sections 5 and 6, respectively. Incomplete moments and mean deviations are determined in Section 7. Expansions for the density of the order statistics and their moments are given in Section 8. Maximum likelihood estimation is discussed in Section 9. In Section 10, we illustrate the importance of the new distribution applied to a real data set. Finally, concluding remarks are addressed in Section 11.

2 Expansion for the density function

For any real non-integer β, we consider the power series

$$(1-z)^{\beta-1} = \sum_{k=0}^{\infty} (-1)^k \binom{\beta-1}{k} z^k, \tag{2.1}$$

which is valid for $|z| < 1$. Applying (2.1) in equation (1.2) twice gives

$$F(x) = \sum_{j=0}^{\infty} w_{j+1} H_{j+1}(x), \tag{2.2}$$

where $H_{j+1}(x) = G(x)^{j+1}$ and the coefficients w_j are

$$w_{j+1} = w_{j+1}(a,b) = \sum_{k=1}^{\infty} (-1)^{k+j+1} \binom{b}{k} \binom{ka}{j+1}.$$

Equation (2.2) gives the generated c.d.f. $F(x)$ distribution as a linear combination of exp-G c.d.f.s with positive powers $1, 2, \ldots$. By differentiating (2.2), we obtain

$$f(x) = \sum_{j=0}^{\infty} w_{j+1} h_{j+1}(x), \tag{2.3}$$

where $h_{j+1}(x) = (j+1) g(x) G(x)^j$ is the exp-G p.d.f. with power parameter $j+1$.

Equation (2.3) reveals that the EG p.d.f. is a linear combination of exp-G p.d.f.s. Thus, some structural properties of the EG class can be obtained from well-established properties of the exp-G distributions, which have been studied by many authors (see Nadarajah and Kotz, 2006).

Let T_{j+1} be a random variable having the exponentiated-GHN (exp-GHN) distribution with power parameter $j+1$. The density function of T_{j+1} is obtained from (1.3) and (1.4) as

$$h_{(j+1)}(t) = (j+1)\sqrt{\frac{2}{\pi}}\left(\frac{\alpha}{t}\right)\left(\frac{t}{\theta}\right)^{\alpha}\exp\left[-\frac{1}{2}\left(\frac{t}{\theta}\right)^{2\alpha}\right]\operatorname{erf}\left(\frac{(t/\theta)^{\alpha}}{\sqrt{2}}\right)^{j}. \tag{2.4}$$

Based on the density function (2.3), the p.d.f. of X can be expressed as

$$f(x) = g(x)\sum_{r=0}^{\infty} t_{r+1}\left\{\operatorname{erf}\left(\frac{(x/\theta)^{\alpha}}{\sqrt{2}}\right)\right\}^{r}, \tag{2.5}$$

where $t_{r+1} = t_{r+1}(a,b) = (r+1)w_{r+1}$. Equations (2.4) and (2.5) are the main expansions to obtain structural properties of the EGHN model. They and other expansions in this paper can be evaluated in classic symbolic computation software which have the ability to deal with complex expressions.

3 Quantile measures

The effect of the shape parameters a and b on the skewness and kurtosis of the EGHN distribution can be investigated based on quantile measures determined from (1.8). The Bowley skewness (Kenney and Keeping, 1962) is one of the earliest skewness measures defined by

$$B = \frac{Q(3/4) + Q(1/4) - 2Q(1/2)}{Q(3/4) - Q(1/4)}.$$

The Moors kurtosis (see Moors, 1998) based on octiles is defined by

$$M = \frac{Q(7/8) - Q(5/8) + Q(3/8) - Q(1/8)}{Q(6/8) - Q(2/8)}.$$

Figures 3 and 4 display plots of the measures B and M for the EGHN distribution for some parameter values. The upper plots indicate (b fixed) that the Bowley skewness can increase or decrease, and then increase for increasing values of a, whereas the under plots indicate (a fixed) that the Bowley skewness can decrease, or decrease and then increase for increasing values of a. On the other hand, the Moors kurtosis can increase or decrease, and then increase (b fixed) for increasing values of a, and decrease and then increase, or, increase and then decrease (a fixed) for increasing values of b. So, these plots indicate that both measures can be very sensitive to the extra shape parameters, thus indicating the importance of the EGHN distribution.

