CompAMa
www.compama.co.usb.ve
Bull. Comput. Appl. Math., Vol.6, No.1, 2018
ISSN 2244-8659

A new flexible extension of the generalized half-normal lifetime model with characterizations and regression modeling

Emrah Altun [1], *Haitham M. Yousof* [2], *G.G. Hamedani* [3].

CompAMa Vol.6, No.1, pp.83-115, 2018 - Accepted February 16, 2018

Abstract

In this paper, we introduced a new flexible extension of the Generalized Half-Normal lifetime model as well as a new log-location regression model based on the proposed model. Some useful characterization results are presented and some mathematical properties are derived. The maximum likelihood method is used to estimate the model parameters by means of a graphical Monte Carlo simulation study. We show that the new log-location regression lifetime model can be very useful in analysing real data and provide more realistic fits than other regression models. Index plot of the modified deviance residual and Q-Q plot for modified deviance residual are presented to illustrate that our new model is more appropriate to HIV data set than other competitive models like log-odd log-logistic generalized half-normal regression model and log-generalized half-normal regression model. The sensitivity analysis is used via the index plot of generalized cook distance to discover the possible influential observations.

Keywords: Regression model, generalized Cook distance, residual analysis, influential diagnostics, Monte Carlo simulation.

[1]Department of Statistics, Hacettepe University, Turkey (emrahaltun@hacettepe.edu.tr).

[2]Department of Statistics, Mathematics and Insurance, Benha University, Benha, Egypt (haitham.yousof@fcom.bu.edu.eg).

[3]Department of Mathematics, Statistics and Computer Science, Marquette University, USA. (gholamhoss.hamedani@marquette.edu).

1 Introduction

The Generalized Half-Normal (GHN) density function (Cooray and Ananda [1]) with shape parameter $\lambda > 0$ and scale parameter $\theta > 0$ is given by (for $x > 0$)

$$g(x;\lambda,\theta) = \sqrt{\frac{2}{\pi}}\,(\lambda/x)\,(x/\theta)^{\lambda}\exp\left[-\frac{1}{2}\,(x/\theta)^{2\lambda}\right].$$

The corresponding cumulative distribution function (cdf) depends on the error function

$$G(x;\lambda,\theta) = \left\{2\Phi\left[(x/\theta)^{\lambda}\right]-1\right\} = erf\left[(x/\theta)^{\lambda}/\sqrt{2}\right], \tag{1}$$

where

$$\Phi\,(x) = \frac{1}{2}\left[1+erf\left(x/\sqrt{2}\right)\right],$$

and

$$erf\,(x) = \frac{2}{\sqrt{\pi}}\int_0^x \exp(-t^2)dt.$$

Its nth moment is given by (Cooray and Ananda, 2008) as

$$E(X^n) = \Gamma\,(n+\lambda/2\lambda)\,\theta^n\sqrt{\frac{2^{\frac{n}{\lambda}}}{\pi}},$$

where $\Gamma\,(.)$ is the gamma function. The Half-Normal (HN) distribution is a sub-model when $\lambda = 1$.

The goal of this paper is to propose the first generalization of the generalized half-normal distribution using the BurrX-G ("BrX-G" for short) family of distributions. For an arbitrary baseline cdf $G(x)$, Yousof et al. [2] proposed the probability density function (pdf) $f(x)$ and the cdf $F(x)$ of the BrX-G family of distributions with an additional shape parameter $\delta > 0$ defined (for $x \geq 0$) by

$$F(x;\delta,\xi) = \left(1-\exp\left\{-\left[\frac{G(x;\xi)}{\overline{G}(x;\xi)}\right]^2\right\}\right)^{\delta}. \tag{2}$$

The BrX-G density function is

$$
\begin{aligned}
f(x;\delta,\xi) =& \frac{2\delta g(x;\xi)G(x;\xi)}{\overline{G}(x;\xi)^3} \\
& \times \exp\left\{-\left[\frac{G(x;\xi)}{\overline{G}(x;\xi)}\right]^2\right\}\left(1-\exp\left\{-\left[\frac{G(x;\xi)}{\overline{G}(x;\xi)}\right]^2\right\}\right)^{\delta-1},
\end{aligned}
$$

where $\delta > 0$ is the shape parameter and $\xi = \xi_k = (\ \xi_1, \xi_2, ...)$ is a parameter vector. Based on the BrX-G family, we construct a new GHN distribution and provide a comprehensive description of some of its mathematical properties. We prove empirically that the BrXGHN model provides better fits than other competitive models, each one having the same number of parameters, by means of two applications to real data. We hope that the new distribution will attract wider applications in reliability, engineering and other areas of research. Inserting (1) into (2) we get

$$
F(x;\delta,\lambda,\theta) = \left[1-\exp\left(-\left\{\frac{2\Phi\left[(x/\theta)^\lambda\right]-1}{2-2\Phi\left[(x/\theta)^\lambda\right]}\right\}^2\right)\right]^\delta, x \geq 0. \quad (3)
$$

The corresponding pdf can be expressed as

$$
\begin{aligned}
f(x;\delta,\lambda,\theta) =& 2\delta\lambda\theta^{-\lambda}\sqrt{\frac{2}{\pi}}x^{\lambda-1}\exp\left[-\frac{1}{2}\left(\frac{x}{\theta}\right)^{2\lambda}\right]\left\{2\Phi\left[(x/\theta)^\lambda\right]-1\right\} \\
& \times\left\{2-2\Phi\left[(x/\theta)^\lambda\right]\right\}^{-3}\exp\left(-\left\{\frac{2\Phi\left[(x/\theta)^\lambda\right]-1}{2-2\Phi\left[(x/\theta)^\lambda\right]}\right\}^2\right) \quad (4)\\
& \times\overbrace{\left[1-\exp\left(-\left\{\frac{2\Phi\left[(x/\theta)^\lambda\right]-1}{2-2\Phi\left[(x/\theta)^\lambda\right]}\right\}^2\right)\right]^{\delta-1}}^{A_i}, x > 0.
\end{aligned}
$$

The justification for the practicality of the BrXGHN lifetime model is based on the fatigue crack growth under variable stress or cyclic load. Also we are motivated to introduce the BrXGHN lifetime model because it exhibits increasing as well as bathtub hazard rates as illustrated in Figure 2. It is shown in Subsection 3.1 that the BrXGHN lifetime model can be viewed as a mixture of the two-parameter GHN distributions introduced by Cooray and Ananda (2008). It can be viewed as a suitable model for fitting the unimodal data. The BrXGHN lifetime model outperforms several of the well-known lifetime distributions with respect to two real data applications as illustrated in Section 6. The new log-location regression model based on the BrXGHN distribution provides better fits than the log-odd log-logistic generalized half-normal and log-generalized half-normal regression models for HIV data set.

Many extension of the GHN model can be cited as follows: the beta generalized half-Normal distribution with applications to myelogenous leukemia data by Pescim et al. [3], Kumaraswamy generalized half-normal distribution for censored data by Cordeiro et al. [4], a log-linear regression model based on the beta generalized half-Normal distribution by Pescim et al. [5], the beta generalized half normal geometric distribution by Ramires et al. [6]. Merovci et al. [7] defined and applied the exponentiated transmuted generalized half-normal for a data set of the life of fatigue fracture of Kevlar 373/epoxy that are subject to constant pressure at the 90% stress level until all had failed.

The rest of the paper is organized as follows. Section 2 deals with some useful characterization results of the proposed model. In Section 3, we derived some of its mathematical properties. In Section 4, the maximum likelihood method is used to estimate the model parameters by means of a Monte Carlo simulation study. A new log-location regression model as well as residual analysis are presented in Section 5. Section 6 is devoted to applications to real data sets to prove empirically the importance of new the model. Finally, some conclusions and future work are given in Section 7.

2 Characterizations

This section deals with certain characterizations of BrXGHN distribution. These characterizations are in terms of: (i) the truncated moment involving two functions; (ii) a simple relationship between two truncated moments; (iii) the hazard function and (iv) certain function of the random variable. One of the advantages of characterization (ii) is that the cdf is not required to have a closed form. We present our characterizations $(i) - (iv)$ in four subsections.

2.1 Characterizations based on truncated moment involving two functions

Our first characterization is based on the following Proposition.

Proposition 1. *Let* $X : \Omega \to \mathbb{R}$ *be a continuous random variable with cdf* F *. Let* $\psi(x)$ *and* $\varphi(x)$ *be two differentiable functions on* $\mathbb{R}$ *such that* $\int_{-\infty}^{\infty} \frac{\varphi'(t)}{[\varphi(t)-\psi(t)]} dt = \infty$. *Then*

$$E[\psi(X) \mid X \geq x] = \varphi(x) \ , \quad x \in \mathbb{R},$$

implies

$$F(x) = 1 - \exp\left\{ -\int_{-\infty}^{x} \frac{\varphi'(t)}{[\varphi(t) - \psi(t)]} dt \right\}, \quad x \in \mathbb{R}.$$

Proof. If $E[\psi(X) \mid X \geq x] = \varphi(x) \ , \ x \in \mathbb{R}$ holds, then

$$\int_{x}^{\infty} \psi(u) f(u) du = (1 - F(x)) \varphi(x) .$$

Differentiating both sides of the above equation and rearranging the terms, we arrive at

$$\frac{f(x)}{1 - F(x)} = \frac{\varphi'(x)}{\varphi(x) - \psi(x)}, \quad x \in \mathbb{R}.$$

Integrating the last equation with respect to t from $-\infty$ to x , we have

$$-\ln[1 - F(x)] = \int_{-\infty}^{x} \frac{\varphi'(t)}{[\varphi(t) - \psi(t)]} dt,$$

from which we obtain

$$F(x) = 1 - \exp\left\{ -\int_{-\infty}^{x} \frac{\varphi'(t)}{[\varphi(t) - \psi(t)]} dt \right\}.$$

Remark 1. *For* $\delta = 1$, $\psi(x) = 2\varphi(x)$, $\varphi(x) = \exp\left\{ -\left(\left\{ \frac{2\Phi\left[(x/\theta)^{\lambda}\right]-1}{2-2\Phi\left[(x/\theta)^{\lambda}\right]} \right\}^2 \right) \right\}$, $x > 0$ *and the fact that* $\lim_{x \to 0^+} \varphi(x) = 1$, *we have*

$$F(x) = 1 - \exp\left(-\left\{ \frac{2\Phi\left[(x/\theta)^{\lambda}\right] - 1}{2 - 2\Phi\left[(x/\theta)^{\lambda}\right]} \right\}^2 \right), \quad x \geq 0,$$

which is cdf (3) *for* $\delta = 1$.

