

RESEARCH

Open Access



A new generalization of generalized half-normal distribution: properties and regression models

Emrah Altun^{1*}, Haitham M. Yousof² and G.G. Hamedani³

*Correspondence:

emrahaltun@bartin.edu.tr

¹Department of Statistics, Bartin University, Bartin 74100, Turkey
Full list of author information is available at the end of the article

Abstract

In this paper, a new extension of the generalized half-normal distribution is introduced and studied. We assess the performance of the maximum likelihood estimators of the parameters of the new distribution via simulation study. The flexibility of the new model is illustrated by means of four real data sets. A new log-location regression model based on the new distribution is also introduced and studied. It is shown that the new log-location regression model can be useful in the analysis of survival data and provides more realistic fits than other competitive regression models.

Keywords: Regression, Residuals, Simulation

AMS 2010 Subject Classification: 60E05, 62J05

Introduction

The generalized half-normal (GHN) distribution has been widely modified and studied in recent years and various authors developed new generalizations of it. Following an idea due to Eugene et al. (2002), Pescim et al. (2017) introduced the beta generalized half-Normal (BGHN) distribution with applications to myelogenous leukemia data. Cordeiro et al. (2012) defined the Kumaraswamy generalized half-normal (KwGHN) distribution for censored data. More recently, Cordeiro et al. (2013) studied some of the mathematical properties of the BGHN distribution proposed by Pescim et al. (2010b). Pescim et al. (2013) proposed the log-linear regression model based on the BGHN distribution, while Ramires et al. (2013) defined the beta generalized half-normal geometric (BGHNG) distribution in order to achieve wider diversity among the density and failure rate functions.

The GHN density function (Cooray and Ananda 2008) with shape parameter $\lambda > 0$ and scale parameter $\theta > 0$ is given (for $x > 0$) by

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

and its cumulative distribution function (cdf) depends on the error function

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

where

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

and

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

The n th moment of the random variable X with cdf (2) is

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

where $\Gamma(\cdot)$ is the gamma function. The HN distribution is a sub-model of GHN when $\lambda = 1$.

The goal of this paper is to propose the first generalization of the generalized half-normal distribution using the Zografos–Balakrishnan Odd Log-Logistic-G (“ZBOLL-G” for short) family of distributions. For an arbitrary baseline cdf $G(x)$, Cordeiro et al. (2015) proposed the probability density function (pdf) $f(x)$ and the cdf $F(x)$ of the ZBOLL-G family of distributions with two additional shape parameters $\beta > 0$ and $\alpha > 0$ as

$$f(x; \beta, \alpha, \xi) = \frac{\alpha}{\Gamma(\beta)} \frac{g(x; \xi) G^{\alpha-1}(x; \xi) \bar{G}^{\alpha-1}(x; \xi)}{[G^\alpha(x; \xi) + \bar{G}^\alpha(x; \xi)]^2} \left\{ -\log \left[\frac{\bar{G}^\alpha(x; \xi)}{G^\alpha(x; \xi) + \bar{G}^\alpha(x; \xi)} \right] \right\}^{\beta-1} \tag{3}$$

and

$$F(x; \beta, \alpha, \xi) = \frac{1}{\Gamma(\beta)} \gamma \left(\beta, -\log \left[1 - \frac{G^\alpha(x; \xi)}{G^\alpha(x; \xi) + \bar{G}^\alpha(x; \xi)} \right] \right), \quad x \in \mathfrak{N}, \tag{4}$$

where ξ denotes the parameter vector of the baseline distribution. We use Eqs. (1), (2) and (3) to obtain the four-parameter ZBOLLGHN pdf (for $x > 0$)

$$f(x) = \frac{\alpha \sqrt{\frac{2}{\pi}} \left(\frac{x}{\theta}\right)^\lambda \exp \left[-\frac{1}{2} \left(\frac{x}{\theta}\right)^{2\lambda} \right] \left\{ 2\Phi \left[\left(\frac{x}{\theta}\right)^\lambda \right] - 1 \right\}^{\alpha-1} \left\{ 2 - 2\Phi \left[\left(\frac{x}{\theta}\right)^\lambda \right] \right\}^{\alpha-1}}{\Gamma(\beta) \left(\left\{ 2\Phi \left[\left(\frac{x}{\theta}\right)^\lambda \right] - 1 \right\}^\alpha + \left\{ 2 - 2\Phi \left[\left(\frac{x}{\theta}\right)^\lambda \right] \right\}^\alpha \right)^2} \times \left[-\log \left(\frac{\left\{ 2 - 2\Phi \left[\left(\frac{x}{\theta}\right)^\lambda \right] \right\}^\alpha}{\left\{ 2\Phi \left[\left(\frac{x}{\theta}\right)^\lambda \right] - 1 \right\}^\alpha + \left\{ 2 - 2\Phi \left[\left(\frac{x}{\theta}\right)^\lambda \right] \right\}^\alpha} \right) \right]^{\beta-1}, \tag{5}$$

where $\alpha > 0, \beta > 0, \lambda > 0$ are shape parameters and θ is the scale parameter. The corresponding cdf is given by

$$F(x) = \frac{1}{\Gamma(\beta)} \gamma \left[\beta, -\log \left(1 - \frac{\left\{ 2\Phi \left[\left(\frac{x}{\theta}\right)^\lambda \right] - 1 \right\}^\alpha}{\left\{ 2\Phi \left[\left(\frac{x}{\theta}\right)^\lambda \right] - 1 \right\}^\alpha + \left\{ 2 - 2\Phi \left[\left(\frac{x}{\theta}\right)^\lambda \right] \right\}^\alpha} \right) \right], \quad x \geq 0, \tag{6}$$

where $\gamma(\beta, z) = \int_z^\infty t^{\beta-1} \exp(-t) dt$ denotes the complementary incomplete gamma function. Henceforth, we denote a random variable X with pdf (5) by $X \sim \text{ZBOLL-GHN}(\beta, \alpha, \lambda, \theta)$. The sub-models of (5) are given in Table 1.

Table 1 Submodels of ZBOLL-GHN distribution

Distribution	β	α	λ	θ	Author
Gamma-GHN	β	1	λ	θ	New
Gamma-HN	β	1	1	θ	New
OLL-GHN	1	α	λ	θ	Cordeiro et al. (2016b)
OLL-HN	1	α	1	θ	Cordeiro et al. (2016b)
GHN	1	1	λ	θ	Cooray and Ananda (2008)
HN	1	1	1	θ	Cooray and Ananda (2008)

We investigate the possible hazard rate function (hrf) and pdf shapes of ZBOLL-GHN distribution. Figure 1 displays the pdf shapes of ZBOLL-GHN distribution. Based on the Fig. 1, ZBOLL-GHN pdf has the following shapes: left-skewed, right-skewed, symmetric and bimodal. Figure 2 displays the hrf shapes of ZBOLL-GHN distribution. From Fig. 2, we conclude that the ZBOLL-GHN hrf has the following shapes: increasing, decreasing, upside-down and bathtub.