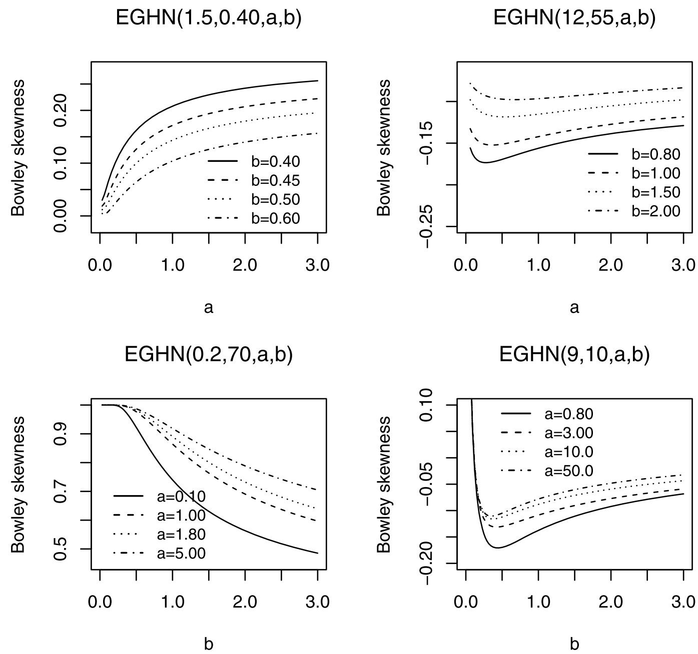

Figure 3 *The EGHN Bowley skewness as a function of b for some values of a and as a function of a for some values of b.*

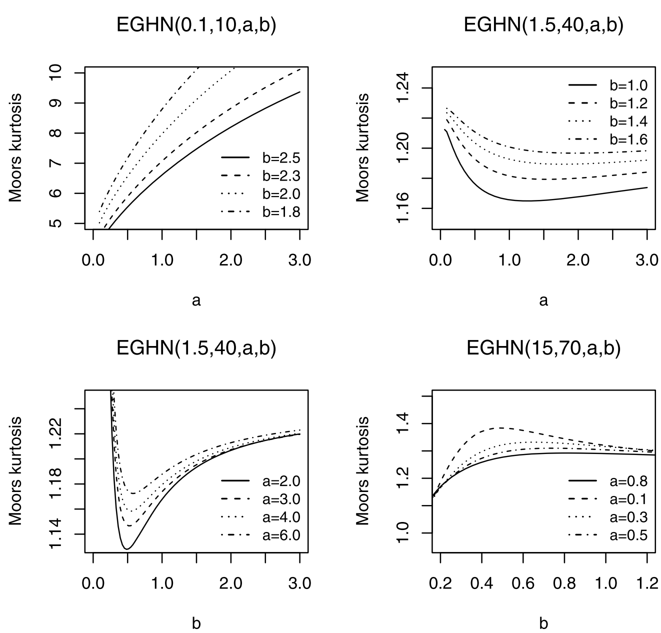

Figure 4 *The EGHN Moors kurtosis as a function of b for some values of a and as a function of a for some values of b.*

4 Moments

The nth moment of X follows from (2.5) by setting $u=(x/\theta)^\alpha$ as

$$\mathrm{E}(X^n)=\theta^n\sqrt{\frac{2}{\pi}}\sum_{r=0}^{\infty}t_{r+1}I\left(\frac{n}{\alpha},r\right),$$

where

$$I\left(\frac{n}{\alpha},r\right)=\int_0^\infty u^{n/\alpha}\mathrm{e}^{-u^2/2}\left[\mathrm{erf}\left(\frac{u}{\sqrt{2}}\right)\right]^r du.$$

Then

$$\mathrm{E}(X^n)=\theta^n\sqrt{\frac{2}{\pi}}\sum_{r=0}^{\infty}t_{r+1}I\left(\frac{n}{\alpha},r\right), \tag{4.1}$$

where

$$\begin{aligned} I\left(\frac{n}{\alpha},r\right) &= \pi^{-r/2}2^{r+(n/(2\alpha))-1/2} \\ &\quad\times\sum_{m_1,\ldots,m_r=0}^{\infty}\frac{(-1)^{m_1+\cdots+m_r}\Gamma(m_1+\cdots+m_r+(r+n/\alpha+1)/2)}{(m_1+1/2)\cdots(m_r+1/2)m_1!\cdots m_r!}. \end{aligned} \tag{4.2}$$

Further, if $r+\frac{n}{\alpha}$ is even, the integral $I(\frac{n}{\alpha},r)$ can be expressed in terms of the Lauricella function of type A (Aarts, 2000) defined by

