Pakistan Journal of Statistics and Operation Research

Bayesian Inference for a Weighted Bilal Distribution: Regression Model

Yupapin Atikankul[1]*

*Corresponding author

1. Department of Mathematics and Statistics, Rajamangala University of Technology Phra Nakhon, Thailand, yupapin.a@rmutp.ac.th

Abstract

Weighted distributions play an important role when observations from a sample are recorded with unequal probabilities. They are useful for the efficient modeling of statistical data when the original distributions are not appropriate. In this paper, a new weighted distribution is proposed. Various statistical properties of the proposed distribution such as survival function, hazard rate function, mean residual life function, moments, moment generating function, Bonferroni curve, Lorenz curve, and order statistic are presented. The Bayesian estimator of the distribution parameter is derived. The behavior of the Bayesian estimator is assessed by a simulation study. Furthermore, a regression model is developed based on the proposed distribution. Some real data applications are analyzed to show the potentiality of the proposed models.

Key Words: Weighted Distribution; Bilal Distribution; Bayesian Approach; Regression Model.

Mathematical Subject Classification: 60E05, 62JO5.

1. Introduction

Weighted distributions are widely used in many areas such as engineering, medicine, economics, and biological science. When observations are collected by nature according to a certain stochastic model, weighted distributions can be used for modeling data. They can extend distributions by adding flexibility and are helpful for a better understanding of the original distributions.

Suppose Y is a non-negative random variable with probability density function (pdf) $g(y)$, the pdf of a weighted distribution can be defined by

$$g_w(y) = \frac{w(y)g(y)}{\mathrm{E}(w(y))},$$

where $w(y)$ is a non-negative weighted function and $E(w(Y)) < \infty$.

Patil and Ord (1976) proposed $w(y) = y$, called length-biased distribution. The pdf of length-biased distribution can be defined by

$$g_l(y) = \frac{yg(y)}{\mathrm{E}(Y)}. \tag{1}$$

Abd-Elrahman (2013) introduced a new distribution by utilizing order statistics for lifetime data, called the Bilal distribution. The pdf of the Bilal distribution with a parameter θ is given by

$$g(y) = \frac{6}{\theta} e^{\frac{-2y}{\theta}} (1 - e^{-\frac{y}{\theta}}), \quad y > 0, \theta > 0. \tag{2}$$

The cumulative distribution function (cdf) of the Bilal distribution is

$$G(y) = 1 - e^{\frac{-2y}{\theta}} (3 - 2e^{\frac{y}{\theta}}).$$

The rth moment about the origin of the Bilal distribution can be written by

$$\mathrm{E}\left(Y^r\right) = r! \left(\frac{\theta}{6}\right)^r \left(3^{r+1} - 2^{r+1}\right).$$

If $r = 1$, the mean of the Bilal distribution is given by

$$\mathrm{E}\left(Y\right) = \frac{5\theta}{6}. \tag{3}$$

The Bilal distribution is in explicit form with a parameter θ. Several generalizations of the Bilal distribution for lifetime data are proposed such as the new two parameter generalized Bilal distribution (Abd-Elrahman, 2017), the power Bilal distribution (Riad et al., 2022).

A mixture model is a flexible and effective method for analyzing data from multiple populations (Peel and MacLahlan, 2000). In this paper, a mixture of the Bilal distribution and the length-biased Bilal distribution, called the weighted Bilal distribution, is proposed. The proposed distribution has a closed-form pdf with a parameter θ. We study shape and some properties of the proposed distribution. Moreover, a regression model based on the weighted Bilal distribution is constructed. Since the maximum likelihood estimation is not suitable for small sample sizes, in this article the Bayesian approach is applied for unknown parameters. The weighted Bilal distribution and the weighted Bilal regression model are shown the potential by fitting with two real data sets, and they are compared with other models. The results indicate that the proposed models outperform compared models.