2.2 Characterizations based on a simple relationship between two truncated moments

In this subsection we present characterizations of BrXGHN distribution in terms of the ratio of two truncated moments. This characterization result employs a theorem due to Glanzel [8], see Theorem 1 of Appendix A. Note that the result holds also when the interval H is not closed. Moreover, as mentioned above, it could also be applied when the cdf F does not have a closed form. As shown in Glanzel [9], this characterization is stable in the sense of weak convergence.

Proposition 2. *Let $X : \Omega \to (0, \infty)$ be a continuous random variable and let*

$$q_1(x) = \left[1 - \exp\left(-\left\{\frac{2\Phi\left[(x/\theta)^{\lambda}\right] - 1}{2 - 2\Phi\left[(x/\theta)^{\lambda}\right]}\right\}^2\right)\right]^{1-\delta}$$

and

$$q_2(x) = \exp\left(-\left\{\frac{2\Phi\left[(x/\theta)^{\lambda}\right] - 1}{2 - 2\Phi\left[(x/\theta)^{\lambda}\right]}\right\}^2\right)$$

for $x > 0$. The random variable X has pdf (4) if and only if the function η defined in Theorem 1 has the form

$$\eta(x) = \frac{1}{2}\exp\left(-\left\{\frac{2\Phi\left[(x/\theta)^{\lambda}\right] - 1}{2 - 2\Phi\left[(x/\theta)^{\lambda}\right]}\right\}^2\right), \quad x > 0.$$

Proof. Let X be a random variable with pdf (4), then

$$(1 - F(x))\, E\left[q_1(X) \mid X \geq x\right] = \exp\left(-\left\{\frac{2\Phi\left[(x/\theta)^{\lambda}\right] - 1}{2 - 2\Phi\left[(x/\theta)^{\lambda}\right]}\right\}^2\right), \quad x > 0,$$

and

$$(1-F(x))\,E\left[q_2(X)\mid X\geq x\right]=$$
$$\frac{1}{2}\exp\left(-2\left\{\frac{2\Phi\left[(x/\theta)^{\lambda}\right]-1}{2-2\Phi\left[(x/\theta)^{\lambda}\right]}\right\}^2\right)\quad x>0,$$

and finally

$$\eta(x)\,q_1(x)-q_2(x)=$$
$$-\frac{q_1(x)}{2}\exp\left(-\left\{\frac{2\Phi\left[(x/\theta)^{\lambda}\right]-1}{2-2\Phi\left[(x/\theta)^{\lambda}\right]}\right\}^2\right)<0,\quad \mathit{for}\quad x>0.$$

Conversely, if η is given as above, then

$$s'(x)=\frac{\eta'(x)q_1(x)}{\eta(x)q_1(x)-q_2(x)}=2\lambda\theta^{-\lambda}\sqrt{\frac{2}{\pi}}x^{\lambda-1}$$
$$\times\frac{\left\{2\Phi\left[(x/\theta)^{\lambda}\right]-1\right\}}{\left\{2-2\Phi\left[(x/\theta)^{\lambda}\right]\right\}^3}\exp\left[-\frac{1}{2}(x/\theta)^{2\lambda}\right]\quad x>0.$$

Now, in view of Theorem 1, X has density (4).

Corollary 1. *Let $X:\Omega\to(0,\infty)$ be a continuous random variable and let $q_1(x)$ be as in Proposition 2 The pdf of X is (4) if and only if there exist functions q_2 and η defined in Theorem 1 satisfying the differential equation*

$$\frac{\eta'(x)\,q_1(x)}{\eta(x)\,q_1(x)-q_2(x)}=$$
$$2\lambda\theta^{-\lambda}\sqrt{\frac{2}{\pi}}x^{\lambda-1}\frac{\left\{2\Phi\left[(x/\theta)^{\lambda}\right]-1\right\}}{\left\{2-2\Phi\left[(x/\theta)^{\lambda}\right]\right\}^3}\exp\left[-\frac{1}{2}(x/\theta)^{2\lambda}\right]\quad x>0.$$

The general solution of the differential equation in Corollary 2.2 is

$$\eta(x) = \exp\left(\left\{\frac{2\Phi\left[(x/\theta)^{\lambda}\right]-1}{2-2\Phi\left[(x/\theta)^{\lambda}\right]}\right\}^{2}\right) \times \left[\begin{array}{c} -\int 2\lambda\theta^{-\lambda}\sqrt{\frac{2}{\pi}}x^{\lambda-1}\frac{\left\{2\Phi\left[(x/\theta)^{\lambda}\right]-1\right\}}{\left\{2-2\Phi\left[(x/\theta)^{\lambda}\right]\right\}^{3}}\exp\left[-\frac{1}{2}(x/\theta)^{2\lambda}\right]\times \\ \exp\left(-\left\{\frac{2\Phi\left[(x/\theta)^{\lambda}\right]-1}{2-2\Phi\left[(x/\theta)^{\lambda}\right]}\right\}^{2}\right)(q_1(x))^{-1}q_2(x)+D \end{array}\right],$$

where D is a constant. Note that a set of functions satisfying the above differential equation is given in Proposition 2 with $D = 0$. However, it should be also noted that there are other triplets (q_1, q_2, η) satisfying the conditions of Theorem 1.

2.3 Characterization based on hazard function

It is known that the hazard function, h_F, of a twice differentiable distribution function, F, satisfies the first order differential equation

$$\frac{f'(x)}{f(x)} = \frac{h'_F(x)}{h_F(x)} - h_F(x).$$

For many univariate continuous distributions, this is the only characterization available in terms of the hazard function. The following characterization establishes a non-trivial characterization of BrXGHN distribution for $\delta = 1$.

Proposition 3. *Let $X : \Omega \to (0,\infty)$ be a continuous random variable. The pdf of X is (4), for $\delta = 1$, if and only if its hazard function $h_F(x)$ satisfies the differential equation*

$$h'_F(x) + \lambda\theta^{-\lambda}x^{\lambda-1}h_F(x) = 2\lambda\theta^{-\lambda}\sqrt{\frac{2}{\pi}}\exp\left[-\frac{1}{2}(x/\theta)^{2\lambda}\right] \times\frac{d}{dx}\left\{\frac{x^{\lambda-1}\left\{2\Phi\left[(x/\theta)^{\lambda}\right]-1\right\}}{\left\{2-2\Phi\left[(x/\theta)^{\lambda}\right]\right\}^{3}}\right\}, \quad x > 0.$$

Proof. If X has pdf (4), then clearly the above differential equation holds. Now, if this differential equation holds, then

$$\frac{d}{dx}\left\{h_F(x)\exp\left[\frac{1}{2}(x/\theta)^{2\lambda}\right]\right\}=$$
$$2\lambda\theta^{-\lambda}\sqrt{\frac{2}{\pi}}\frac{d}{dx}\left\{\frac{x^{\lambda-1}\left\{2\Phi\left[(x/\theta)^{\lambda}\right]-1\right\}}{\left\{2-2\Phi\left[(x/\theta)^{\lambda}\right]\right\}^{3}}\right\},\quad x>0,$$

from which, we obtain

$$h_F(x)=\frac{2\lambda\theta^{-\lambda}\sqrt{\frac{2}{\pi}}x^{\lambda-1}\exp\left[-\frac{1}{2}(x/\theta)^{2\lambda}\right]\left\{2\Phi\left[(x/\theta)^{\lambda}\right]-1\right\}}{\left\{2-2\Phi\left[(x/\theta)^{\lambda}\right]\right\}^{3}},\quad x>0,$$

which is the hazard function of BrXGHN distribution for $\delta=1$.

2.4 Characterizations Based on Conditional Expectation

The following proposition has already appeared in [10], so we will just state it here which can be used to characterize the BrXGHN distribution for $\delta=1$.

Proposition 4. *Let* $X:\Omega\rightarrow(a,b)$ *be a continuous random variable with cdf* F *. Let* $\psi(x)$ *be a differentiable function on* (a,b) *with* $\lim_{x\rightarrow a^{+}}\psi(x)=1$. *Then for* $\gamma\neq 1$,

$$E\left[\psi(X)\mid X\geq x\right]=\gamma\psi(x),\quad x\in(a,b),$$

if and only if

$$\psi(x)=(1-F(x))^{\frac{1}{\gamma}-1},\quad x\in(a,b).$$

Remark 2. *For* $\psi(x)=\exp\left(-\left\{\frac{2\Phi\left[(x/\theta)^{\lambda}\right]-1}{2-2\Phi\left[(x/\theta)^{\lambda}\right]}\right\}^{2}\right)$, $\delta=1$, $\gamma=\frac{1}{2}$ *and* $(a,b)=(0,\infty)$, *Proposition 4 provides a characterization of BrXGHN distribution. Of course there are other suitable functions than the one we mentioned above, which is chosen for simplicity.*

3 Mathematical and statistical properties

In this section we will provide some mathematical and statistical properties of the BrXGHN distribution.