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

where $(a)_k=a(a+1)\cdots(a+k-1)$ 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 Mathematica (Trott, 2006). Hence, $\mathrm{E}(X^n)$ can be given in terms of the Lauricella functions of type A

$$\mathrm{E}(X^n)=\theta^n\sqrt{\frac{2}{\pi}}\sum_{r=0}^{\infty}C_rF_A^{(r)}\left(\frac{r+n/\alpha+1}{2};\frac{1}{2},\ldots,\frac{1}{2};\frac{3}{2},\ldots,\frac{3}{2};-1,\ldots,-1\right),$$

where

$$C_r=t_{r+1}\pi^{-r/2}2^{2r+n/(2\alpha)-1/2}\Gamma\left(\frac{r+n/\alpha+1}{2}\right).$$

This equation is an infinite sum of Lauricella functions of type A which vanish when $r+n/\alpha$ is odd.

An alternative expression for $\mathrm{E}(X^n)$ can be obtained from (2.3) as

$$\mathrm{E}(X^n) = \sum_{j=0}^{\infty} w_{j+1} \mathrm{E}(T_{j+1}^n), \tag{4.3}$$

where $T_{j+1} \sim$ exp-GHN$(j+1)$ and w_{j+1} is defined in Section 2.

Equation (4.3) gives the EGHN moments in terms of an infinite linear combination of exp-GHN moments.

5 Quantile expansion

Here, we obtain a power series for the q.f. of X. By expanding the binomial terms in (1.8), we can write

$$\frac{1}{2}\{2 - (1 - u^{1/b})]^{1/a}\} = \sum_{k=0}^{\infty} m_k u^{k/b}, \tag{5.1}$$

where $m_0 = 1/2$ and $m_k = [(-1)^k \binom{1/a}{k+1}]/2$ for $k \geq 1$.

Following Steinbrecher (2002), the standard normal q.f. can be expanded as

$$Q_N(u) = \sum_{k=0}^{\infty} b_k w^{2k+1}, \tag{5.2}$$

where $w = \sqrt{2\pi}(u - 1/2)$ and the quantities b_k can be calculated recursively from

$$b_{k+1} = \frac{1}{2(2k+3)} \sum_{r=0}^{k} \frac{(2r+1)(2k-2r+1) b_r b_{k-r}}{(r+1)(2r+1)}.$$

Here, $b_0 = 1$, $b_1 = 1/6$, $b_2 = 7/120$, $b_3 = 127/7560, \ldots$. The function $Q_N(u)$ can be expressed as a power series

$$Q_N(u) = \sum_{r=0}^{\infty} d_r u^r, \tag{5.3}$$

where

$$d_r = \sum_{k=r}^{\infty} \left(\frac{-1}{2}\right)^{k-r} \binom{k}{r} e_k,$$

and the quantities e_k are defined from the coefficients in (5.2) by $e_k = 0$ for $k = 0, 2, 4, \ldots$ and $e_k = (2\pi)^{k/2} b_{(k-1)/2}$ for $k = 1, 3, 5, \ldots$.

Combining (5.1) and (5.3), we obtain

$$Q_N(\{2 - [1 - u^{1/b}]^{1/a}\}/2) = \sum_{r=0}^{\infty} d_r \left(\sum_{k=0}^{\infty} m_k u^{k/b}\right)^r. \tag{5.4}$$

We use throughout an equation of Gradshteyn and Ryzhik (2000) for a power series raised to a positive integer j

$$\left(\sum_{i=0}^{\infty} a_i x^i\right)^j = \sum_{i=0}^{\infty} c_{j,i} x^i, \tag{5.5}$$

where the coefficients $c_{j,i}$ (for $i = 1, 2, \ldots$) are determined from the recurrence equation

$$c_{j,i} = (i a_0)^{-1} \sum_{m=1}^{i} [m(j+1) - i] a_m c_{j,i-m}, \tag{5.6}$$

and $c_{j,0} = a_0^j$. From equations (5.4) and (5.5), we have

$$Q_N(\{2 - [1 - u^{1/b}]^{1/a}\}/2) = \sum_{k=0}^{\infty} p_k u^{k/b},$$

where $p_k = \sum_{r=0}^{\infty} d_r g_{r,k}$ and the quantities $g_{r,k}$ come from (5.6) as $g_{r,0} = m_0^r$ and $g_{r,k} = (k m_0)^{-1} \sum_{s=1}^{k} [s(r+1) - k] m_s g_{r,k-s}$ for $k \geq 1$. The argument of the standard normal q.f. implies that the sum $\sum_{k=0}^{\infty} p_k u^{k/b}$ belongs to the interval $(-4, 4)$. Setting $h_k = p_k/5$, the EGHN q.f. becomes