The organization of this paper is as follows. In Section 2, we introduce the weighted Bilal distribution and its shape. Reliability measures and various properties of the proposed distribution are presented in Section 3, In Section 4, the unknown parameter of the proposed distribution is derived by the Bayesian approach. The simulation study is carried out to assess the Bayesian estimator in Section 5. In Section 6, a regression model based on the proposed distribution is presented. In section 7, applications of the weighted Bilal distribution and the weighted Bilal regression model are presented. Finally, we summarize the paper in Section 8.

2. The Weighted Bilal Distribution

In this section, the weighted Bilal distribution is defined and its shape is studied.

Theorem 2.1. *A random variable Y follows the weighted Bilal distribution with a parameter $\theta > 0$ if it has the pdf*

$$f(y) = \frac{6e^{\frac{-3y}{\theta}}(e^{\frac{y}{\theta}} - 1)(5\theta^2 + 6y)}{5\theta^2(\theta + 1)}, \quad y > 0. \tag{4}$$

Proof. The pdf of the length-biased Bilal distribution can be obtained by substituting Equation (2) and Equation (3) into Equation (1). Hence

$$g_l(y) = \frac{36ye^{-\frac{2y}{\theta}}(1 - e^{-\frac{y}{\theta}})}{5\theta^2}. \tag{5}$$

The weighted Bilal distribution can be expressed as a mixture of two component

$$f(y) = pg(y) + (1 - p)g_l(y), \tag{6}$$

where $p = \frac{\theta}{1+\theta}$, $g(y)$ is the pdf of the Bilal distribution, given by Equation (2); and $g_l(y)$ is the pdf of the length-biased Bilal distribution, given by Equation (5). Finally, Equation (4) is obtained.

□

The corresponding cumulative distribution function is

$$F(y) = 1 - \frac{e^{-\frac{3y}{\theta}} \left(\left(18y + 15\theta^2 + 9\theta\right) e^{\frac{y}{\theta}} - 12y - 10\theta^2 - 4\theta \right)}{5\theta\left(\theta + 1\right)}. \quad (7)$$

Proposition 2.1. *The pdf of the weighted Bilal distribution is log-concave.*

Proof. The first and the second derivative of logarithm of $f(y)$ are

$$\frac{d}{dy}\log f(y) = \frac{6e^{-\frac{3y}{\theta}} e^{\frac{3y}{\theta}}}{6y + 5\theta^2} - \frac{3e^{-\frac{3y}{\theta}} e^{\frac{3y}{\theta}}}{\theta} + \frac{e^{\frac{y}{\theta}}}{\theta\left(e^{\frac{y}{\theta}} - 1\right)},$$

$$\frac{d^2}{dy^2}\log f(y) = -\frac{e^{\frac{2y}{\theta}}}{\theta^2\left(e^{\frac{y}{\theta}} - 1\right)^2} + \frac{e^{\frac{y}{\theta}}}{\theta^2\left(e^{\frac{y}{\theta}} - 1\right)} - \frac{36}{\left(6y + 5\theta^2\right)^2}.$$

Since $\frac{d^2}{dy^2}\log f(y) \leq 0$, the weighted Bilal density is log-concave. □

Figure 1 shows the pdf and cdf plots of the weighted Bilal ditribution with different parameter values. The pdf is either a decreasing or a unimodal function and the cdf decreases with increasing value of θ.