Following Cordeiro et al. (2016a), equation (6) can be expressed as

$$F(x) = \sum_{w=0}^{\infty} b_w \Pi_w(x; \lambda, \theta),$$

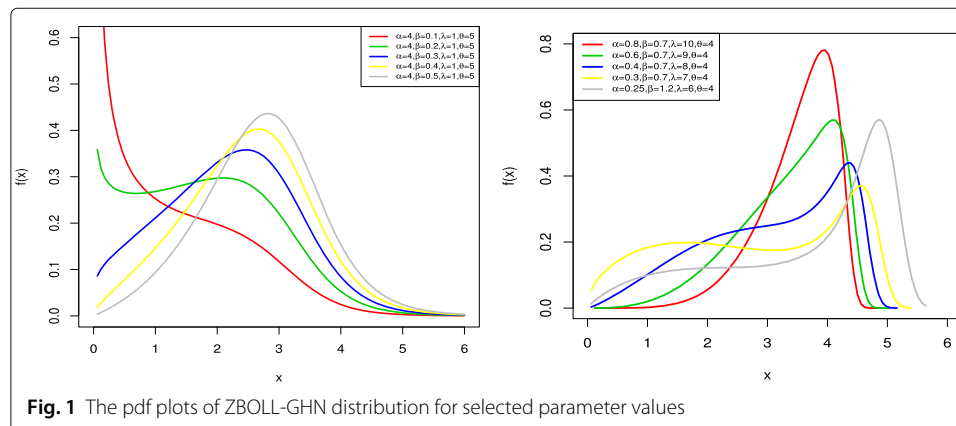
where

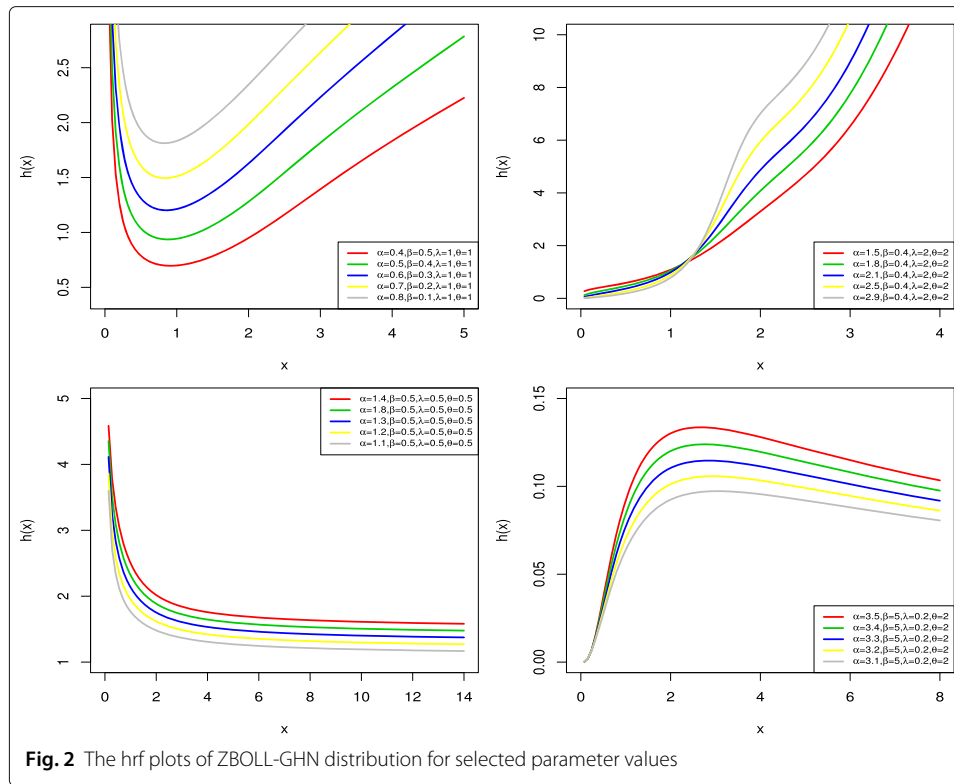
$$b_w = \frac{1}{\Gamma(\beta)} \sum_{i,k=0}^{\infty} \sum_{j=0}^k \frac{(-1)^{i+j+k}}{(\beta+i-j) i!} p_{j,k} a_w(\beta, \alpha, i, k) \binom{k-\beta-i}{k} \binom{k}{j}$$

and $\Pi_w(x; \lambda, \theta) = [G(x; \lambda, \theta)]^w$ denotes the cdf of the exp-GHN distribution with the power parameter w . The pdf (5) reduces to

$$f(x) = \sum_{w=0}^{\infty} b_{w+1} \pi_{w+1}(x; \lambda, \theta), \tag{7}$$

where $\pi_{w+1}(x; \lambda, \theta) = (w+1)g(x; \lambda, \theta)[G(x; \lambda, \theta)]^w$ denotes the pdf of the exp-GHN distribution with the power parameter $w+1$. For the definitions of $p_{j,k}$ and $a_w(\beta, \alpha, i, k)$, please see Cordeiro et al. (2016a). Equation (7) reveals that the density function of X is a linear combination of the exp-GHN densities. Thus, some of the structural properties of the ZBOLL-GHN distribution such as ordinary and incomplete moments and generating function can be obtained from well-established properties of the exp-GHN distribution.





We are motivated to introduce the ZBOLL-GHN distribution since it contains a number of aforementioned known lifetime models as illustrated in Table 1. The new distribution exhibits increasing, decreasing, upside-down as well as bathtub hazard rates as illustrated in Fig. 2. It is shown that the new distribution can be viewed as a mixture of the two-parameter GHN model. It can also be viewed as a suitable model for fitting the left-skewed, right-skewed, symmetric and bimodal data. The ZBOLL-GHN distribution outperforms several of the well-known lifetime distributions with respect to four real data applications as illustrated in “Applications” section. The new log-location regression model based on the ZBOLL-GHN distribution provides better fits than log BGHN, log GHN and log-Weibull models for volatage data set. Based on the residual analysis (martingale and modified deviance residuals) for the new log-location regression model (log ZBOLL-GHN), we conclude that none of the observed values appear as possible outliers. Thus, it is clear that the fitted model is appropriate for the voltage data set.

The rest of the paper is organized as follows. In “Estimation” section, the maximum likelihood method is used to estimate the model parameters. The performance of maximum likelihood estimators of the model parameters are investigated by means of a Monte Carlo simulation study when n is finite. A new log-location regression model as well as residual analysis are presented in “A new log-location regression model” section. Four applications to real data sets illustrate empirically the importance of the new model in “Applications” section. Finally, a summary is provided in “Summary” section.

Estimation

If X follows the ZBOLL-GHN distribution with vector of parameters $\Psi = (\beta, \alpha, \lambda, \theta)^T$. The log-likelihood function for a single observation x of X is given by

$$\begin{aligned} \ell(\Psi) = & \log(\alpha) + \log\left(\sqrt{\frac{2}{\pi}}\right) + \log(\lambda) - \log(x) + \lambda \log(x\theta^{-1}) - \frac{1}{2}(x\theta^{-1})^{2\lambda} \\ & + (\alpha - 1) \log(w - 1) + (\alpha - 1) \log(2 - w) - \log[\Gamma(\beta)] \\ & - 2 \log[(w - 1)^\alpha + (2 - w)^\alpha] + (\beta - 1) \log\left[-\log\left(\frac{(2 - w)^\alpha}{(w - 1)^\alpha + (2 - w)^\alpha}\right)\right]. \end{aligned}$$

The components of the unit score vector $U = U(\Psi) = (\partial\beta/\partial\ell, \partial\alpha/\partial\ell, \partial\lambda/\partial\ell, \partial\theta/\partial\ell)^T$ are available if needed. For a random sample $x = (x_1, \dots, x_n)^T$ of size n from X , the total log-likelihood is

$$\ell_n(\Psi) = \sum_{i=0}^n \ell^{(i)}(\Psi),$$

where $\ell^{(i)}(\Psi)$ is the log-likelihood for the i^{th} observation. The total score function is

$$U_n = \sum_{i=0}^n U^{(i)},$$

where $U^{(i)}$ has the form given before. Maximization of $\ell(\Psi)$ (or $\ell_n(\Psi)$) can be easily performed using well-established routines such as the **nlm** or **optim** in the R statistical package. Setting these equations equal to zero, $U(\Psi) = 0$, and solving them simultaneously gives the MLE $\hat{\Psi}$ of Ψ . These equations cannot be solved analytically and statistical software can be used to evaluate them numerically using iterative techniques such as the Newton-Raphson algorithm.