$$Q(u) = \theta 5^{1/\alpha} \left(\sum_{k=0}^{\infty} h_k u^{k/b}\right)^{1/\alpha}, \tag{5.7}$$

and then it does involve a power series in the interval $(0, 1)$. We can obtain an expansion for $G(x)^\beta$ ($\beta > 0$ real non-integer) given by

$$G(x)^\beta = \sum_{r=0}^{\infty} s_r(\beta) G(x)^r, \tag{5.8}$$

where

$$s_r(\beta) = \sum_{j=r}^{\infty} (-1)^{r+j} \binom{\beta}{j} \binom{j}{r}.$$

Using (5.8), we can write

$$\left(\sum_{k=0}^{\infty} h_k u^{k/b}\right)^{1/\alpha} = \sum_{r=0}^{\infty} s_r(\alpha^{-1}) \left(\sum_{k=0}^{\infty} h_k u^{k/b}\right)^r,$$

where $s_r(\alpha^{-1}) = \sum_{j=r}^{\infty} (-1)^{r+j} \binom{\alpha^{-1}}{j} \binom{j}{r}$. Combining this equation and (5.5) gives a very neat way of writing the q.f. of X, namely

$$Q(u) = \sum_{k=0}^{\infty} v_k u^{k/b}, \tag{5.9}$$

where $v_k = \theta 5^{1/\alpha} \sum_{r=0}^{\infty} s_r(\alpha^{-1}) q_{r,k}$ for $k \geq 0$, $q_{r,k} = (kh_0)^{-1} \sum_{m=1}^{k} [m(r+1) - k] h_m q_{r,k-m}$ for $k \geq 1$ and $q_{r,0} = h_0^r$. Equation (5.9) is the main result of this section.

6 Generating function

The moment generating function (mgf) of X can be obtained from (2.5) by setting $u = (x/\theta)^\alpha$ as

$$M(s) = \sqrt{\frac{2}{\pi}} \sum_{r=0}^{\infty} t_{r+1} \sum_{m=0}^{\infty} \frac{\theta^m s^m}{m!} \int_0^{\infty} u^{m/\alpha} \mathrm{e}^{-u^2/2} \left[\mathrm{erf}\left(\frac{u}{\sqrt{2}} \right) \right]^r du.$$

Following similar lines of Section 4, $M(s)$ can be expanded as

$$M(s) = \sum_{m=0}^{\infty} \frac{A_m s^m}{m!}, \tag{6.1}$$

where

$$A_m = \sqrt{\frac{2}{\pi}} \theta^m \sum_{r=0}^{\infty} t_{r+1} \pi^{-r/2} 2^{r+m/(2\alpha)-1/2} I\left(\frac{m}{\alpha}, r \right),$$

where $I(\frac{m}{\alpha}, r)$ is given by (4.2). Evidently, A_m denotes a second representation for the mth moment of X.

An alternative equation for $M(s)$ can be derived using equation (5.9). We can write from (5.5)

$$M(s) = \int_0^1 \exp\{s\, Q(u)\}\, du = \int_0^1 \sum_{k=0}^{\infty} \frac{s^k (\sum_{n=0}^{\infty} v_n u^{n/b})^k}{k!} du = \sum_{k=0}^{\infty} \frac{B_k s^k}{k!}, \tag{6.2}$$

where $B_k = \sum_{n=0}^{\infty} d_{k,n}/(n+1)$ for $k = 0, 1, \ldots$, and $d_{k,n}$ can be determined from (5.6) as $d_{k,0} = v_0^k$ and $d_{k,n} = (kv_0)^{-1} \sum_{m=1}^{n} [m(k+1) - n] v_m d_{k,n-m}$ for $k \geq 1$. Clearly, B_k gives a third representation for the kth moment of X. Equations (6.1) and (6.2) are the main results of this section.