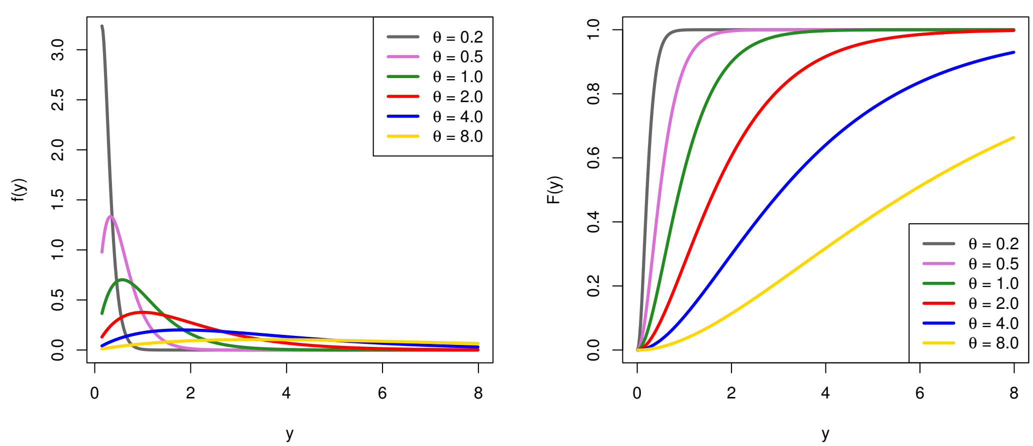

Figure 1: Pdf plot (left) and cdf plot (right) of the weighted Bilal distribution with some parameter values.

3. Reliability Measures and Statistical Properties

In this section, reliability measures and statistical properties like survival function, hazard rate function, mean residual life function, moments, moment generating function, Bonferroni curve, Lorenz curve and order statistic of the weighted Bilal distribution are presented.

3.1. Survival Function, Hazard Rate Function and Mean Residual Life Function

The survival function, hazard rate function and mean residual life function are applied frequently in survival or reliability studies. Let Y be a random variable with weighted Bilal pdf, the survival function of Y is given by

$$S(y) = \frac{e^{-\frac{3y}{\theta}}\left(\left(18y + 15\theta^2 + 9\theta\right)e^{\frac{y}{\theta}} - 12y - 10\theta^2 - 4\theta\right)}{5\theta\left(\theta + 1\right)}.$$

The hazard rate function of Y can be defined by $h(y) = \frac{f(y)}{S(y)}$. Therefore, the hazard rate function of the weighted Bilal distribution is given by

$$h(y) = \frac{6(e^{\frac{y}{\theta}} - 1)(5\theta^2 + 6y)}{\theta(e^{\frac{y}{\theta}}(15\theta^2 + 9\theta + 18y) - 10\theta^2 - 4\theta - 12y)}.$$

Proposition 3.1. *The hazard rate function of the weighted Bilal distribution is an increasing function for all θ.*

Proof. The shape of the hazard rate function is studied by using the Glaser's lemma (Glaser, 1980).

Let $\eta(y) = -\frac{d\log f(y)}{dy}$, then

$$\frac{d\eta(y)}{dy} = \frac{36\theta^2 e^{\frac{2y}{\theta}} + \left(36y^2 + 60\theta^2 y + 25\theta^4 - 72\theta^2\right)e^{\frac{y}{\theta}} + 36\theta^2}{\theta^2\left(6y + 5\theta^2\right)^2\left(e^{\frac{y}{\theta}} - 1\right)^2} > 0$$

□

Thus, $h(y)$ is increasing function.

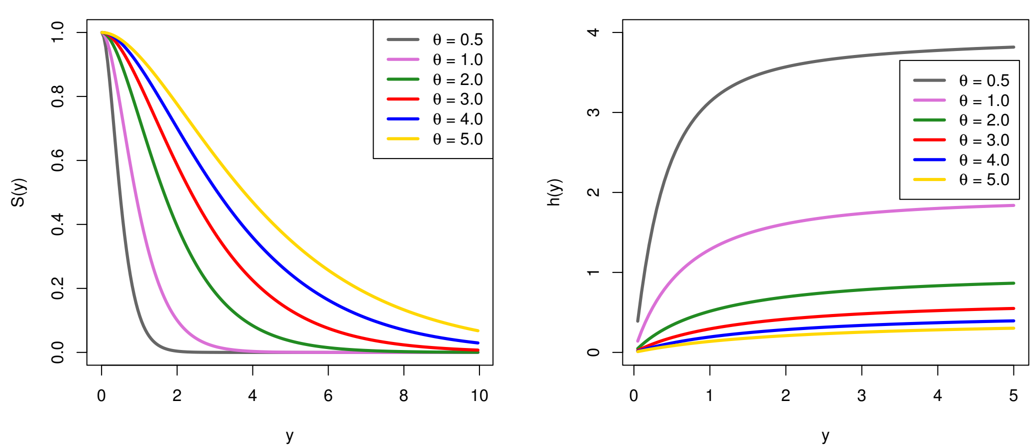

Figure 2: Survival function plot (left) and hazard rate function plot (right) of the weighted Bilal distribution with various parameter values.