3.1 Linear representation

In this sub-section, we provide a very useful linear representation of the BrXGHN model. If $|z| < 1$ and $b > 0$ is a real non-integer, the following power series holds

$$(1-z)^{b-1} = \sum_{i=0}^{\infty} \frac{(-1)^i \, \Gamma(b)}{i! \, \Gamma(b-i)} \, z^i. \tag{5}$$

Applying (5) to the term A_i of (4), Equation (5) reduces to

$$\begin{aligned} f(x) = {} & 2\delta\lambda\theta^{-\lambda}\sqrt{\frac{2}{\pi}}x^{\lambda-1}\frac{\exp\left[-\frac{1}{2}(x/\theta)^{2\lambda}\right]\left\{2\Phi\left[(x/\theta)^{\lambda}\right]-1\right\}}{\left\{2-2\Phi\left[(x/\theta)^{\lambda}\right]\right\}^{3}} \\ & \times \sum_{i=0}^{\infty}\frac{(-1)^i\,\Gamma(\delta)}{i!\,\Gamma(\delta-i)} \underbrace{\exp\left(-(i+1)\left\{\frac{2\Phi\left[(x/\theta)^{\lambda}\right]-1}{2-2\Phi\left[(x/\theta)^{\lambda}\right]}\right\}^{2}\right)}_{B_i}. \end{aligned} \tag{6}$$

Applying the power series to the term B_i, Equation (6) becomes

$$\begin{aligned} f(x) = {} & 2\theta\sqrt{\frac{2}{\pi}}\left(\frac{\lambda}{x}\right)(x/\theta)^{\lambda}\exp\left[-\frac{1}{2}(x/\theta)^{2\lambda}\right] \\ & \times \sum_{i,j=0}^{\infty}\frac{(-1)^{i+j}\,(i+1)^j\,\Gamma(\delta)}{i!\,j!\Gamma(\delta-i)} \underbrace{\frac{\left\{2\Phi\left[(x/\theta)^{\lambda}\right]-1\right\}^{2j+1}}{\left\{2-2\Phi\left[(x/\theta)^{\lambda}\right]\right\}^{2j+3}}}_{C_i}. \end{aligned} \tag{7}$$

Consider the series expansion

$$(1-z)^{-b} = \sum_{k=0}^{\infty}\frac{\Gamma(b+k)}{k!\Gamma(b)}z^k, \ \ |z|<1, \ b>0. \tag{8}$$

Applying the expansion in (8) to (7) for the term C_i, Equation (7) becomes

$$f(x) = 2\delta \sum_{i,j,k=0}^{\infty} \frac{(-1)^{i+j}\ (i+1)^j\, \Gamma(\delta)\, \Gamma(2j+k+3)\, [2j+k+2]}{i!\, j!k!\Gamma(\delta-i)\, \Gamma(2j+3)\, [2j+k+2]}$$
$$\times \underbrace{\sqrt{\frac{2}{\pi}}\, (\lambda/x)\, (x/\theta)^{\lambda} \exp\left[-\frac{1}{2}\,(x/\theta)^{2\lambda}\right]}_{g(x;\lambda,\theta)} \underbrace{\left\{2\Phi\left[(x/\theta)^{\lambda}\right]-1\right\}^{2j+k+1}}_{G(x;\lambda,\theta)}.$$

This can be written as

$$f(x) = \sum_{j,k=0}^{\infty} \Omega_{j,k}\, \pi_{2j+k+2}(x;\lambda,\theta), \tag{9}$$

where

$$\Omega_{j,k} = 2\delta\, \frac{(-1)^j\, \Gamma(\delta)\, \Gamma(2j+k+3)}{j!k!\Gamma(2j+3)\,(2j+k+2)} \sum_{i=0}^{\infty} \frac{(-1)^i\ (i+1)^j}{i!\, \Gamma(\delta-i)}$$

and $\pi_{2j+k+2}(x;\lambda,\theta) = (2j+k+2)\, g(x;\lambda,\theta)\, G(x;\lambda,\theta)^{2j+k+1}$ is the pdf of the exponentiated-GHN (Exp-GHN) distribution with the power parameter $2j+k+2$. Equation (9) reveals that the density of X can be expressed as a linear mixture of exp-G densities. So, several mathematical properties of the new family can be obtained from those of the Exp-GHN distribution. Similarly, the cdf of the BrXGHN model can be expressed as a mixture of Exp-GHN cdfs given by

$$F(x) = \sum_{j,k=0}^{\infty} \Omega_{j,k}\ \Pi_{2j+k+2}(x;\lambda,\theta),$$

where $\Pi_{2j+k+2}(x;\lambda,\theta) = G(x;\lambda,\theta)^{2j+k+1}$ is the cdf of the Exp-GHN distribution with the power parameter $2j+k+2$.

3.2 Moments and generating function

By setting $u = \left(\frac{x}{\theta}\right)^{\lambda}$ and considering the error function as the cdf of the GHN distribution, the nth moment of X can be obtained from equation (9) as

$$\mu_n' = E(X^n) = \theta^n \sqrt{\frac{2}{\pi}} \sum_{j,k=0}^{\infty} \Omega_{j,k} I(n/\lambda, k),$$

where

$$I\left(n/\lambda,k\right)=\int_{0}^{\infty}u^{\frac{n}{\lambda}}\exp\left(-\frac{1}{2}u^{2}\right)\left[erf\left(\frac{u}{\sqrt{2}}\right)\right]^{k}du.$$

Inserting the power series for the error function

$$erf\left(x\right)=\frac{2}{\sqrt{\pi}}\sum_{m=0}^{\infty}\left(-1\right)^{m}\left(m!\right)^{-1}\left(2m+1\right)^{-1}x^{2m+1}$$

in the last equation and computing the integral, we have (for any real $k+n/\lambda$) we get

$$E\left(X^{n}\right)=\theta^{n}\sqrt{2/\pi}\sum_{j,k=0}^{\infty}\Omega_{j,k}I\left(n/\lambda,k\right),$$

where

$$I\left(n/\lambda,k\right)=2^{-\frac{1}{2}+k+\frac{n}{2\lambda}}\ \pi^{-\frac{1}{2}k}\sum_{m_1,\ldots,m_k=0}^{\infty}\frac{\left(-1\right)^{m_1+\ldots+m_k}\Gamma\left(m_1+\ldots+m_k+\frac{1}{2}\left[1+k+n/\lambda\right]\right)}{m_1!\left(m_1+\frac{1}{2}\right)m_2!\left(m_2+\frac{1}{2}\right)\ldots m_k!\left(m_k+\frac{1}{2}\right)}.$$

Moreover, for the very special case when $k+n/\lambda$ is even, the integral $I\left(n/\lambda,k\right)$ can be expressed in terms of the Lauricella function of type A (Exton [11]; Aarts [12]) defined by

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

where $(a)_k=a\left(a+1\right)\ldots\left(a+k-1\right)$ is the ascending factorial (with the convention that $(a)_0=1$). Numerical technics for the direct computation of the Lauricella function of type A are available, see Exton [11] and Mathematica (Trott [13]). Hence, $E\left(X^{n}\right)$ can be expressed in terms of the Lauricella functions of type A

$$E\left(X^{n}\right)=\theta^{n}\sqrt{\frac{2}{\pi}}\sum_{j,k=0}^{\infty}a_{j,k}F_A^{(k)}\left(\frac{1}{2}\left[2j+k+2+n/\lambda\right];\frac{1}{2},\ldots,\frac{1}{2};\frac{3}{2},\ldots,\frac{3}{2};-1,\ldots,-1\right),$$

where

$$a_{j,k}=2^{-\frac{1}{2}+k+\frac{n}{2\lambda}}\pi^{-\frac{1}{2}k}\Omega_{j,k}\,\Gamma\left(\frac{1}{2}\left[2j+k+2+n/\lambda\right]\right).$$

The central moments (μ_n) and cumulants (κ_n) of X are determined using $E\left(X^n\right)$ as

$$\mu_n=\sum_{k=0}^{n}\binom{n}{k}(-1)^k\;\mu_1^{\prime k}\;\mu_{n-k}^{\prime}$$

and

$$\kappa_n=\mu_n^{\prime}-\sum_{k=1}^{n-1}\binom{n-1}{k-1}\;\kappa_k\;\mu_{n-k}^{\prime},$$

respectively, where $\kappa_1=\mu_1^{\prime}$. The skewness $\beta_1=\kappa_3/\kappa_2^{3/2}$ and kurtosis $\beta_2=\kappa_4/\kappa_2^2$ are obtained from the third and fourth standardized cumulants.The moment generating function (mgf) of $of\;\;X,\quad say\;M_X\left(t\right)=E\left(e^{tX}\right)$, is given by $M_X\left(t\right)=\sum\limits_{n=0}^{\infty}t^n\left(n!\right)^{-1}E\left(X^n\right)$. The characteristic function (cf) of X, $\phi\left(t\right)$ $=E\left(e^{it\,X}\right)$, and the cumulant generating function (cgf) of X, $K\left(t\right)=\log\phi\left(t\right)$ can be obtained from the well known relationships, where $i=\sqrt{-1}$.