7 Incomplete moments

The nth incomplete moments of X is given by $J(q,n) = \int_0^q x^n f(x)\, dx$. This integral can be obtained from (2.5) by setting $u = (x/\theta)^\alpha$ as

$$J(q,n) = \theta^n \sqrt{\frac{2}{\pi}} \sum_{r=0}^{\infty} t_{r+1} \int_0^{(q/\theta)^\alpha} u^{n/\alpha} \mathrm{e}^{-u^2/2} \left[\mathrm{erf}\left(\frac{u}{\sqrt{2}} \right) \right]^r du.$$

The error function $\mathrm{erf}(\frac{u}{\sqrt{2}})$ admits the expansion $\sum_{k=0}^{\infty} a_k u^k$, where $a_{2k+1} = \frac{(-1)^k 2^{1-k}}{\sqrt{2\pi}(2k+1)k!}$ and $a_{2k} = 0$ for $k \in \mathbb{N}$. Thus,

$$J(q,n) = \theta^n \sqrt{\frac{2}{\pi}} \sum_{r,k=0}^{\infty} t_{r+1} c_{r,k} \int_0^{(q\theta^{-1})^\alpha} u^{k+n/\alpha} \mathrm{e}^{-u^2/2}\, du,$$

where the quantities $c_{r,k}$ are determined from the a_k's above using (5.6). Setting $v = u^2/2$, we write

$$J(q,n) = \frac{\theta^n}{\sqrt{\pi}} \sum_{r,k=0}^{\infty} t_{r+1} 2^{(k+n\alpha^{-1})/2} c_{r,k} \int_0^{(q\theta^{-1})^{2\alpha}/2} v^{(k+n\alpha^{-1}-1)/2} \mathrm{e}^{-v}\, dv.$$

For $\lambda > 0$,

$$\int_0^x v^{\lambda-1} \mathrm{e}^{-\alpha v}\, dv = \alpha^{-\lambda} \gamma(\lambda, \alpha x),$$

where $\gamma(a,x) = \int_0^x t^{a-1} \mathrm{e}^{-t}\, dt$ is the incomplete gamma function and then

$$J(q,n) = \frac{\theta^n}{\sqrt{\pi}} \sum_{r,k=0}^{\infty} t_{r+1} 2^{(k+n\alpha^{-1})/2} c_{r,k} \gamma\left[(k+n\alpha^{-1}+1)/2, \left(q\theta^{-1}\right)^{2\alpha}/2\right]. \quad (7.1)$$

We can derive the mean deviations about the mean $\mu = \mathrm{E}(X)$ and about the median M from

$$\delta_1 = 2\left[\mu F(\mu) - J(\mu,1)\right] \quad \text{and} \quad \delta_2 = \mu - 2J(M,1), \quad (7.2)$$

where M is the solution of the non-linear equation $Q(M) = 1/2$ obtained from (1.8) and $J(q,n)$ is given by (7.1). From equations (7.1) and (7.2), we obtain the mean deviations. The first incomplete moment gives the Bonferroni and Lorenz curves which have applications in several areas. For a given probability π, they are defined by $B(\pi) = J(q,1)/(\pi\mu)$ and $L(\pi) = J(q,1)/\mu$, respectively, where $q = Q(\pi)$. Using (7.1) with $n = 1$, we have $B(\pi)$ and $L(\pi)$.

Figure 5 displays the $B(\pi)$ and $L(\pi)$ curves. The $B(\pi)$ curve can increase or decrease depending on the parameter values. On the other hand, the $L(\pi)$ curve increases for any parameters values.

8 Order statistics

The density function $f_{i:n}(x)$ of the ith order statistic, for $i = 1, \ldots, n$, from i.i.d. random variables $X_1, \ldots, X_n$ following any EG-G distribution is simply given by

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

where $B(\cdot,\cdot)$ denotes the beta function.

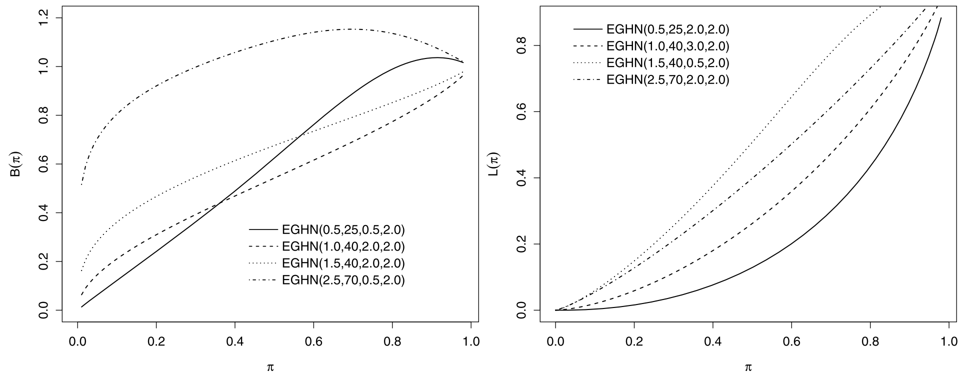

Figure 5 *The $B(\pi)$ and $L(\pi)$ curves of the EGHN model.*

Cordeiro, Ortega and da Cunha (2013) presented the density of the EG-G order statistics as a linear combination of exp-G densities. This result enables us to derive the ordinary moments of the order statistics as an infinite weighted sum of probability weighted moments (PWMs) of the G distribution. They demonstrated that