The parameter estimation procedure of ZBOLL-GHN model can be summarized as follows:

- The **optim** function of R software is used to minimize the minus log-likelihood function of GHN model by means of the Nelder-Mead (NM) optimization method. There is no need to provide the derivatives of the objective function for NM method.
- The estimated parameters of GHN distribution are used as initial values of the ZBOLL-GHN model. The initial values of the additional parameters α and β are chosen as 1. Note that the ZBOLL-GHN model reduces to GHN model when the parameters $\alpha = \beta = 1$. Then, the parameter estimation of ZBOLL-GHN model are obtained with the **optim** function as given in the first step.
- The inverse of estimated Hessian matrix is used to obtain the corresponding standard errors.

Simulation study

In this subsection, the performance of the maximum likelihood estimators of the ZBOLL-GHN parameters are evaluated via a Monte Carlo simulation study with 10,000 replications. The coverage probabilities (CPs), mean square errors (MSES) and the bias of the parameter estimates, estimated average lengths (ALs) are calculated by means of R software. We generate $N = 10,000$ samples of sizes $n = 50, 55, \dots, 500$ from the ZBOLL-GHN distribution with $\alpha = 0.8, \beta = 7, \lambda = 9, \theta = 4$. Let $(\hat{\alpha}, \hat{\beta}, \hat{\lambda}, \hat{\theta})$ be the MLEs of the new model parameters and $(s_{\hat{\alpha}}, s_{\hat{\beta}}, s_{\hat{\lambda}}, s_{\hat{\theta}})$ be the standard errors of the MLEs. The estimated biases and MSEs are given by

$$\widehat{Bias}_\epsilon(n) = \frac{1}{N} \sum_{i=1}^N (\hat{\epsilon}_i - \epsilon)$$

and

$$\widehat{MSE}_\epsilon(n) = \frac{1}{N} \sum_{i=1}^N (\hat{\epsilon}_i - \epsilon)^2,$$

for $\epsilon = \alpha, \beta, \lambda, \theta$. The CPs and ALs are given, respectively, by

$$CP_\epsilon(n) = \frac{1}{N} \sum_{i=1}^N I(\hat{\epsilon}_i - 1.95996s_{\hat{\epsilon}_i}, \hat{\epsilon}_i + 1.95996s_{\hat{\epsilon}_i})$$

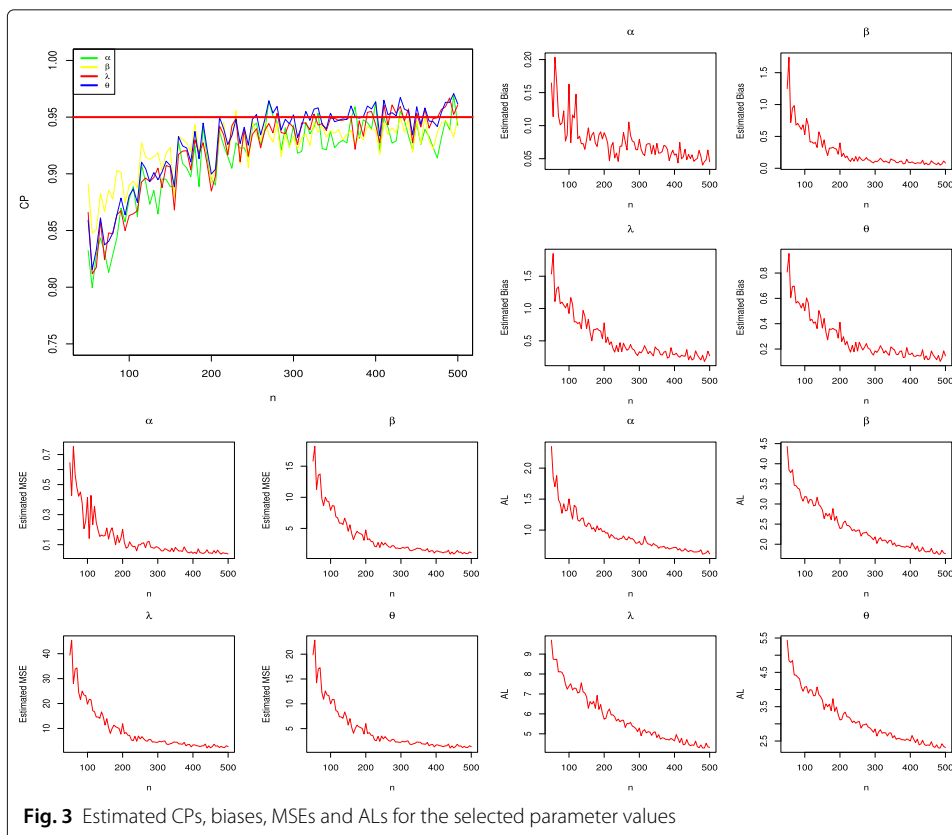
and

$$AL_\epsilon(n) = \frac{3.919928}{N} \sum_{i=1}^N s_{\hat{\epsilon}_i}.$$

Figure 3 displays the numerical results for the above measures. We list below the results from these plots:

- ✓ The estimated biases decrease when the sample size n increases,
- ✓ The estimated MSEs decay toward zero as n increases,
- ✓ The CPs are near 0.95 and approach the nominal value when the sample size n increases,
- ✓ The ALs decrease for all parameters when the sample size n increases.

These results reveal the consistency property of the MLEs.



A new log-location regression model

Let X denote a random variable following the ZBOLL-GHN distribution (5) and let $Y = \log(X)$. The density function of Y (for $y \in \Re$) and replacing $\mu = \log(\theta)$, $\sigma = \sqrt{2}/2\lambda$ can be expressed as

$$\begin{aligned}
 f(y) = & \frac{\alpha}{\Gamma(\beta)\sigma\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \exp \left[\left(\frac{y-\mu}{\sigma} \right) \sqrt{2} \right] + \left(\frac{y-\mu}{\sigma} \right) \frac{\sqrt{2}}{2} \right\} \left[2\Phi \left[\exp \left[\left(\frac{y-\mu}{\sigma} \right) \frac{\sqrt{2}}{2} \right] \right] - 1 \right]^{\alpha-1} \\
 & \times \left(1 - \left[2\Phi \left[\exp \left[\left(\frac{y-\mu}{\sigma} \right) \frac{\sqrt{2}}{2} \right] \right] - 1 \right] \right)^{\alpha-1} \\
 & \times \left[\left[2\Phi \left[\exp \left[\left(\frac{y-\mu}{\sigma} \right) \frac{\sqrt{2}}{2} \right] \right] - 1 \right]^{\alpha} + \left(1 - \left[2\Phi \left[\exp \left[\left(\frac{y-\mu}{\sigma} \right) \frac{\sqrt{2}}{2} \right] \right] - 1 \right] \right)^{\alpha-2} \right. \\
 & \left. \times \left\{ -\log \left[\frac{\left(1 - \left[2\Phi \left[\exp \left[\left(\frac{y-\mu}{\sigma} \right) \frac{\sqrt{2}}{2} \right] \right] - 1 \right)^\alpha}{\left[2\Phi \left[\exp \left[\left(\frac{y-\mu}{\sigma} \right) \frac{\sqrt{2}}{2} \right] \right] - 1 \right]^\alpha + \left(1 - \left[2\Phi \left[\exp \left[\left(\frac{y-\mu}{\sigma} \right) \frac{\sqrt{2}}{2} \right] \right] - 1 \right)^\alpha} \right] \right\}^{\beta-1} \right.
 \end{aligned} \tag{8}$$