3.3 Probability weighted moments

The probability weighted moment (PWM)s are expectations of certain functions of a random variable and they can be defined for any random variable whose ordinary moments exist. The PWM method can generally be used for estimating parameters of a distribution whose inverse form cannot be expressed explicitly. The (s,r)th PWM of X following the BrXGHN model, say $\rho_{s,r}$, is formally defined by

$$\rho_{s,r}=E\left\{X^s\,F(X)^r\right\}=\int_{-\infty}^{\infty}x^s\,F(x)^r\,f\left(x\right)\,dx.$$

Using equations (3) and (4) we can write

$$f\left(x\right)\;F(x)^r=\sum_{j,k=0}^{\infty}a_{j,k}\pi_{2j+k+2}\left(x\right),$$

where

$$a_{j,k}=\frac{2\delta\left(-1\right)^j\Gamma\left(2j+k+3\right)}{j!k!\Gamma\left(2j+3\right)\left(2j+k+2\right)}\sum_{i=0}^{\infty}\left(-1\right)^i\left(i+1\right)^j\binom{\delta\left(r+1\right)-1}{i}.$$

Then, the (s,r)th PWM of X can be expressed as

$$\rho_{s,r} = \theta^s \sqrt{\frac{2}{\pi}} \sum_{j,k=0}^{\infty} b_{j,k} F_A^{(k)} \left(\frac{1}{2} \left[2j+k+2+s/\lambda \right]; \frac{1}{2}, ..., \frac{1}{2}; \frac{3}{2}, ..., \frac{3}{2}; -1, ..., -1 \right).$$

where

$$b_{j,k} = 2^{-\frac{1}{2}+k+\frac{n}{2\lambda}} \pi^{-\frac{1}{2}k} a_{j,k} \, \Gamma \left(\frac{1}{2} \left[2j+k+2+s/\lambda \right] \right).$$

3.4 Stress-strength model

Let X_1 and X_2 be two independent random variables with BrX-GHN(δ_1, ξ) and BrX-GHN(δ_2, ξ) distributions, respectively. Then, the reliability is defined by

$$\mathbf{R} = \int_0^{\infty} f_1 \left(x; \delta_1, \xi \right) F_2 \left(x; \delta_2, \xi \right) dx.$$

We can write

$$\mathbf{R} = \sum_{j,k,w,m=0}^{\infty} s_{j,k,w,m} \int_0^{\infty} \pi_{2j+2w+k+m+4} \left(x \right) dx,$$

where

$$\begin{aligned} s_{j,k,w,m} &= 4\delta_1\delta_2 \sum_{i,h=0}^{\infty} \frac{(-1)^{i+h} (i+1)^j \, (h+1)^w \binom{\delta_1-1}{i} \binom{\delta_2-1}{h}}{(2w+m+2)(2j+k+2w+m+4)} \\ &\times \sum_{j,k,w,m=0}^{\infty} \frac{(-1)^{j+w} \, \Gamma(2j+k+3) \, \Gamma(2w+m+3)}{j!k!\, w!m! \Gamma(\delta_2-h) \, \Gamma(2j+3) \, \Gamma(2w+3)}. \end{aligned}$$

Thus, the reliability, $\mathbf{R}$, can be expressed as

$$\mathbf{R} = \sum_{j,k,w,m=0}^{\infty} s_{j,k,w,m}.$$

3.5 Order statistics

Order statistics make their appearance in many areas of statistical theory and practice. Let $X_1, \ldots, X_n$ be a random sample from the BrXGHN distributions and let $X_{1:n}, \ldots, X_{n:n}$ be the corresponding order statistics. The pdf of ith order statistic, $X_{i:n}$, can be written as

$$f_{i:n}(x) = [\mathrm{B}(i, n-i+1)]^{-1} \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} f(x)\ F^{j+i-1}(x), \qquad (10)$$

where $B(\cdot,\cdot)$ is the beta function. Using (3), (4) and (10) we have

$$f(x)\ F(x)^{j+i-1} = \sum_{w,k=0}^{\infty} t_{w,k} \pi_{2w+k+2}(x),$$

where

$$t_{w,k} = \frac{2\delta(-1)^w \Gamma(2w+k+3)}{w!k!\Gamma(2w+3)(2w+k+2)} \sum_{m=0}^{\infty} (-1)^m (m+1)^w \binom{\delta(j+i)-1}{m}.$$

The pdf of $X_{i:n}$ can be expressed as

$$f_{i:n}(x) = \sum_{w,k=0}^{\infty} \sum_{j=0}^{n-i} (-1)^j\ [\mathrm{B}(i, n-i+1)]^{-1} \binom{n-i}{j} t_{w,k} \pi_{2w+k+2}(x).$$

Then, the density function of the BrX-GHN order statistics is a mixture of exp-G densities. Based on the last equation, we note that the properties of $X_{i:n}$ follow from those properties of Y_{2w+k+2}. For example, the moments of $X_{i:n}$ can be expressed as

$$E(X_{i:n}^q) = \sum_{w,k=0}^{\infty} \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} [\mathrm{B}(i, n-i+1)]^{-1} t_{w,k} E\left(Y_{2w+k+2}^q\right).$$

$$E(X_{i:n}^q) = \theta^q \sqrt{\frac{2}{\pi}} \sum_{w,k=0}^{\infty} c_{j,k} F_A^{(k)} \left(\frac{1}{2}[2w+k+2+q/\lambda]; \frac{1}{2}, ..., \frac{1}{2}; \frac{3}{2}, ..., \frac{3}{2}; -1, ..., -1\right),$$

where

$$c_{j,k} = 2^{-\frac{1}{2}+k+\frac{n}{2\lambda}}\ \pi^{-\frac{1}{2}k} \times \sum_{j=0}^{n-i} (-1)^j \binom{n-i}{j} [\mathrm{B}(i, n-i+1)]^{-1} t_{w,k} \Gamma\left(\frac{1}{2}[2w+k+2+q/\lambda]\right).$$

4 Estimation

4.1 Maximum likelihood estimation

Several approaches for parameter estimation were proposed in the literature but the maximum likelihood method is the most commonly employed. So, we consider the estimation of the unknown parameters of this family from complete samples only by maximum likelihood. Let $x_1, \ldots, x_n$ be a random sample from the BrX-GHN model with parameters δ, λ and θ. Let $\Theta = (\delta, \lambda, \theta)^\intercal$ be the $p \times 3$ parameter vector. The log-likelihood function is

$$\begin{aligned}\ell &= \ell(\Theta) = n\log 2 + n\log\delta + n\log\sqrt{\frac{2}{\pi}} + n\log\lambda + n\lambda\log\theta + \sum_{i=1}^{n}\log x_i^{\lambda-1} \\ &-\frac{1}{2}\sum_{i=1}^{n}(x_i/\theta)^{2\lambda} + \sum_{i=1}^{n}\log(\tau_i-1) - 3\sum_{i=1}^{n}\log(2-\tau_i) \\ &-\sum_{i=1}^{n}s_i^2 + (\delta-1)\sum_{i=1}^{n}\log\left[1-\exp\left(-s_i^2\right)\right],\end{aligned}$$

where $s_i = \frac{\tau_i - 1}{2-\tau_i}$ and $\tau_i = 2\Phi\left[(x_i/\theta)^\lambda\right]$. The components of the score vector, $\mathbf{U}(\Theta) = \frac{\partial\ell}{\partial\Theta} = \left(\frac{\partial\ell}{\partial\delta}, \frac{\partial\ell}{\partial\lambda}, \frac{\partial\ell}{\partial\theta}\right)^\intercal$, are

$$U_\delta = \frac{n}{\delta} + \sum_{i=1}^{n}\log\left[1-\exp\left(-s_i^2\right)\right],$$

where

$$\begin{aligned}U_\lambda &= \frac{n}{\lambda} + +n\log\theta + \sum_{i=1}^{n}\log x_i - \sum_{i=1}^{n}(x_i/\theta)^{2\lambda}\log(x_i/\theta) \\ &+\sum_{i=1}^{n}\frac{\zeta_{\tau_i}^{(\lambda)}}{\tau_i-1} + 3\sum_{i=1}^{n}\frac{\zeta_{\tau_i}^{(\lambda)}}{2-\tau_i} - 2\sum_{i=1}^{n}s_i\frac{\zeta_{\tau_i}^{(\lambda)}}{(2-\tau_i)^2} \\ &+(\delta-1)\sum_{i=1}^{n}\frac{2s_i\zeta_{\tau_i}^{(\lambda)}(2-\tau_i)^{-2}\exp(-s_i^2)}{1-\exp(-s_i^2)}\end{aligned}$$

and

$$U_\theta = \frac{n\lambda}{\theta} - \lambda \sum_{i=1}^{n} \theta^{-2\lambda-1} + \sum_{i=1}^{n} \frac{\zeta_q^{(\theta)}}{\tau_i - 1} + 3\sum_{i=1}^{n} \frac{\zeta_q^{(\theta)}}{2-\tau_i}$$
$$-2\sum_{i=1}^{n} s_i \frac{\zeta_{\tau_i}^{(\theta)}}{(2-\tau_i)^2} + (\delta - 1)\sum_{i=1}^{n} \frac{2s_i\zeta_{\tau_i}^{(\theta)}(2-\tau_i)^{-2}\exp(-s_i^2)}{1-\exp(-s_i^2)},$$

where

$$\zeta_{\tau_i}^{(\lambda)} = \frac{2}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(x_i/\theta)^{2\lambda}\right](x_i/\theta)^{\lambda}\log(x_i/\theta)$$

and

$$\zeta_{\tau_i}^{(\theta)} = \frac{2}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(x_i/\theta)^{2\lambda}\right](x_i/\theta)^{\lambda}(\lambda/\theta).$$

Setting the nonlinear system of equations $U_\theta = 0$ and $U_\xi = \mathbf{0}$ and solving them simultaneously yields the MLE $\widehat{\Theta} = (\widehat{\theta}, \widehat{\xi}^{\intercal})^{\intercal}$. To solve these equations, it is usually more convenient to use nonlinear optimization methods such as the quasi-Newton algorithm to numerically maximize ℓ.

4.2 Simulation study

In this section, we conduct a simulation study to evaluate the performance of MLEs of BrX-GHN model. We generate 10,000 samples of size, n =50, 250 and 500 from BrXGHN model using the inverse transform method. The evaluation of estimates was based on the averages of estimates (AEs) and mean squared errors (MSEs). The empirical study was conducted with software R.

The empirical results are given in Table 1. The values in Table 1 indicate that the estimates are quite stable and, more importantly, are close to the true values for these sample sizes. The simulation study shows that the maximum likelihood method is appropriate for estimating the BrXGHN parameters. In fact, the MSEs tend to be closer to zero when n increases. This fact supports that the asymptotic normal distribution provides an adequate approximation to the finite sample distribution of the MLEs.

Tab. 1: Estimated AEs and MLEs for several parameter values of BrXGHN distribution.