$$f_{i:n}(x) = \frac{ab}{B(i, n-i+1)} g(x) \sum_{l=0}^{\infty} s_l G(x)^l, \tag{8.1}$$

where the coefficients s_l are given by

$$s_l = \sum_{k=0}^{n-i} \sum_{r=0}^{\infty} (-1)^{k+r+l} \binom{n-i}{k} \binom{(i+k)b-1}{r} \binom{(r+1)a-1}{l}.$$

Equation (8.1) can be rewritten in terms of the exp-G density functions as

$$f_{i:n}(x) = \frac{ab}{B(i, n-i+1)} \sum_{l=0}^{\infty} v_{l+1} h_{l+1}(x), \tag{8.2}$$

where $v_{l+1} = s_l/(l+1)$.

For example, the sth moment of the order statistics, say $\mathrm{E}(X_{i:n}^s)$, can be obtained from (8.2) as

$$\mathrm{E}(X_{i:n}^s) = \frac{ab}{B(i, n-i+1)} \sum_{l=0}^{\infty} v_{l+1} \mathrm{E}(T_{l+1}^s),$$

where $T_{l+1} \sim \mathrm{EGHN}(l+1)$ and the constants v_{l+1} are defined in (2.4) and (8.2), respectively. Clearly, equations (8.1) and (8.2) should be used numerically with a large number instead of infinity.

9 Estimation and inference

We consider the estimation of the parameters of the proposed model by the method of maximum likelihood. Several approaches for parameter point estimation were proposed in the literature but the maximum likelihood method is the most commonly employed. The maximum likelihood estimates (MLEs) enjoy desirable properties and can be used when constructing confidence intervals and regions and also in test statistics. Large sample theory for these estimates delivers simple approximations that work well in finite samples. Let Y have the EGHN distribution with vector of parameters $\boldsymbol{\lambda} = (\alpha, \theta, a, b)^T$. The log-likelihood for the model parameters from a single observation y of Y is given by

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

The components of the unit score vector $\mathbf{U} = (\frac{\partial \ell}{\partial \alpha}, \frac{\partial \ell}{\partial \theta}, \frac{\partial \ell}{\partial a}, \frac{\partial \ell}{\partial b})^T$ are given in the Appendix.

For a sample $y = (y_1, \dots, y_n)^T$ 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 ($i = 1, \dots, n$). 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 $i = 1, \dots, n$. The MLE $\hat{\boldsymbol{\lambda}}$ of $\boldsymbol{\lambda}$ is the solution of the system of non-linear equations $\mathbf{U}_n = \mathbf{0}$. These equations cannot be solved analytically, and statistical softwares are required to solve them numerically. It is usually more convenient to adopt non-linear optimization algorithms such as quasi-Newton algorithm to numerically maximize the log-likelihood function or functions such as the BFGS, L-BFGS-B, Nelder–Mead and simulated annealing methods. Alternatively, we can use the `AdequacyModel` script version 1.0.8 available for the programming language `R`. The script is currently maintained by one of the authors of this paper and more information can be obtained from http://cran.rstudio.com/web/packages/AdequacyModel/index.html. The package is distributed under the terms of the licenses GNU General Public License (GPL-2 or GPL-3). We can take as starting values the estimates for the HN model. The final MLEs are usually robust to these initial values.

For interval estimation and tests of hypotheses on the parameters in $\boldsymbol{\lambda}$, we require the 4×4 unit observed information matrix $\mathbf{K} = \mathbf{K}(\boldsymbol{\lambda}) = \{\kappa_{ij}\}$, where

$i, j = \alpha, \theta, a, b$, whose elements can be computed numerically. The method of the resampling bootstrap can be used for correcting the biases of the MLEs of the model parameters. Good interval estimates may also be obtained through the bootstrap percentile method.