Figure 2 displays plots of the survival function and the hazard rate function with some parameter values. The figure confirms that the hazard rate function of the weighted Bilal distribution is an increasing function. For fixed y, the value of $S(y)$ is increasing with increasing θ but the value of $h(y)$ is decreasing with increasing θ.

The mean residual life function can be defined by

$$\begin{aligned} m(y) &= E(Y-y|Y>y) \\ &= \frac{1}{1-F(y)}\int_y^\infty [1-F(t)]dt \\ &= \frac{\theta\left((9e^{\frac{y}{\theta}}-4)(5\theta^2+6y)+2\theta(27e^{\frac{y}{\theta}}-8)\right)}{6\left((3e^{\frac{y}{\theta}}-2)(5\theta^2+6y)+\theta(9e^{\frac{y}{\theta}}-4)\right)}. \end{aligned}$$

3.2. Moments and Moment Generating Function

Moments are important properties for any distribution. They can be used to study characteristics of distribution and for estimation.

Proposition 3.2. *Let Y be a weigthed Bilal random variable, then the rth moment about the origin of Y is given by*

$$\mu_r' = \frac{6^{-r}\theta^r\left(3^{r+1}(5\theta+3r+3)-2^{r+1}(5\theta+2r+2)\right)\Gamma(r+1)}{5\theta+1}. \tag{8}$$

Proof.

$$\begin{aligned} \mu_r' &= \mathrm{E}(Y^r) \\ &= \int_0^\infty y^r \frac{6e^{\frac{-3y}{\theta}}(e^{\frac{y}{\theta}}-1)(5\theta^2+6y)}{5\theta^2(\theta+1)}dy \\ &= \int_0^\infty \Big[-\frac{36e^{\frac{-3y}{\theta}}y^{r+1}}{5\theta^2(\theta+1)}+\frac{36e^{\frac{-2y}{\theta}}y^{r+1}}{5\theta^2(\theta+1)}-\frac{6e^{\frac{-3y}{\theta}}y^r}{\theta+1}+\frac{6e^{\frac{-2y}{\theta}}y^r}{\theta+1}\Big]dy \\ &= \frac{6^{-r}\theta^r\left(3^{r+1}(5\theta+3r+3)-2^{r+1}(5\theta+2r+2)\right)\Gamma(r+1)}{5(\theta+1)}. \end{aligned}$$

□

Proposition 3.3. *If Y be a weigthed Bilal random variable, then the moment generating function of Y can be obtained by*

$$M_Y(t) = \frac{6(5\theta^3t^2-25\theta^2t-12\theta t+30\theta+30)}{5(\theta+1)(\theta t-3)^2(\theta t-2)^2}. \tag{9}$$

Proof.

$$\begin{aligned} M_Y(t) &= \mathrm{E}(e^{tY}) \\ &= \int_0^\infty e^{ty}\frac{6e^{\frac{-3y}{\theta}}(e^{\frac{y}{\theta}}-1)(5\theta^2+6y)}{5\theta^2(\theta+1)}dy \\ &= \int_0^\infty \Big[-\frac{36ye^{ty-\frac{3y}{\theta}}}{5\theta^2(\theta+1)}+\frac{36ye^{ty-\frac{2y}{\theta}}}{5\theta^2(\theta+1)}-\frac{6e^{ty-\frac{3y}{\theta}}}{\theta+1}+\frac{6e^{ty-\frac{2y}{\theta}}}{\theta+1}\Big]dy \\ &= \frac{6(5\theta^3t^2-25\theta^2t-12\theta t+30\theta+30)}{5(\theta+1)(\theta t-3)^2(\theta t-2)^2}. \end{aligned}$$