Parameters		AE			MSE		
	n	δ	λ	θ	δ	λ	θ
(2,2,0.5)	50	3.01134	2.27575	0.49430	10.84800	2.33700	0.00017
	250	2.07400	2.05300	0.49900	0.37500	0.11200	0.00002
	500	2.02300	2.03400	0.49970	0.15400	0.05600	0.00001
(0.5,2,2)	50	0.55419	2.89561	1.97113	0.18602	5.36315	0.00526
	250	0.50072	2.12757	1.99380	0.01453	0.23766	0.00089
	500	0.50021	2.04821	1.99676	0.00268	0.04086	0.00019
(2,4,5)	50	3.01081	4.90053	4.97673	16.71033	32.58943	0.00439
	250	2.10414	4.06892	4.99638	0.36515	0.44431	0.00040
	500	2.03973	4.06787	4.99713	0.19537	0.24999	0.00022

5 Log-BrXGHN regression model

Consider the BrXGHN distribution with three parameters given in (4). Henceforth, X denotes a random variable following the BrXGHN distribution (4) and let $Y = \log(X)$. The density function of Y (for $y \in \Re$) obtained by replacing $\lambda = \frac{\sqrt{2}}{2\sigma}$ and $\theta = \exp(\mu)$ can be expressed as

$$f(y) = \frac{\frac{2\delta}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2}\exp\left[\left(\frac{y-\mu}{\sigma}\right)\sqrt{2}\right]+\left(\frac{y-\mu}{\sigma}\right)\frac{\sqrt{2}}{2}\right\}\left[2\Phi\left[\exp\left[\left(\frac{y-\mu}{\sigma}\right)\frac{\sqrt{2}}{2}\right]\right]-1\right]}{\left(2-2\Phi\left[\exp\left[\left(\frac{y-\mu}{\sigma}\right)\frac{\sqrt{2}}{2}\right]\right]\right)^{3}}$$
$$\times\exp\left\{-\left[\frac{\left[2\Phi\left[\exp\left[\left(\frac{y-\mu}{\sigma}\right)\frac{\sqrt{2}}{2}\right]\right]-1\right]}{\left(2-2\Phi\left[\exp\left[\left(\frac{y-\mu}{\sigma}\right)\frac{\sqrt{2}}{2}\right]\right]\right)}\right]^{2}\right\}\left(1-\exp\left\{-\left[\frac{\left[2\Phi\left[\exp\left[\left(\frac{y-\mu}{\sigma}\right)\frac{\sqrt{2}}{2}\right]\right]-1\right]}{\left(2-2\Phi\left[\exp\left[\left(\frac{y-\mu}{\sigma}\right)\frac{\sqrt{2}}{2}\right]\right]\right)}\right]^{2}\right\}\right)^{\delta-1}, \quad (11)$$

where $\mu \in \Re$ is the location parameter, $\sigma > 0$ is the scale parameter and $\delta > 0$ is the shape parameter. We refer to equation (11) as the Log-BrXGHN (LBrXGHN) pdf, say $Y \sim \text{LBrXGHN}(\delta, \sigma, \mu)$. The plots in Figure 3 show shapes of density function (11) for selected parameter values. They reveal that this distribution is a good candidate to model left skewed and symmetric lifetime data sets. The survival function corresponding to (11) is given by

$$S(y) = 1 - \left(1-\exp\left\{-\left[\frac{\left[2\Phi\left[\exp\left[\left(\frac{y-\mu}{\sigma}\right)\frac{\sqrt{2}}{2}\right]\right]-1\right]}{\left(2-2\Phi\left[\exp\left[\left(\frac{y-\mu}{\sigma}\right)\frac{\sqrt{2}}{2}\right]\right]\right)}\right]^{2}\right\}\right)^{\delta}, \quad (12)$$

and the hrf is simply $h(y) = f(y)/S(y)$. The standardized random variable $Z = (Y - \mu)/\sigma$ has density function

$$f(z) = \frac{\frac{2\delta}{\sqrt{2\pi}} \exp\left\{ \frac{1}{2} \exp\left[z\sqrt{2}\right] + (z)\frac{\sqrt{2}}{2}\right\} \left[2\Phi\left[\exp\left[z\frac{\sqrt{2}}{2}\right]\right] - 1\right]}{\left(2 - 2\Phi\left[\exp\left[z\frac{\sqrt{2}}{2}\right]\right]\right)^3}$$
$$\times \exp\left\{ -\left[\frac{\left[2\Phi\left[\exp\left[z\frac{\sqrt{2}}{2}\right]\right]-1\right]}{\left(2-2\Phi\left[\exp\left[z\frac{\sqrt{2}}{2}\right]\right]\right)}\right]^2 \right\} \left(1 - \exp\left\{ -\left[\frac{\left[2\Phi\left[\exp\left[z\frac{\sqrt{2}}{2}\right]\right]-1\right]}{\left(2-2\Phi\left[\exp\left[z\frac{\sqrt{2}}{2}\right]\right]\right)}\right]^2 \right\}\right)^{\delta - 1}. \tag{13}$$

5.1 Estimation

5.1.1 Maximum Likelihood Estimation

Based on the LBrXGHN density, we propose a linear location-scale regression model linking the response variable y_i and the explanatory variable vector $\mathbf{v}_i^\intercal = (v_{i1}, \ldots, v_{ip})$ given by

$$y_i = \mathbf{v}_i^\intercal \boldsymbol{\beta} + \sigma z_i, \;\; i = 1, \ldots, n, \tag{14}$$

where the random error z_i has density function (13), $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\intercal$, and $\sigma > 0$ and $\delta > 0$ are unknown parameters. The parameter $\mu_i = \mathbf{v}_i^\intercal \boldsymbol{\beta}$ is the location of y_i. The location parameter vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^\intercal$ is represented by a linear model $\boldsymbol{\mu} = \mathbf{V}\boldsymbol{\beta}$, where $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)^\intercal$ is a known model matrix.

Consider a sample $(y_1, \mathbf{v}_1), \ldots, (y_n, \mathbf{v}_n)$ of n independent observations, where each random response is defined by $y_i = \min\{\log(x_i), \log(c_i)\}$ where x_i and c_i are lifetime and censoring times, respectively. We assume non-informative censoring such that the observed lifetimes and censoring times are independent. Let F and C be the sets of individuals for which y_i is the log-lifetime or log-censoring, respectively. The log-likelihood function for the vector of parameters $\boldsymbol{\tau} = (\theta, \beta, \sigma, \boldsymbol{\beta}^\intercal)^\intercal$ from model (14) has the form $l(\boldsymbol{\tau}) = \sum\limits_{i \in F} l_i(\boldsymbol{\tau}) + \sum\limits_{i \in C} l_i^{(c)}(\boldsymbol{\tau})$, where $l_i(\boldsymbol{\tau}) = \log[f(y_i)]$, $l_i^{(c)}(\boldsymbol{\tau}) = \log[S(y_i)]$, $f(y_i)$ is the density (11) and $S(y_i)$ is the survival function (12) of Y_i. The total log-likelihood function for $\boldsymbol{\tau}$ is given by

$$
\begin{aligned}
\ell(\boldsymbol{\tau}) =& r\log\left(\frac{2\delta}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2}\sum_{i\in F}\exp\left(z_i\sqrt{2}\right) + \frac{\sqrt{2}}{2}\sum_{i\in F} z_i + \sum_{i\in F}\log\left[u_i - 1\right] \\
& - 3\sum_{i\in F}\log\left[2-u_i\right] \\
& - \sum_{i\in F}\left[\frac{u_i-1}{2-u_i}\right]^2 + (\delta-1)\sum_{i\in F}\log\left(1-\exp\left\{-\left[\frac{u_i-1}{2-u_i}\right]^2\right\}\right) \\
& + \sum_{i\in C}\log\left[1-\left(1-\exp\left\{-\left[\frac{u_i-1}{2-u_i}\right]^2\right\}\right)^{\delta}\right]
\end{aligned}
\tag{15}
$$

where $u_i = 2\Phi\left[\exp\left(z_i\sqrt{2}/2\right)\right]$, $z_i = (y_i - \mu_i)/\sigma$ and r is the number of uncensored observations (failures). The MLE $\widehat{\boldsymbol{\tau}}$ of the vector of unknown parameters can be evaluated by maximizing the log-likelihood function (15). The optim function of R software is used to estimate $\widehat{\boldsymbol{\tau}}$. Under the standard regularity conditions, the asymptotic distribution of $(\widehat{\boldsymbol{\tau}} - \boldsymbol{\tau})$ is multivariate normal $N_{p+2}(0, K(\boldsymbol{\tau})^{-1})$, where $K(\boldsymbol{\tau})$ is the expected information matrix. The asymptotic covariance matrix $K(\boldsymbol{\tau})^{-1}$ of $\widehat{\boldsymbol{\tau}}$ can be approximated by the inverse of the $(p+2)\times(p+2)$ observed information matrix $-\ddot{\mathbf{L}}(\boldsymbol{\tau})$, whose elements are evaluated numerically. The approximate multivariate normal distribution $N_{p+2}(0, -\ddot{\mathbf{L}}(\boldsymbol{\tau})^{-1})$ for $\widehat{\boldsymbol{\tau}}$ can be used, in the classical way, to construct approximate confidence intervals for the parameters in $\boldsymbol{\tau}$.

5.2 Sensitivity analysis

A first tool to perform sensitivity analysis, as stated before, is by means of global influence starting from case deletion. Case deletion is a popular method to investigate the influence of taking out the i_{th} case from the data on the parameters estimates. This method compares the $\hat{\boldsymbol{\tau}}$ with $\hat{\boldsymbol{\tau}}_{-i}$ where $\hat{\boldsymbol{\tau}}_{-i}$ is the estimated parameters when the i_{th} case is dropped from the data. If there is a big differences betweeb $\hat{\boldsymbol{\tau}}_{-i}$ and $\hat{\boldsymbol{\tau}}$, the dropped observation could be considered as influential observation.