where $\mu \in \Re$ is the location parameter, $\sigma > 0$ is the scale parameter and $\alpha > 0$ and $\beta > 0$ are the shape parameters. We refer to Eq. (8) as the pdf of LZBOLL-GHN distribution, say $Y \sim \text{LZBOLL-GHN}(\alpha, \beta, \mu, \sigma)$. The survival function corresponding to (8) is given by

$$S(y) = 1 - \frac{1}{\Gamma(\beta)} \gamma \left(\beta, -\log \left[1 - \frac{\left[2\Phi \left[\exp \left[\left(\frac{y-\mu}{\sigma} \right) \frac{\sqrt{2}}{2} \right] \right] - 1 \right]^\alpha}{\left[2\Phi \left[\exp \left[\left(\frac{y-\mu}{\sigma} \right) \frac{\sqrt{2}}{2} \right] \right] - 1 \right]^\alpha + \left(1 - \left[2\Phi \left[\exp \left[\left(\frac{y-\mu}{\sigma} \right) \frac{\sqrt{2}}{2} \right] \right] - 1 \right)^\alpha} \right] \right) \tag{9}$$

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

$$\begin{aligned}
 f(z) = & \frac{\alpha}{\Gamma(\beta)\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \exp \left[(z) \sqrt{2} \right] + \left(\frac{y-\mu}{\sigma} \right) \frac{\sqrt{2}}{2} \right\} \left[2\Phi \left[\exp \left[(z) \frac{\sqrt{2}}{2} \right] \right] - 1 \right]^{\alpha-1} \\
 & \times \left(1 - \left[2\Phi \left[\exp \left[(z) \frac{\sqrt{2}}{2} \right] \right] - 1 \right] \right)^{\alpha-1} \\
 & \times \left[\left[2\Phi \left[\exp \left[(z) \frac{\sqrt{2}}{2} \right] \right] - 1 \right]^{\alpha} + \left(1 - \left[2\Phi \left[\exp \left[(z) \frac{\sqrt{2}}{2} \right] \right] - 1 \right] \right)^{\alpha-2} \right. \\
 & \left. \times \left\{ \log \left[\frac{\left(1 - \left[2\Phi \left[\exp \left[(z) \frac{\sqrt{2}}{2} \right] \right] - 1 \right)^\alpha}{\left[2\Phi \left[\exp \left[(z) \frac{\sqrt{2}}{2} \right] \right] - 1 \right]^\alpha + \left(1 - \left[2\Phi \left[\exp \left[(z) \frac{\sqrt{2}}{2} \right] \right] - 1 \right)^\alpha} \right] \right\}^{\beta-1} \right.
 \end{aligned} \tag{10}$$

Figure 4 provides some plots of the density function (8) for selected parameter values. They reveal that this distribution is a good candidate to model left skewed and symmetric data sets.

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

$$y_i = \mathbf{v}_i^T \boldsymbol{\beta} + \sigma z_i, \quad i = 1, \dots, n, \tag{11}$$

where the random error z_i has density function (10), $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$, and $\sigma > 0, \alpha > 0$ and $\beta > 0$ are unknown parameters. The parameter $\mu_i = \mathbf{v}_i^T \boldsymbol{\beta}$ is the location of y_i . The

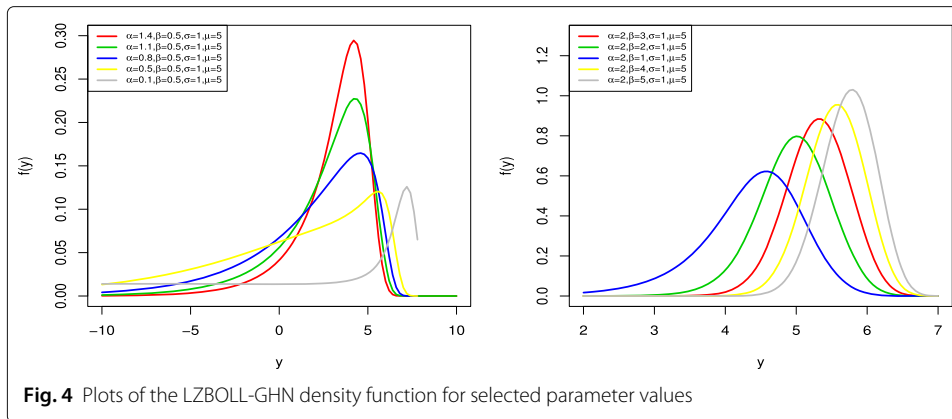


Fig. 4 Plots of the LZBOLL-GHN density function for selected parameter values

location parameter vector $\mu = (\mu_1, \dots, \mu_n)^T$ is represented by a linear model $\mu = V\beta$, where $V = (v_1, \dots, v_n)^T$ is a known model matrix.

The LZBOLL-GHN model (11) provides new opportunities for modeling several types of data sets. This model contains two important regression models as its sub-models: (i) for $\beta = 1$, the LZBOLL-GHN model reduces to log-OLL-GHN regression model introduced by Pescim et al. (2017); (ii) for $\alpha = \beta = 1$, the LZBOLL-GHN model reduces to log-GHN regression model.

Let F and C be the sets of individuals for which y_i is the log-lifetime or log-censoring, respectively. Assume that the observed lifetimes and censoring times are independent. The log-likelihood function for the vector of parameters $\Theta = (\alpha, \beta, \sigma, \beta^T)^T$ from model (11) is given by $l(\Theta) = \sum_{i \in F} l_i(\Theta) + \sum_{i \in C} l_i^{(c)}(\Theta)$, where $l_i(\Theta) = \log[f(y_i)]$, $l_i^{(c)}(\Theta) = \log[S(y_i)]$. The $f(y_i)$ and $S(y_i)$ are defined in(8) and (9), respectively. The total log-likelihood function for Θ is given by

$$\begin{aligned} \ell(\Theta) = & r \log \left(\frac{\alpha}{\Gamma(\beta) \sigma \sqrt{2\pi}} \right) - \frac{1}{2} \sum_{i \in F} \exp(z_i \sqrt{2}) + \frac{\sqrt{2}}{2} \sum_{i \in F} z_i \\ & + (\alpha - 1) \sum_{i \in F} \log[u_i - 1] + (\alpha - 1) \sum_{i \in F} \log(2 - u_i) - 2 \sum_{i \in F} \log \left[[u_i - 1]^\alpha + (2 - u_i)^\alpha \right] \\ & + (\beta - 1) \sum_{i \in F} \log \left\{ -\log \left[\frac{(2 - u_i)^\alpha}{[u_i - 1]^\alpha + (2 - u_i)^\alpha} \right] \right\} \\ & + \sum_{i \in C} \log \left\{ 1 - \frac{1}{\Gamma(\beta)} \gamma \left(\beta, -\log \left[1 - \frac{[u_i - 1]^\alpha}{[u_i - 1]^\alpha + (2 - u_i)^\alpha} \right] \right) \right\}, \end{aligned} \tag{12}$$

where $u_i = 2\Phi[\exp(z_i \sqrt{2}/2)]$, $z_i = (y_i - \mu_i)/\sigma$, and r is the number of uncensored observations (failures). The MLE $\hat{\Theta}$ of the vector of unknown parameters can be evaluated by maximizing the log-likelihood (12). The R software is used to estimate unknown parameters of LZBOLL-GHN regression model