The estimated multivariate normal $\mathrm{N}_4(\mathbf{0}, n^{-1}K(\hat{\boldsymbol{\lambda}})^{-1})$ distribution can be used to construct approximate confidence intervals for the model parameters. An asymptotic confidence interval with significance level γ for each parameter λ_r is given by

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

where $\hat{\kappa}^{\lambda_r,\lambda_r}$ is the rth diagonal element of $K(\hat{\boldsymbol{\lambda}})^{-1}$, 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 comparing the new distribution with some of its special models. For example, we may use the LR statistic to check if the fit using the EGHN distribution is statistically "superior" to a fit using the GHN distribution for a given data set. In any case, considering the partition $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1^T, \boldsymbol{\lambda}_2^T)^T$, 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 using 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, $w \xrightarrow{d} \chi^2_q$, 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^2_q distribution.

10 Application

Here, for the purpose of illustration, we analyze the data given by Sharafi and Behboodian (2008), which were compiled in 1971 by a large insurance company in order to investigate its selection procedures for claims adjusters. The data set concerns OTIS IQ Scores for 52 minority (non-white) males hired by the company. One of the key questions of the study was the predictability of job performance when the OTIS test was applied. Following them, we fit the GHN, BGHN, Kumaraswamy generalized half-normal (KwGHN) and EGHN distributions to these data.

The BGHN p.d.f. with the four parameters α, θ a and b, say BGHN(α, θ, a, b), is defined by

$$\pi_1(x) = \frac{\sqrt{2/\pi}\,(\alpha/x)(x/\theta)^\alpha \mathrm{e}^{-(1/2)(x/\theta)^{2\alpha}}}{B(a,b)} \times \left[2\Phi\left(\left(\frac{x}{\theta}\right)^\alpha\right) - 1\right]^{a-1}\left[2 - 2\Phi\left(\left(\frac{x}{\theta}\right)^\alpha\right)\right]^{b-1},$$

whereas the KwGHN p.d.f. with the same parameters is given by

$$\pi_2(x) = ab\sqrt{\frac{2}{\pi}}\left(\frac{\alpha}{x}\right)\left(\frac{x}{\theta}\right)^{\alpha}\exp\left[-\frac{1}{2}\left(\frac{x}{\theta}\right)^{2\alpha}\right]\left\{2\Phi\left[\left(\frac{x}{\theta}\right)^{\alpha}\right]-1\right\}^{a-1}$$
$$\times\left[1-\left\{2\Phi\left[\left(\frac{x}{\theta}\right)^{\alpha}\right]-1\right\}^{a}\right]^{b-1}.$$

We fit the above models to the current data and compute the MLEs, their standard errors (given in parentheses) and the following statistics: Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Consistent Akaike Information Criterion (CAIC) and Crámer–von Mises (W) and Anderson–Darling (A) (Chen and Balakrishnan, 1995). The computations are performed using the `AdequacyModel` script in R package. We also perform a non-parametric bootstrap simulation with 500 trials to compute the biases of the MLEs and of the statistics AIC, BIC, CAIC, W and A. Table 1 provides the MLEs, SEs, the biases of the estimates and of the test statistics from the fitted EGHN, KwGHN, BGHN and GHN models. The better the fit of the model, the smaller the values of the corresponding statistics. The biases of the MLEs and of the test statistics are also reported in this table. The results indicate that the EGHN model has the smallest values of these statistics among all fitted models. So, it could be chosen as the best fitted model. The biases of the test statistics are quite small.

Further, we compute the corrected W^* and A^* statistics (Chen and Balakrishnan, 1995) given by:

$$W^* = W^2(1+0.5/n) \quad \text{and} \quad A^* = A^2(1+0.75/n+2.25/n^2).$$

The corrected versions of the statistics W^* and A^* are adopted since they are more

Table 1 *MLEs, their SEs and biases, and some statistics for the fitted models to the current data*

Model	a	b	α	θ	AIC	BIC	CAIC	W	A
EGHN	0.225	15.287	2.644	58.299	373.727	381.532	374.578	0.109	0.652
(SE)	(0.045)	(5.629)	(0.103)	(0.078)					
Bias	0.238	−0.567	0.449	6.426	0.303	0.303	0.303	0.042	0.283
KwGHN	5.077	0.041	1.186	20.909	507.776	515.581	508.627	0.131	0.759
(SE)	(0.031)	(0.006)	(0.001)	(0.007)					
Bias	0.000	0.004	−0.005	0.004	−0.369	−0.369	−0.369	0.044	0.279
BGHN	7.874	0.040	0.767	8.289	549.782	557.587	550.632	0.119	0.701
(SE)	(5.244)	(0.005)	(0.004)	(0.003)					
Bias	0.440	0.006	0.070	1.559	−5.104	−5.104	−5.104	0.048	0.309
GHN	1.000	1.000	9.956	112.425	383.409	387.311	383.654	0.382	2.150
(SE)	–	–	(1.096)	(1.238)					
Bias	–	–	0.242	0.294	−1.415	−1.415	−1.415	0.041	0.262