Here, generalized cook distance is used to detect the possible influential observations. Generalized Cook distance (GD) is given by

$$
GD_i(\boldsymbol{\tau}) = \left(\hat{\boldsymbol{\tau}}_{-i} - \hat{\boldsymbol{\tau}}\right)^T\left[-\ddot{L}\left(\hat{\boldsymbol{\tau}}\right)\right]\left(\hat{\boldsymbol{\tau}}_{-i} - \hat{\boldsymbol{\tau}}\right), \tag{16}
$$

where $-\ddot{L}(\hat{\boldsymbol{\tau}})$ is the observed information matrix.

5.3 Residual analysis

Residual analysis has critical role in checking the adequacy of the fitted model. In order to analyse departures from error assumption, two types of residuals are considered: martingale and modified deviance residuals.

5.3.1 Martingale residual

The martingale residuals is defined in counting process and takes values between $+1$ and$-\infty$ (see, Fleming and Harrington[14] for details). The martingale residuals for LOLLBXII model is,

$$r_{M_i} = \begin{cases} 1+\log\left\{1-\left(1-\exp\left\{-\left[\frac{u_i-1}{2-u_i}\right]^2\right\}\right)^\delta\right\} & \text{if} i \in F, \\ \log\left\{1-\left(1-\exp\left\{-\left[\frac{u_i-1}{2-u_i}\right]^2\right\}\right)^\delta\right\} & \text{if} i \in C, \end{cases} \tag{17}$$

where $u_i = 2\Phi\left[\exp\left(z_i\sqrt{2}/2\right)\right]$ and $z_i = (y_i - \mu_i)/\sigma$.

5.3.2 Modified deviance residual

The main drawback of martingale residual is that when the fitted model is correct, it is not symmetrically distributed about zero. To overcome this problem, modified deviance residual was proposed by Therneau et al. [15]. The modified deviance residual is given by

$$r_{D_i} = \begin{cases} sign\left(r_{M_i}\right)\left\{-2\left[r_{M_i}+\log\left(1-r_{M_i}\right)\right]\right\}^{1/2}, & \text{if} i \in F \\ sign\left(r_{M_i}\right)\left\{-2r_{M_i}\right\}^{1/2}, & \text{if} i \in C, \end{cases} \tag{18}$$

where $\hat{r}_{M_i}$ is the martingale residual.

6 Applications

In this section, we provide two applications to real data sets to illustrate the flexibility of the BrxGHN distribution and BrxGHN regression model. The statistical software R is used for all numerical computations. The following goodness-of-fit measures are used to compare fitted models: Cramer von Mises (W*), Anderson Darling (A*), estimated $-\ell$ and Akaike Information Criteria (AIC). In general, the smaller the values of these statistics, the better the fit to the data.

We compare the BrxGHN distribution with another extension of GHN distribution introduced by Cordeiro et al. [10], named odd log-logistic generalized half-normal (OLLGHN) distribution. The cdf of OLLGHN distribution is given by

$$F_{OLLGHN}(x;\alpha,\lambda,\theta)=\frac{\left\{2\Phi\left[\left(\frac{x}{\theta}\right)^{\lambda}\right]-1\right\}^{\alpha}}{\left\{2\Phi\left[\left(\frac{x}{\theta}\right)^{\lambda}\right]-1\right\}^{\alpha}+\left\{2-2\Phi\left[\left(\frac{x}{\theta}\right)^{\lambda}\right]\right\}^{\alpha}}, \quad (19)$$

where $\alpha>0$ is additional shape parameter. Note that when $\alpha=1$, OLLGHN distribution reduces to GHN distribution.

6.1 Univariate data modeling

The first data set refers to a lifetime taken from Gross and Clark [16]. The data are: 1.1,1.4,1.3,1.7,1.9,1.8,1.6,2.2,1.7,2.7,4.1,1.8,1.5,1.2,1.4,3,1.7,2.3,1.6, Table 2 shows the estimated parameters and their standard errors, $-\ell$ and AIC values. Based on the figures in Table 2, BrxGHN distribution provides better fits than OLL-GHN distribution for used data set. Figure 4(a) displays the histogram with fitted pdfs and Figure 4(b) displays the fitted hrf and P-P plot of BrXGHN distribution. These figures reveal that BrXGHN model provides superior fits to used data set.

Tab. 2: MLEs and their SEs of the fitted models and goodness-of-fit statistics for second data set

Models	α	δ	λ	θ	$-\ell$	AIC	$A^\star$	$W^\star$
BrXGHN		16188.74	0.0771	0.214	15.545	37.091	0.256	0.047
		(33.442)	(0.022)	(0.214)				
OLL-GHN	34.380		0.099	92.226	16.574	39.148	0.400	0.071
	(11.502)		(0.027)	(101.786)				
GHN			1.955	2.302	22.452	48.905	1.391	0.242
			(0.487)	(0.137)				

6.2 HIV data set

The performance of LBrXGHN regression model is compared with log-odd log-logistic generalized half-normal (LOLLGHN) regression model, introduced by Pescim et al. [17], and log-generalized half-normal (LGHN) regression model. The survival functions of LOLLGHN and LGHN regression models are given by, respectively,

$$\begin{aligned} S_{LOLLGHN}(y) &= \frac{\left(2-\Phi\left[\exp\left[\left(\frac{y-\mu}{\sigma}\right)\frac{\sqrt{2}}{2}\right]\right]\right)^{\theta}}{\left[2\Phi\left[\exp\left[\left(\frac{y-\mu}{\sigma}\right)\frac{\sqrt{2}}{2}\right]\right]-1\right]^{\theta}+\left(2-\Phi\left[\exp\left[\left(\frac{y-\mu}{\sigma}\right)\frac{\sqrt{2}}{2}\right]\right]\right)^{\theta}}, \\ S_{LGHN}(y) &= 2-2\Phi\left[\exp\left[\left(\frac{y-\mu}{\sigma}\right)\frac{\sqrt{2}}{2}\right]\right]. \end{aligned} \tag{20}$$

The hypothetical dataset contains 100 observations on HIV+ subjects belonging to an Health Maintenance Organization(HMO). The HMO wants to evaluate the survival time of these subjects. In this hypothetical data set, subjects were enrolled from January 1, 1989 until December 31, 1991. Study follow up then ended on December 31, 1995. This data set are reported in Hosmer and Lemeshow [18] and also can be found in R package *Bolstad2*. We adopt the LBrXGHN regression model to analyze this dataset. The variables involved in the study are: y_i - observed survival time (in months); $cens_i$ - censoring indicator (0= alive at study end or lost to follow-up,1=death due to AIDS or AIDS related factors), $x_{i1}(1 = yes, 0 = no)$ represents the history of drug use and x_{i2} represents the ages of patients.

We consider the following regression model

$$y_i = \boldsymbol{\beta_0} + \boldsymbol{\beta_1} x_{i1} + \boldsymbol{\beta_2} x_{i2} + \sigma z_i,$$

where y_i has the LBrXGHN density, for $i = 1, \ldots, 100$.

6.2.1 Maximum Likelihood Estimation

The MLE method is used to estimate unknown parameters of LBrxGHN, LOLLGHN and LGHN regression models. Table 3 shows the MLEs of the model parameters fitted regression models, estimated log-likelihood values and AIC values. These results indicate that the LBrxGHN regression model has the lowest values of these statistics, and so LBrxGHN model provides better fitting than LOLLGHN and LGHN models for used data set. For the fitted regression models, note that β_0, β_1 and β_2 are marginally significant at the 1% level.

Tab. 3: MLEs of the parameters, their standard errors and p-values, the estimated $-\ell$ and AIC statistic.

Models	Parameters	δ	θ	σ	β_0	β_1	β_2	$-\ell$	AIC
LGHN	Estimates			0.757	6.347	-0.091	-1.126	130.590	269.180
	Std. Errors			0.067	0.487	0.013	0.177		
	p values				$<$0.001	$<$0.001	$<$0.001		
LOLLGHN	Estimates		2.448	1.691	6.686	-0.091	-0.965	128.228	266.455
	Std. Errors		1.72504	1.1841	0.7582	0.01427	0.2097		
	p values				$<$0.001	$<$0.001	$<$0.001		
LBrXGHN	Estimates	5.064		5.066	7.085	-0.089	-0.962	127.585	265.171
	Std. Errors	3.988		1.981	0.612	0.015	0.211		
	p values				$<$0.001	$<$0.001	$<$0.001		

6.2.2 Sensitivity Analysis

Here, possible influential observations are analysed with measure described in Section 5.2. Figure 5 displays the results of generalized Cook distance,$GD_i\left(\boldsymbol{\tau}\right)$. Based on Figure 5, cases 41 and 48 can be considered as possible influential observations.

6.2.3 Residual Analysis

Figure 6 displays the index plot of the modified deviance residuals and its Q-Q plot against to $N(0,1)$ quantiles. Based on Figure 6, we conclude that none of observed values appears as possible outliers. Therefore, the fitted model is appropriate for these data set.

7 Conclusion

In this study, we introduced a new flexible extension of the Generalized Half-Normal lifetime model as well as a new log-location regression model based on the proposed model. Some useful characterization results are presented and some mathematical properties are derived. The maximum likelihood method is used to estimate the model parameters by means of a graphical Monte Carlo simulation study. We show that the new log-location regression lifetime model can be very useful in analysing real data and provide more realistic fits than other regression models. Index plot of the modified deviance residual and Q-Q plot for modified deviance residual are presented to illustrate that our new model is more appropriate to HIV data set than other competitive models like log-odd log-logistic generalized half-normal regression model and log-generalized half-normal regression model. The sensitivity analysis is used via the index plot of generalized cook distance to discover the possible influential observations.

References

[1] K. Cooray and M.M.A. Ananda. A generalization of the half-normal distribution with applications to lifetime data. *Communication in Statistics - Theory and Methods*, 37(9):1323–1337, 2008. doi: 10.1080/03610920701826088.