The likelihood ratio (LR) statistic can be used for comparing some sub-models of LZBOLL-GHN regression model. For example, the LR statistic can be used to discriminate between the LZBOLL-GHN and LGHN regression models since they are nested models, or equivalently to test $H_0 : \alpha = \beta = 1$. The LR statistic reduces to $w = 2 \left[\ell(\hat{\alpha}, \hat{\beta}, \hat{\sigma}, \hat{\beta}) - \ell(1, 1, \tilde{\sigma}, \tilde{\beta}) \right]$, where $(\hat{\alpha}, \hat{\beta}, \hat{\sigma}, \hat{\beta})$ are the unrestricted MLEs and $(1, 1, \tilde{\sigma}, \tilde{\beta})$ are the restricted estimates under H_0 . The statistic w is asymptotically (as

Table 2 MLEs and their SEs of the fitted models and goodness-of-fit statistics for first data set

Models	α	β	λ	θ	$-\ell$	AIC	A*	W*	K-S	p-value
ZBOLL-GHN	0.143	1.360	4.049	8.243	73.053	154.107	0.620	0.098	0.160	0.383
	0.022	0.177	0.003	0.003						
B-GHN	0.233	1.327	4.876	13.924	81.868	171.737	2.157	0.348	0.315	0.004
	0.047	0.439	0.004	0.004						
Gamma-GHN		0.238	4.945	13.941	81.851	169.702	2.193	0.354	0.305	0.005
		0.042	0.003	0.003						
OLL-GHN	0.165		4.016	8.488	75.522	157.043	0.750	0.117	0.294	0.008
	0.032		0.141	0.133						
GHN			1.491	10.226	87.927	179.853	3.074	0.521	0.278	0.014
			0.255	0.903						

$n \rightarrow \infty$) distributed as χ_k^2 , where k is difference of two parameter vectors of nested models. For example, take $k = 2$ for the above hypothesis test.

Residual analysis

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

Martingale residual

The martingale residuals is defined in counting process and takes values between +1 and $-\infty$ (see for details, Fleming and Harrington (1994)). The martingale residuals for LZBOLL-GHN model is,

$$r_{Mi} = \begin{cases} 1 + \log \left\{ 1 - \frac{1}{\Gamma(\beta)} \gamma \left(\beta, -\log \left[1 - \frac{[u_i-1]^\alpha}{[u_i-1]^\alpha + (2-u_i)^\alpha} \right] \right) \right\} & \text{if } i \in F, \\ \log \left\{ 1 - \frac{1}{\Gamma(\beta)} \gamma \left(\beta, -\log \left[1 - \frac{[u_i-1]^\alpha}{[u_i-1]^\alpha + (2-u_i)^\alpha} \right] \right) \right\} & \text{if } i \in C, \end{cases} \tag{13}$$

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

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

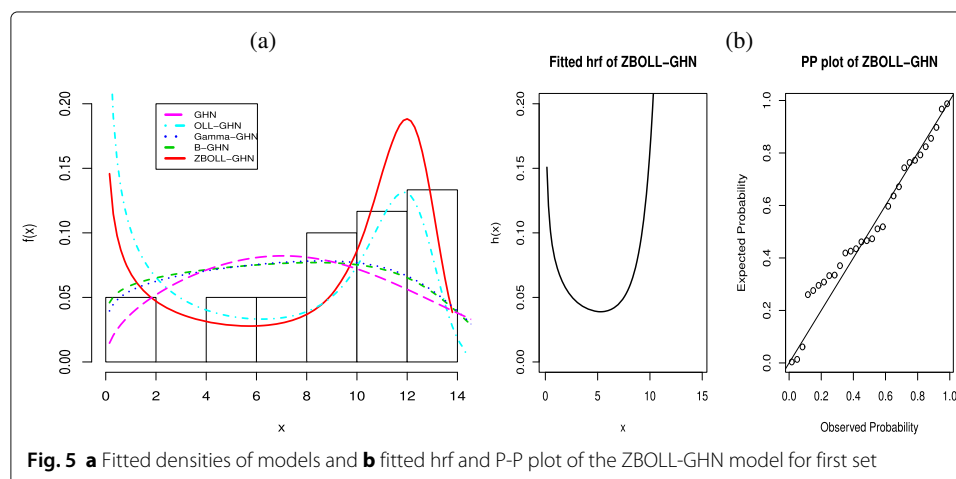


Fig. 5 **a** Fitted densities of models and **b** fitted hrf and P-P plot of the ZBOLL-GHN model for first set

Table 3 The LR test results for first data set

	Hypotheses	LR	p-value
ZBOLL-GHN versus OLL-GHN	$H_0 : \beta = 1$	4.936	0.0262
ZBOLL-GHN versus Gamma-GHN	$H_0 : \alpha = 1$	17.595	< 0.0001
ZBOLL-GHN versus GHN	$H_0 : \alpha = \beta = 1$	29.746	< 0.0001

residual was proposed by Therneau et al. (1990). The modified deviance residual for LZBOLL-GHN model is,

$$r_{D_i} = \begin{cases} \text{sign}(\hat{r}_{M_i}) \left\{ -2 \left[\left(1 + \log \left\{ 1 - \frac{1}{\Gamma(\beta)} \gamma \left(\beta, -\log \left[1 - \frac{[u_i-1]^\alpha}{[u_i-1]^\alpha + (2-u_i)^\alpha} \right] \right) \right\} \right) \right] \right\}^{\frac{1}{2}} & \text{if } i \in F, \\ \text{sign}(\hat{r}_{M_i}) \left\{ -2 \left[1 + \log \left\{ 1 - \frac{1}{\Gamma(\beta)} \gamma \left(\beta, -\log \left[1 - \frac{[u_i-1]^\alpha}{[u_i-1]^\alpha + (2-u_i)^\alpha} \right] \right) \right\} \right] \right\}^{\frac{1}{2}} & \text{if } i \in C, \end{cases} \quad (14)$$

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

Applications

In this section, four real data sets are used to compare ZBOLL-GHN model with its sub-models and beta-GHN model introduced by Pescim et al. (2013). The first three data sets are used to demonstrate the univariate data fitting performance of ZBOLL-GHN distribution. The fourth data set is used to investigate the usefulness of the proposed distribution in survival analysis. The **optim** function is used to estimate the unknown model parameters. The MLEs and corresponding standard errors, estimated $-\ell$, Kolmogorov-Smirnov (K-S) statistic and corresponding p-value, Cramér-von Mises (W^*), Anderson-Darling (A^*) statistics and Akaike Information Criteria (AIC) are reported in Tables 2, 4 and 6. The lower the values of these criteria show the better fitted model on the data sets. The