□

If Y be a weigthed Bilal random variable, the first four moments about the origin of Y are, respectively

$$\begin{aligned}
\mu_1' &= \frac{\theta(25\theta+38)}{30(\theta+1)},\\
\mu_2' &= \frac{\theta^2(19\theta+39)}{18(\theta+1)},\\
\mu_3' &= \frac{\theta^3(325\theta+844)}{180(\theta+1)},\\
\mu_4' &= \frac{\theta^4(211\theta+665)}{54(\theta+1)}.
\end{aligned}$$

The moments about the mean of Y are given by

$$\begin{aligned}
\mu_r &= \mathrm{E}(Y-\mu)^r\\
&= \sum_{k=0}^{r}\binom{r}{k}\mu_k'(-\mu)^{r-k}.
\end{aligned}$$

The mean plot and the variance plot of the weighted Bilal distribution are shown in figure 3. We can see that the mean and variance are increasing as θ increases.

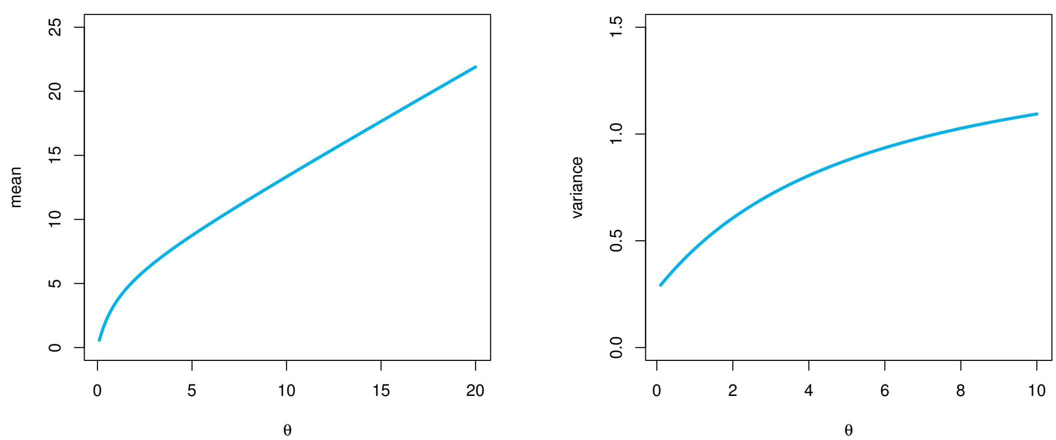

Figure 3: Mean plot (left) and variance plot (right) of the weighted Bilal distribution with some parameter values.

3.3. Bonferroni and Lorenz Curves

Bonferroni and Lorenz Curves are well-known measures of inequality of income, insurance and reliability. Let Y be a weighted Bilal random variable, the Bonferroni curve for a random variable Y can be defined by

$$\begin{aligned}
B(p) &= \frac{\int_0^q yf(y)dy}{p\int_0^\infty yf(y)dy}\\
&= \frac{1}{p\mu}\left(\int_0^\infty yf(y)dy - \int_q^\infty yf(y)dy\right)\\
&= \frac{1}{p\mu}\left(\mu - \int_q^\infty yf(y)dy\right),
\end{aligned}$$

where $q = F^{-1}(p)$.