[2] H.M. Yousof, A.Z. Afify, G.G. Hamedani, and G. Aryal. The burr x generator of distributions for lifetime data. *Journal of Statistical Theory and Applications*, 16(3):288–305, 2017. http://download.atlantis-press.com/php/download_paper.php?id=25883863.

[3] R.R. Pescim, C.G.B. Demetrio, G.M. Cordeiro, E.M.M. Ortega, and M.R. Urbano. The beta generalized half-normal distribution. *Computational Statistics and Data Analysis*, 54(4):945–957, 2010. doi: 10.1016/j.csda.2009.10.007.

[4] G.M. Cordeiro, R.R. Pescim, and E.M.M. Ortega. The kumaraswamy generalized halfnormal distribution for skewed positive data. *Journal of Data Science*, 10:195–224, 2012. http://www.jds-online.com/files/JDS-1010.pdf.

[5] R.R. Pescim, E.M.M. Ortega, G.M. Cordeiro, C.G.B. Demetrio, and G.G. Hamedani. The log-beta generalized half-normal regression model. *Journal of Statistical Theory and Applications*, 12:330–347, 2013. doi: 10.2991/jsta.2013.12.4.2.

[6] T.G. Ramires, E.M.M. Ortega, G.M. Cordeiro, and G.G. Hamedani. The beta generalized half-normal geometric distribution. *Studia Scientiarum Mathematicarum Hungarica*, 50(4):523–554, 2013. doi: 10.1556/SScMath.50.2013.4.1258.

[7] F. Merovci, M. Alizadeh, H.M. Yousof, and Hamedani G.G. The exponentiated transmuted-g family of distributions: theory and applications.

Communications in Statistics-Theory and Method, 46(21):10800–10822, 2017. doi: 10.1080/03610926.2016.1248782.

[8] W. Glanzel. A characterization theorem based on truncated moments and its application to some distribution families. *In: Mathematical Statistics and Probability Theory, Bauer P., Konecny F., Wertz W. (eds). Springer, Dordrecht*, B:75–84, 1987. doi: 10.1007/978-94-009-3965-3_8.

[9] W. Glanzel. Some consequences of a characterization theorem based on truncated moments. *Statistics: A Journal of Theoretical and Applied Statistics*, 21(4):613–618, 1990. doi: 10.1080/02331889008802273.

[10] G.M. Cordeiro, M. Alizadeh, R.R. Pescim, and E.M.M. Ortega. The odd log-logistic generalized half-normal lifetime distribution: properties and applications. *Communication in Statistics- Theory and Methods*, 46(9):4195–4214, 2016. doi: 10.1080/03610926.2015.1080841.

[11] H. Exton. *Handbook of hypergeometric integrals: theory, applications, tables, computer programs*. Halsted Press, New York, 1978. ISBN 978-0853121220.

[12] R.M Aarts. Lauricella functions. *From MathWorld - A Wolfram Web Resource, created by Eric W. Weisstein*, 2000. http://mathworld.wolfram.com/LauricellaFunctions.html.

[13] M. Trott. *The Mathematica Guidebook for Symbolics*. Springer-Verlag, New York, 2006. ISBN 978-0-387-28815-4.

[14] T.R. Fleming and D.P. Harrington. *Counting process and survival analysis*. Series in Probability and Statistics. Wiley, New York, 2013. ISBN 978-0-471-76988-0.

[15] T.M. Therneau, P.M. Grambsch, and T.R. Fleming. Applications of soft sets in ideal theory of bck/bci-algebras. *Biometrika*, 77(1):147–160, 1990. doi: 10.2307/2336057.

[16] A.J. Gross and V.A. Clark. *Survival Distributions: Reliability Applications in the Biomedical Sciences*. Wiley, 1975. ISBN 978-0471328179.

[17] R.R. Pescim, E.M. Ortega, G.M. Cordeiro, and M. Alizadeh. A new log-location regression model: estimation, influence diagnostics and residual analysis. *Journal of Applied Statistics*, 44(2):233–252, 2017. doi: 10.1080/02664763.2016.1168368.

[18] D.W. Hosmer, S. Lemeshow, and S. May. *Applied Survival Analysis: Regression Modeling of Time-to-Event Data, 2 edition.* Wiley, 1998. ISBN 978-0-471-75499-2.

Appendix A

Theorem 1. *Let* $(\Omega, \mathcal{F}, \mathbf{P})$ *be a given probability space and let* $H = [a, b]$ *be an interval for some* $d < b$ $(a = -\infty,\ b = \infty$ *might as well be allowed*$)$. *Let* $X : \Omega \rightarrow H$ *be a continuous random variable with the distribution function* F *and let* q_1 *and* q_2 *be two real functions defined on* H *such that*

$$\mathbf{E}\left[q_2(X) \mid X \geq x\right] = \mathbf{E}\left[q_1(X) \mid X \geq x\right] \eta(x), \quad x \in H,$$

is defined with some real function η. *Assume that* $q_1, q_2 \in C^1(H)$, $\eta \in C^2(H)$ *and* F *is twice continuously differentiable and strictly monotone function on the set* H. *Finally, assume that the equation* $\eta q_1 = q_2$ *has no real solution in the interior of* H. *Then* F *is uniquely determined by the functions* q_1, q_2 *and* η , *particularly*

$$F(x) = \int_a^x C \left| \frac{\eta'(u)}{\eta(u)\, q_1(u) - q_2(u)} \right| \exp\left(-s(u)\right)\, du\ ,$$

where the function s *is a solution of the differential equation* $s' = \frac{\eta'\, q_1}{\eta q_1 - q_2}$ *and* C *is the normalization constant, such that* $\int_H dF = 1$.

Appendix B

R code for parameter estimation of BrXGHN distribution.

```
library(AdequacyModel)
cdf=function(par,x)
{ gam=par[1]
  lambda=par[2]
  theta=par[3]
```

```
  y=x
  G=2*pnorm((y/theta)^lambda)-1
  g=sqrt(2/pi)*(lambda/y)*(y/theta)^(lambda)
  *exp((-1/2)*(y/theta)^(2*lambda))

  f=(1-exp(-(G/(1-G))^2))^gam
  return(f)}

pdf=function(par,x)
{ gam=par[1]
  lambda=par[2]
  theta=par[3]

  y=x
  G=2*pnorm((y/theta)^lambda)-1
  g=sqrt(2/pi)*(lambda/y)*(y/theta)^(lambda)
  *exp((-1/2)*(y/theta)^(2*lambda))

  f=((2*gam*g*G)/(1-G)^3)*exp(-(G/(1-G))^2)
  *(1-exp(-(G/(1-G))^2))^(gam-1)
  return(f)}

fit=goodness.fit(pdf=pdf, cdf=cdf,
starts = c(gam1,lambda1,theta1), data = data,
method="N", domain=c(0,Inf))
```

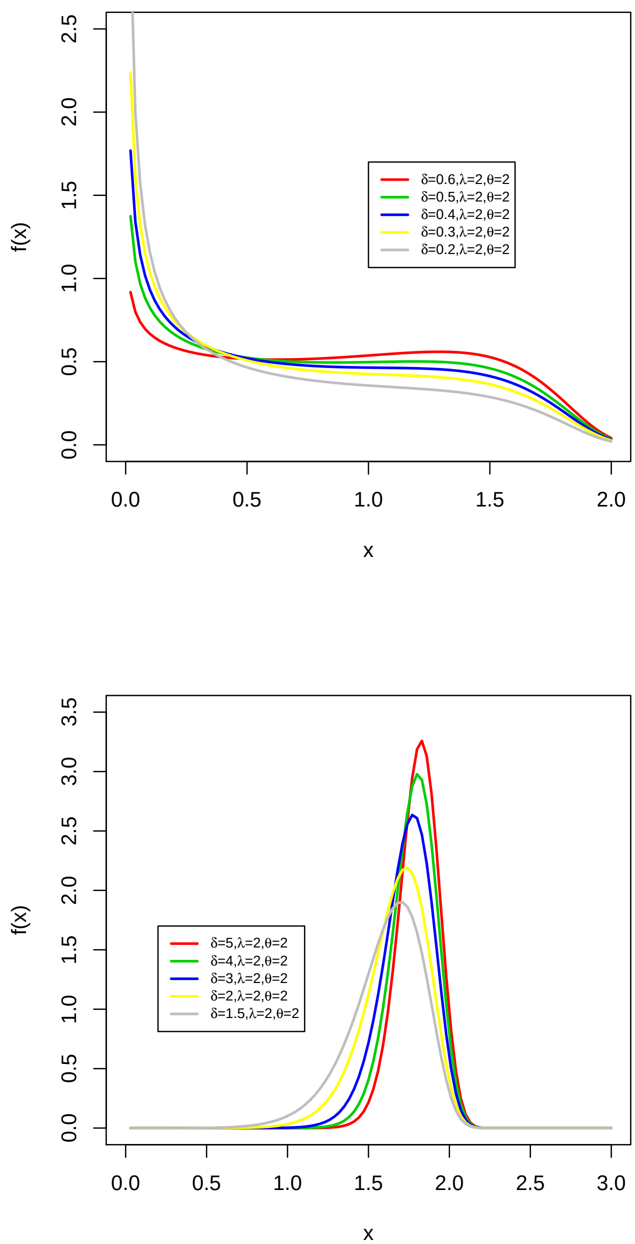


Fig. 1: The pdf plots of BrXGHN distribution for several parameter values.