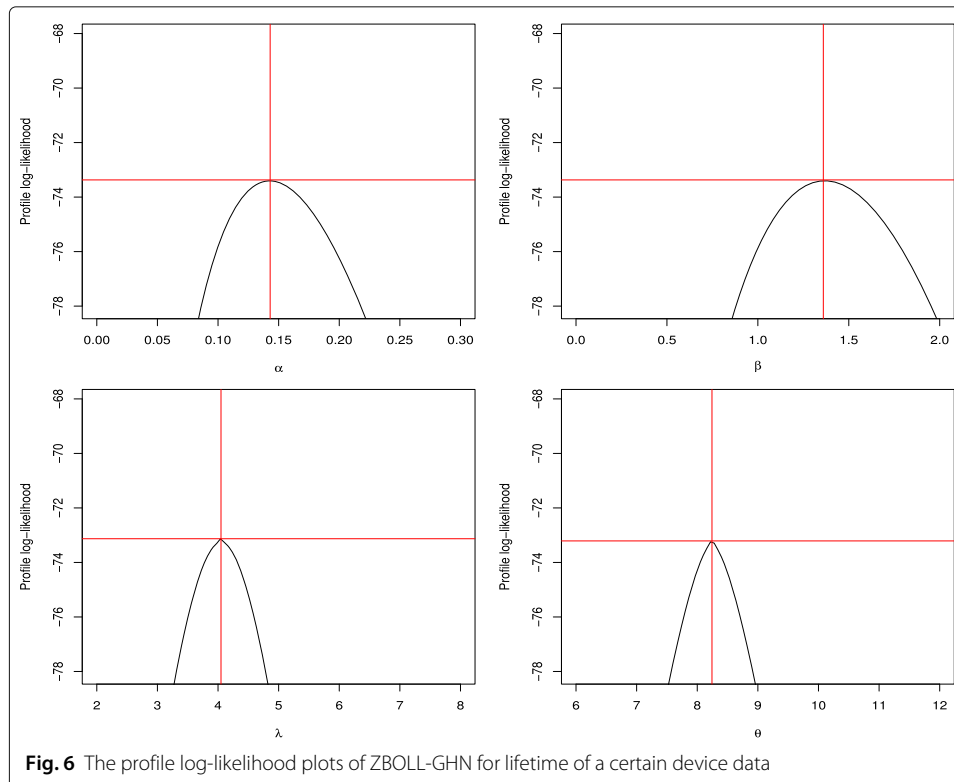


Fig. 6 The profile log-likelihood plots of ZBOLL-GHN for lifetime of a certain device data

Table 4 MLEs and their SEs of the fitted models and goodness-of-fit statistics for second data set

Models	α	β	λ	θ	$-\ell$	AIC	A*	W*	K-S	p-value
ZBOLL-GHN	0.531 0.065	0.387 0.059	7.807 0.003	4.079 0.003	126.286	260.573	0.566	0.069	0.066	0.847
B-GHN	0.240 0.029	1.704 0.375	6.370 0.057	4.527 0.056	128.668	265.336	1.005	0.164	0.091	0.476
Gamma-GHN		0.356 0.472	4.438 5.055	4.153 0.833	128.774	263.548	0.895	0.141	0.085	0.569
OLL-GHN	0.868 0.231		2.145 0.487	3.080 0.137	129.197	264.394	0.656	0.089	0.072	0.758
GHN			1.917 0.175	3.107 0.135	129.328	262.656	0.600	0.078	0.067	0.834

histograms with fitted pdfs are provided for visual comparison of the fitted distribution functions. Moreover, fitted hrfs and P-P plots of the best fitted models are displayed in Figs. 5, 7 and 9.

Lifetime of device data

The first data set is given by Sylwia (2007) on the lifetime of a certain device. Table 2 shows the estimated parameters and their standard errors, $-\ell$, A*, W*, K-S and its corresponding p-value and AIC values. Based on the figures in Table 2, it is clear that ZBOLL-GHN model provides the best fit for this data set. Figure 5a displays the estimated pdfs of the fitted models. Figure 5b displays the P-P plot of ZBOLL-GHN distribution and its fitted hrf. Figure 5 shows that ZBOLL-GHN distribution provides superior fit to the left-skewed data set.

Table 3 shows the LR statistics and the corresponding p-values for the first data set. From Table 3, the computed p-values are smaller than 0.05, so the null hypotheses are rejected for all sub-models. We conclude that the ZBOLL-GHN model fits the first data better than its sub-models according to the LR test results.

In addition, the profile log-likelihood functions of the ZBOLL-GHN distribution are plotted in Fig. 6. These plots reveal that the likelihood functions of the ZBOLL-GHN distribution have solutions that are maximizers.

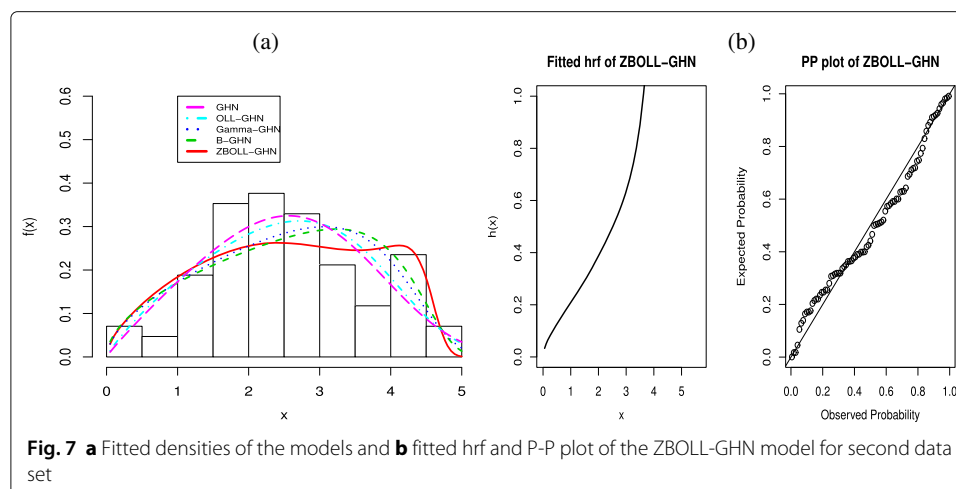


Table 5 The LR test results for second data set

	Hypotheses	LR	<i>p</i> -value
ZBOLL-GHN versus OLL-GHN	$H_0 : \beta = 1$	5.822	0.016
ZBOLL-GHN versus Gamma-GHN	$H_0 : \alpha = 1$	4.976	0.026
ZBOLL-GHN versus GHN	$H_0 : \alpha = \beta = 1$	6.084	0.047

Failure times of wind-shield data

The second data set represents the failure times for a particular wind-shield model including 85 observations that are classified as failed times of wind-shields (Murthy et al. 2004). Table 4 shows the estimated parameters and their standard errors, $-\ell$ and AIC values. Based on the figures in Table 4, ZBOLL-GHN distribution provides the best fit among others. Figure 7a displays the histogram with fitted pdfs and Fig. 7b displays the fitted hrf and P-P plot of ZBOLL-GHN distribution. These figures reveal that ZBOLL-GHN model provides superior fit to the second data set.

Table 5 shows the LR statistics and the corresponding *p*-values for the second data set. From Table 5, the computed *p*-values are smaller than 0.05, so the null hypotheses are rejected for all sub-models. We conclude that the ZBOLL-GHN model fits the first data better than its sub-models according to the LR test results.

The profile log-likelihood functions of the ZBOLL-GHN distribution are plotted but not included here. These plots reveal that the likelihood functions of the ZBOLL-GHN distribution have solutions that are maximizers.

Strengths of glass fibres data

The third data set obtained from Smith and Naylor (1987) represents the strengths of 1.5 cm glass fibres, measured at the National Physical Laboratory, England. Unfortunately, the units of measurement are not given in the paper. This data set have been analyzed recently with the beta generalized exponential distribution, which was introduced and studied by Barreto-Souza et al. (2010). Table 6 shows the estimated parameters and their standard errors, $-\ell$ and AIC values. Based on the figures in Table 6, ZBOLL-GHN distribution provides the best fit among others. Figure 8a displays the histogram with fitted pdfs and Fig. 8b displays the fitted hrf and P-P plot of ZBOLL-GHN distribution. These figures reveal that ZBOLL-GHN model provides superior fit to the third data set.