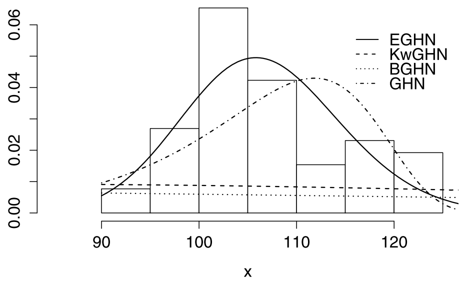

Figure 6 *Histogram of the data and four fitted densities.*

appropriate to discriminate among non-nested models fitted to the same data set and become more widely used nowadays. Then we obtain

- EGHN: $W^* = 0.0044$, $A^* = 0.1379$;
- KwGHN: $W^* = 0.0076$, $A^* = 0.2333$;
- BGHN: $W^* = 0.0050$, $A^* = 0.1559$;
- GHN: $W^* = 0.1170$, $A^* = 3.618$.

We can note that the GHN distribution gives the worst fit, and the EGHN distribution provides the best fit among the four fitted models.

Figure 6 displays the histogram of the data and the fitted EGHN, KwGHN, BGHN and GHN densities.

A non-parametric bootstrap simulation is performed to verify the precision of the estimates. The initial values used for each distribution in each bootstrap replication are the same. Based on the biases and standard errors of the estimates, we can conclude that the estimates are consistent and independent of the bootstrap replication.

11 Conclusions

We propose a new four-parameter model named the extended generalized half normal (EGHN) distribution to extend the generalized half-normal (GHN) distribution. We derive an expansion for its density function and explicit expressions for the ordinary moments, quantile and generating functions, incomplete moments and order statistics. The model parameters are estimated by maximum likelihood. An application of the new distribution to a real data set demonstrates that it can be used quite effectively to provide better fits than other competing models. We hope that this generalization may attract wider applications in the literature of the fatigue life distributions.

Appendix: The components of the score function

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

where $v = \exp[-\frac{1}{2}(\frac{y}{\theta})^{2\alpha}](\frac{y}{\theta})^{\alpha}$.

Acknowledgments

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

References

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

Chen, G. and Balakrishnan, N. (1995). A general purpose approximate goodness-of-fit test. *Journal of Quality Technology* **27**, 154–161.

Cooray, K. and Ananda, M. M. A. (2008). A generalization of the half-normal distribution with applications to lifetime data. *Communications in Statistics. Theory and Methods* **37**, 1323–1337. MR2526464

Cordeiro, G. M., Ortega, E. M. M. and da Cunha, D. C. C. (2013). The exponentiated generalized class of distributions. *Journal of Data Science* **11**, 1–27. MR2963349

Gradshteyn, I. S. and Ryzhik, I. M. (2000). *Table of Integrals, Series, and Products*, 6th ed. San Diego, CA: Academic Press. MR1773820

Kenney, J. F. and Keeping, E. S. (1962). *Mathematics of Statistics, Pt. 1*, 3rd ed. Princeton, NJ: Van Nostrand.

Moors, J. J. A. (1998). A quantile alternative for kurtosis. *Journal of the Royal Statistical Society. Series D. The Statistician* **37**, 25–32.

Nadarajah, S. and Kotz, S. (2006). The exponentiated-type distributions. *Acta Applicandae Mathematicae* **92**, 97–111. MR2265333

Sharafi, M. and Behboodian, J. (2008). The Balakrishnan skew-normal density. *Statistical Papers* **49**, 769–778. MR2439018

Pescim, R. R., Demétrio, C. G. B., Cordeiro, G. M., Ortega, E. M. M. and Urbano, M. R. (2010). The beta generalized half-normal distribution. *Computational Statistics and Data Analysis* **54**, 945–957. MR2580929

Steinbrecher, G. (2002). Taylor expansion for inverse error function around origin. Working paper, Univ. Craiova.

Trott, M. (2006). *The Mathematica Guidebook for Symbolics*. New York: Springer. MR2208600

J. J. Duarte Sanchez
W. W. da Luz Freitas
G. M. Cordeiro
Departamento de Estatística
Universidade Federal de Pernambuco
Avenida Prof. Moraes Rego 1235
Cidade Universitária
Recife, PE 50670-901
Brasil
E-mail: jjduartes@unal.edu.co
wanyweridiana@gmail.com
gauss@de.ufpe.br