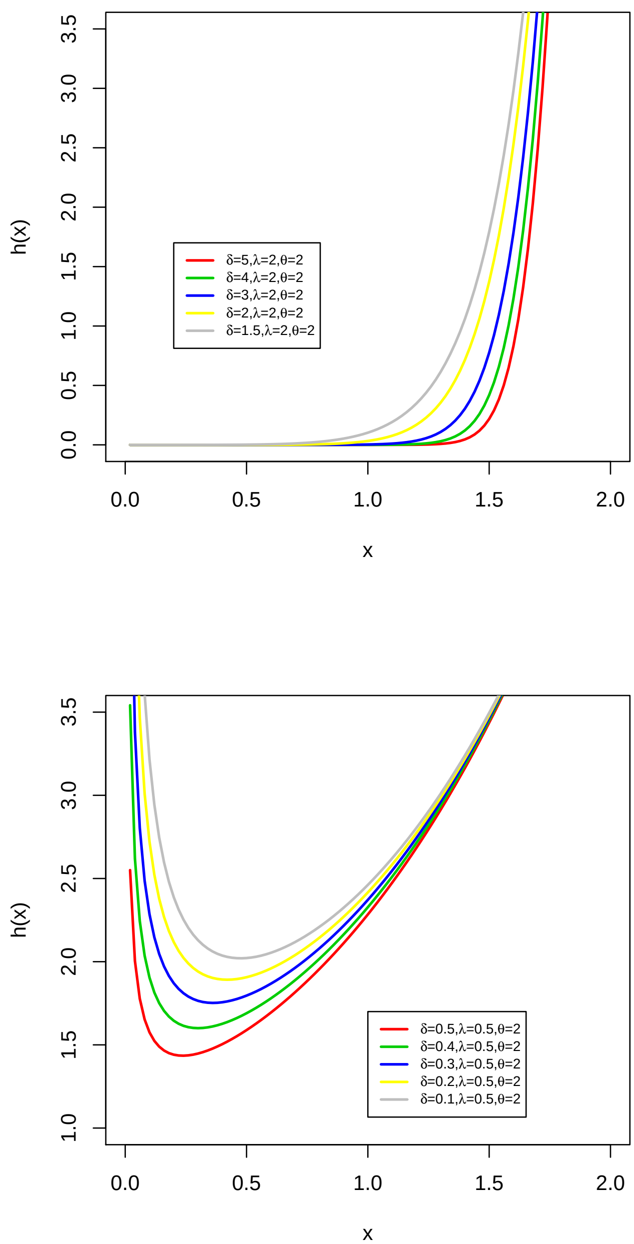
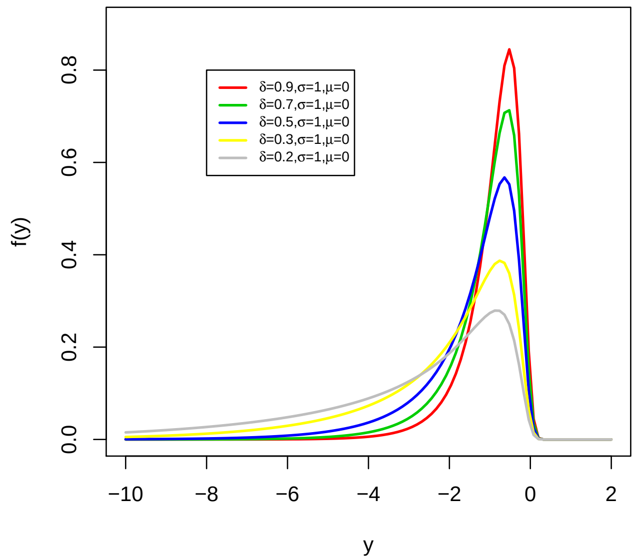


Fig. 2: The hrf plots of BrXGHN distribution for several parameter values.



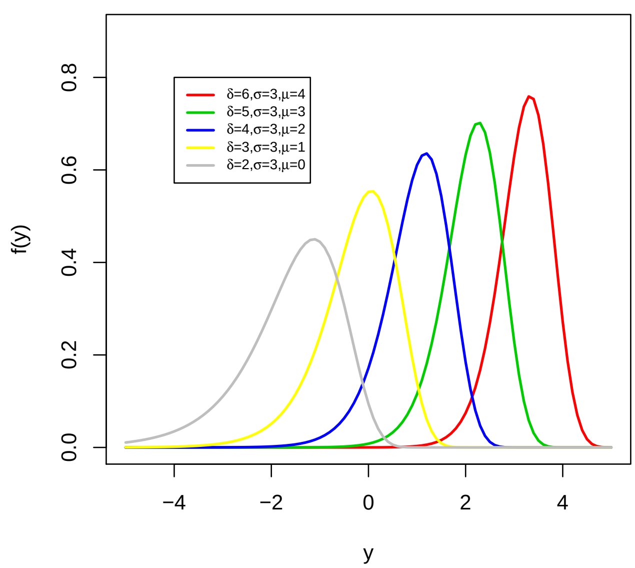


Fig. 3: Plots of the LBrXGHN density function for some parameter values.

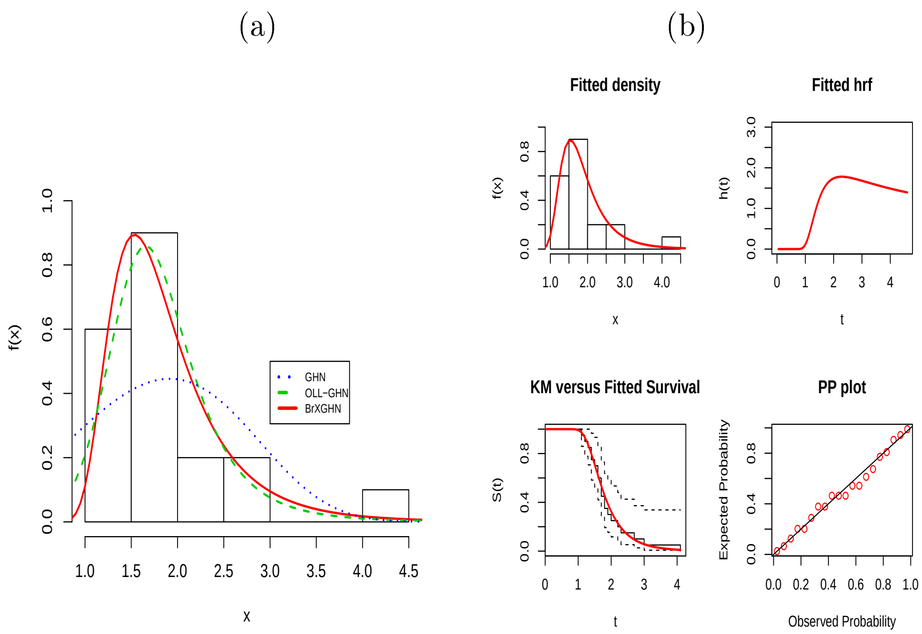


Fig. 4: (a) Fitted densities of fitted models and (b) fitted hrf and P-P plot of the BrXGHN model for used data set.

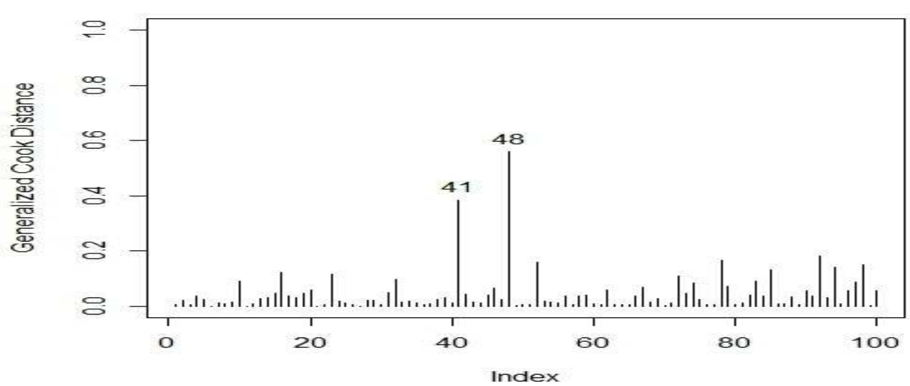
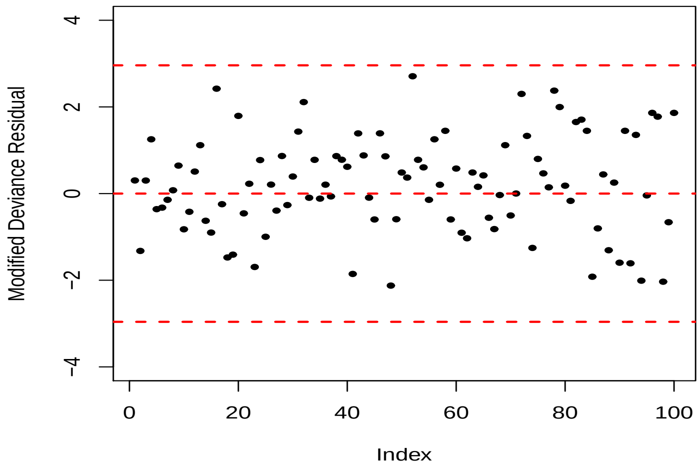
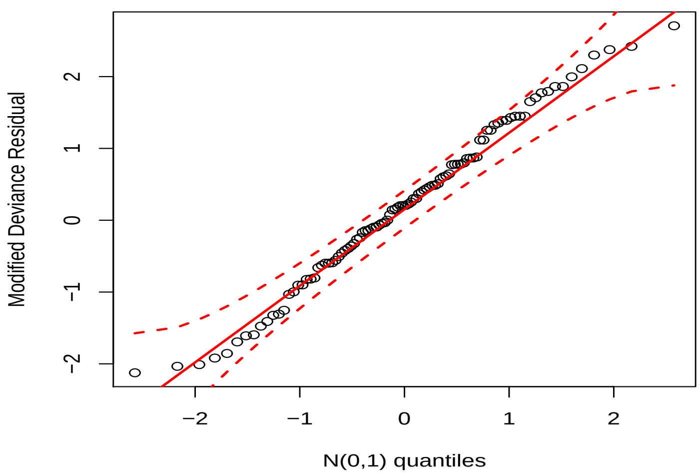


Fig. 5: Index plot of generalized cook distance.



(a)



(b)

Fig. 6: (a) Index plot of the modified deviance residual and (b) Q-Q plot for modified deviance residual.