Table 7 shows the LR statistics and the corresponding *p*-values for the third data set. From Table 7, the computed *p*-values are smaller than 0.05, so the null hypotheses are

Table 6 MLEs and their SEs of the fitted models and goodness-of-fit statistics for third data set

Models	α	β	λ	θ	$-\ell$	AIC	A*	W*	K-S	<i>p</i> -value
ZBOLL-GHN	5.820	0.340	1.723	2.240	11.627	31.254	0.529	0.094	0.115	0.373
	6.976	0.134	1.902	0.611						
B-GHN	1.131	0.298	3.592	1.324	14.113	36.227	0.973	0.174	0.137	0.186
	0.445	0.362	0.529	0.256						
Gamma-GHN		1.316	3.670	1.579	14.513	35.026	1.084	0.195	0.144	0.144
		0.545	1.096	0.172						
OLL-GHN	1.290		3.761	1.709	14.163	34.328	1.065	0.192	0.136	0.188
	0.328		0.775	0.048						
GHN			4.414	1.682	14.740	33.481	1.052	0.187	0.145	0.141
			0.429	0.036						

Table 7 The LR test results for third data set

	Hypotheses	LR	<i>p</i> -value
ZBOLL-GHN versus OLL-GHN	$H_0 : \beta = 1$	5.0716	0.0243
ZBOLL-GHN versus Gamma-GHN	$H_0 : \alpha = 1$	5.7716	0.0162
ZBOLL-GHN versus GHN	$H_0 : \alpha = \beta = 1$	6.2264	0.0444

rejected for all sub-models. We conclude that the ZBOLL-GHN model fits the first data better than its sub-models according to the LR test results.

The profile log-likelihood functions of the ZBOLL-GHN distribution are plotted but not included here. These plots reveal that the likelihood functions of the ZBOLL-GHN distribution have solutions that are maximizers (Fig. 8).

Voltage data

Lawless (2003) reported an experiment in which specimens of solid epoxy electrical-insulation were studied in an accelerated voltage life test. The sample size is $n = 60$, the percentage of censored observations is 10% and there are three levels of voltage: 52.5, 55.0 and 57.5. The variables involved in the study are: x_i - failure times for epoxy insulation specimens (in min); c_i - censoring indicator (0 =censoring, 1 =lifetime observed); v_{i1} - voltage (kV).

The data set was used by Pescim et al. (2013) for illustrating the log-B-GHN (LBGHN) regression model. Pescim et al. (2013) compared the log-B-GHN (LBGHN) regression model with LOLLGHN and log-GHN (LGHN) models. In this section we compare the LZBOLL-GHN regression model with models reported in Pescim et al. (2013). The regression model fitted to the voltage data set is given by

$$y_i = \beta_0 + \beta_1 x_{i1} + \sigma z_i, \tag{15}$$

where the random variable y_i follows the LZBOLL-GHN distribution given in (8). The results are presented in Table 8. The MLEs of the model parameters and their SEs and the values of the AIC and BIC statistics are listed in Table 8.

Based on the figures in Table 8, we conclude that the fitted LZBOLL-GHN regression model has the lowest AIC and BIC values. Figure 9 provides the plots of the empirical and

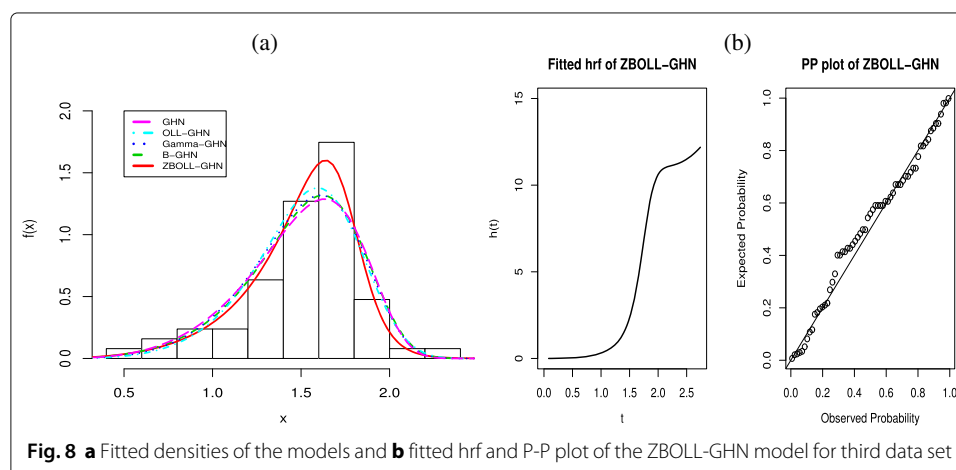


Fig. 8 a Fitted densities of the models and **b** fitted hrf and P-P plot of the ZBOLL-GHN model for third data set

Table 8 MLEs of the parameters to the voltage data for LZBOLL-GHN, LBGHN, LGHN and log-Weibull regression models, the corresponding SEs in second line, p-values in third line and the AIC and BIC statistics

Model	α	β	σ	β_0	β_1	AIC	BIC
LZBOLL-GHN	41.488	10.857	16.021	21.865	-0.177	166.264	176.735
	49.967	11.306	18.828	11.003	0.063		
				0.047	0.005		
LBGHN	102.140	1.564	5.306	10.632	-0.201	167.100	177.500
	3.989	0.672	0.666	3.304	0.056		
				0.002	0.001		
LGHN			0.778	23.637	-0.301	178.800	185.100
			0.089	2.928	0.053		
				<0.001	<0.001		
Log-Weibull			0.845	22.032	-0.275	173.400	179.700
			0.090	3.046	0.055		
				<0.001	<0.001		

estimated survival function for the LZBOLL-GHN regression model. We can conclude from these plots that LZBOLL-GHN regression model provides a good fit to the data.

Residual Analysis of LZBOLL-GHN model

Figure 10 displays the index plot of the modified deviance residuals and its Q-Q plot against $N(0, 1)$ quantiles. Based on the Figure 10, we conclude that none of the observed values appears as a possible outlier. Thus, it is clear that the fitted model is appropriate for these data set (Fig. 10).

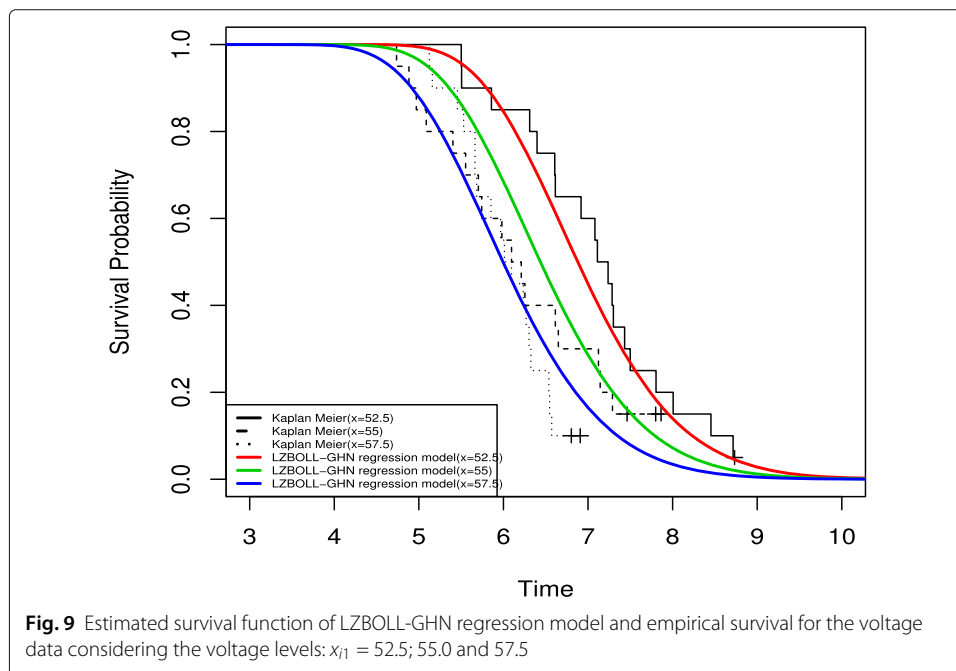
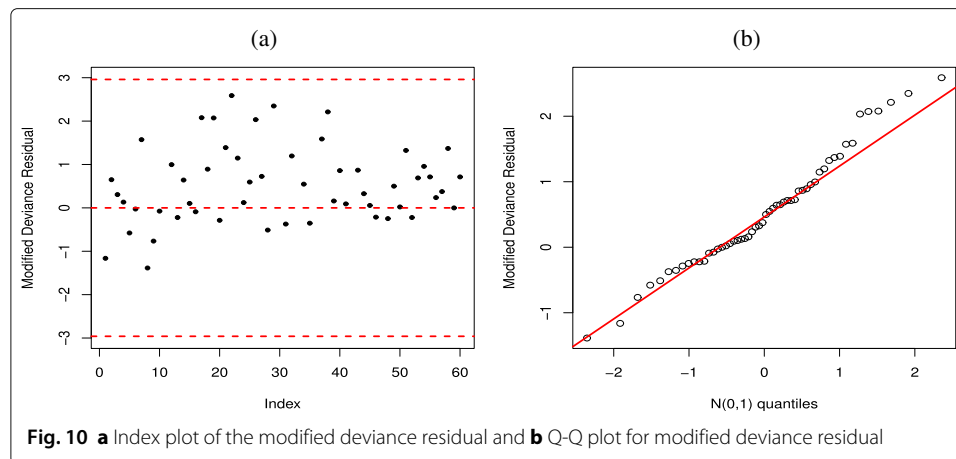


Fig. 9 Estimated survival function of LZBOLL-GHN regression model and empirical survival for the voltage data considering the voltage levels: $x_{f1} = 52.5; 55.0$ and 57.5



Summary

A new model called Zografos-Balarkishnan odd log-logistic generalized half-normal is introduced and studied. We assess the performance of the maximum likelihood estimators of the parameters of the new distribution with respect to the sample size n . The assessment is based on a graphical simulation study. The flexibility of the new model is illustrated by means of the three real data sets. The new model performs much better than beta generalized half-normal, generalized half-normal, odd log-logistic generalized half-normal and the generalized half-normal models. Additionally, a new log-location regression model based on the new distribution is introduced and studied. The martingale residual and the modified deviance residuals to detect outliers and evaluate the model assumptions are defined. We demonstrate that the new log-location regression model can be very useful in the analysis of real data and provide more realistic fits than other regression models such as the log beta generalized half-normal, the log generalized half-normal and the log-Weibull regression models. The potentiality of the new regression model is illustrated by means of a real data.

Abbreviations

ALs: Average lengths; BGHN: Beta generalized half-normal; BGHNG: Beta generalized half-normal geometric; CPs: Coverage probabilities; GHN: Generalized half-normal; HN: Half-normal; KwGHN: Kumaraswamy generalized half-normal; LZBOLLGHN: Log-Zografos-Balarkishnan odd log-logistic generalized half-normal; MLEs: Maximum likelihood estimates; MSEs: Means square errors; OLLGHN: odd log-logistic generalized half-normal; ZBOLL-G: Zografos-Balarkishnan odd log-logistic-G; ZBOLLGHN: Zografos-Balarkishnan odd log-logistic generalized half-normal

Acknowledgments

Not applicable.

Funding

GGH (co-author of the manuscript) is an Associate Editor of JSDA, 100% discount on Article Processing Charge (APC) for accepted article).

Availability of data and material

The used data sets are given in the manuscript.

Authors' contributions

EA, HMY and GGH have contributed jointly to all of the sections of the paper. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Statistics, Bartın University, Bartın 74100, Turkey. ²Department of Statistics, Mathematics and Insurance, Benha University, Benha, Egypt. ³Department of Mathematics, Statistics and Computer Science, Marquette University, Milwaukee, USA.

Received: 1 May 2018 Accepted: 4 November 2018

Published online: 05 December 2018

References

- Aarts, R.M.: Lauricella functions (2000). www.mathworld.com/LauricellaFunctions.html. From MathWorld - A Wolfram Web Resource, created by Eric W. Weisstein
- Barreto-Souza, W., Santos, A.H., Cordeiro, G.M.: The beta generalized exponential distribution. *J. Stat. Comput. Simul.* **80**, 159–172 (2010)
- Cooray, K., Ananda, M.M.A.: A generalization of the half-normal distribution with applications to lifetime data. *Commun. Stat. Theory Methods.* **37**, 1323–1337 (2008)
- Cordeiro, G.M., Alizadeh, M., Ortega, E.M., Serrano, L.H.V.: The Zografos-Balakrishnan odd log-logistic family of distributions: properties and applications. *Hacettepe Res. J. Math. Stat.* **45**, 1781–1803 (2016a)
- Cordeiro, G.M., Alizadeh, M., Pescim, R.R., Ortega, E.M.M.: The odd log-logistic generalized half-normal lifetime distribution: properties and applications. *Commun. Stat. Theory Methods.* **46**, 4195–4214 (2016b)
- Cordeiro, G.M., Pescim, R.R., Ortega, E.M.M.: The Kumaraswamy generalized half-normal distribution for skewed positive data. *J. Data Sci.* **10**, 195–224 (2012)
- Cordeiro, G.M., Pescim, R.R., Ortega, E.M.M., Demétrio, C.G.B.: The beta generalized half-normal distribution: new properties. *J. Probab. Stat.* **2013**, 1–18 (2013)
- Eugene, N., Lee, C., Famoye, F.: Beta-normal distribution and its applications. *Commun. Stat. Theory Methods.* **31**, 497–512 (2002)
- Exton, H.: Handbook of hypergeometric integrals: theory, applications, tables, computer programs. Halsted Press, New York (1978)
- Fleming, T.R., Harrington, D.P.: Counting process and survival analysis. John Wiley, New York (1994)
- Hamedani, G.G.: On certain generalized gamma convolution distributions II (No. 484). Technical Report No. 484. Marquette University, MSCS (2013)
- Lawless, J.F.: Statistical models and methods for lifetime data, Wiley Series in Probability and Statistics. Wiley, Hoboken, NJ, USA (2003). 2nd edition
- Murthy, D.P., Xie, M., Jiang, R.: Weibull models (Vol. 505). Wiley (2004)
- Pescim, R.R., Ortega, E.M., Cordeiro, G.M., Alizadeh, M.: A new log-location regression model: estimation, influence diagnostics and residual analysis. *J. Appl. Stat.* **44**, 233–252 (2017)
- Pescim, R.R., Demétrio, C.G.B., Cordeiro, G.M., Ortega, E.M.M., Urbano, M.R.: The beta generalized half-normal distribution. *Comput. Stat. Data Anal.* **54**, 945–957 (2010b)
- Pescim, R.R., Ortega, E.M.M., Cordeiro, G.M., Demétrio, C.G.B., Hamedani, G.G.: The log-beta generalized half-normal regression model. *J. Stat. Theory Appl.* **12**, 330–347 (2013)
- Ramires, T.G., Ortega, E.M.M., Cordeiro, G.M., Hamedani, G.G.: The beta generalized half-normal geometric distribution. *Stud. Sci. Math. Hung.* **50**, 523–554 (2013)
- Smith, R.L., Naylor, J.C.: A comparison of maximum likelihood and bayesian estimators for the three-parameter Weibull distribution. *Appl. Stat.* **36**, 358–369 (1987)
- Sylwia, K.B.: Makeham's generalised distribution. *Comput. Methods Sci. Tech.* **13**, 113–120 (2007)
- Therneau, T.M., Grambsch, P.M., Fleming, T.R.: Martingale-based residuals for survival models. *Biometrika.* **77**, 147–160 (1990)
- Trott, M.: The mathematica guidebook for symbolics. Springer, New York (2006)

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com