The Lorenz curve is defined by

$$
\begin{aligned}
L(p) &= \frac{\int_0^q yf(y)dy}{\int_0^\infty yf(y)dy} \\
&= \frac{1}{\mu}\left(\int_0^\infty yf(y)dy - \int_q^\infty yf(y)dy\right) \\
&= \frac{1}{\mu}\left(\mu - \int_q^\infty yf(y)dy\right).
\end{aligned}
$$

Since

$$\int_q^\infty yf(y)dy = \frac{e^{-\frac{3q}{\theta}}\left(\left(108q^2 + \left(90\theta^2 + 108\theta\right)q + 45\theta^3 + 54\theta^2\right)e^{\frac{q}{\theta}} - 72q^2 + \left(-60\theta^2 - 48\theta\right)q - 20\theta^3 - 16\theta^2\right)}{30\theta\left(\theta + 1\right)},$$

the Bonferroni and the Lorenz curves of the weighted Bilal are, respectively

$$B(p) = \frac{1}{p}\left(1 - \frac{e^{-\frac{3q}{\theta}}\left(\left(108q^2 + \left(90\theta^2 + 108\theta\right)q + 45\theta^3 + 54\theta^2\right)e^{\frac{q}{\theta}} - 72q^2 + \left(-60\theta^2 - 48\theta\right)q - 20\theta^3 - 16\theta^2\right)}{\theta^2(25\theta + 38)}\right),$$

$$L(p) = 1 - \frac{e^{-\frac{3q}{\theta}}\left(\left(108q^2 + \left(90\theta^2 + 108\theta\right)q + 45\theta^3 + 54\theta^2\right)e^{\frac{q}{\theta}} - 72q^2 + \left(-60\theta^2 - 48\theta\right)q - 20\theta^3 - 16\theta^2\right)}{\theta^2(25\theta + 38)}.$$

3.4. Order Statistic

Let $Y_1, Y_2, \ldots, Y_n$ be independent and identically distributed random sample from the weighted Bilal distribution and let $Y_{(1)} \leq Y_{(2)} \leq \ldots \leq Y_{(n)}$ denotes the order statistics. The pdf of $Y_{(i)}$ can be defined by

$$f_{Y_{(r)}}(y) = \frac{n!}{(i-1)!(n-i)!} f_Y(y)(F_Y(y))^{i-1}(1 - F_Y(y))^{n-i},$$

where $f_Y(y)$ and $F_Y(y)$ are the pdf and cdf of the weighted Bilal distribution given by Equation (4) and Equation (7), respectively. Therefore, the *r*th order statistic of the weighted Bilal distribution is given by

$$
\begin{aligned}
f_{Y_{(r)}}(y) &= \frac{n!}{(i-1)!(n-i)!}\left(\frac{6e^{-\frac{3y}{\theta}}(e^{\frac{y}{\theta}} - 1)(5\theta^2 + 6y)}{5\theta^2(\theta + 1)}\right) \\
&\times \left(1 - \frac{e^{-\frac{3y}{\theta}}\left(\left(18y + 15\theta^2 + 9\theta\right)e^{\frac{y}{\theta}} - 12y - 10\theta^2 - 4\theta\right)}{5\theta\left(\theta + 1\right)}\right)^{i-1} \\
&\times \left(\frac{e^{-\frac{3y}{\theta}}\left(\left(18y + 15\theta^2 + 9\theta\right)e^{\frac{y}{\theta}} - 12y - 10\theta^2 - 4\theta\right)}{5\theta\left(\theta + 1\right)}\right)^{n-i}.
\end{aligned}
$$

4. Bayesian Inference

Suppose $Y_1, Y_2, \ldots, Y_n$ be an iid random variable of size n from the weighted Bilal distribution and $y_1, y_2, \ldots, y_n$ be the observations. Then the likelihood function of the observed sample is given by

$$L(\theta) = \prod_{i=1}^{n} \frac{6e^{\frac{-3y_i}{\theta}}(e^{\frac{y_i}{\theta}} - 1)(5\theta^2 + 6y_i)}{5\theta^2(\theta + 1)}.$$

Based on the Bayesian approach, a parameter is considered as random variable represented by a prior distribution. Following the recommendations of Gelman and Hill (2006), in this paper the half-Cauchy (HC) distribution with a scale 25 is used as the noninformative prior distribution. Thus,

$$\theta \sim \text{HC}(25).$$

The posterior distribution for θ is expressed by

$$p(\theta|y) = \frac{L(y|\theta)p(\theta)}{\int_\theta L(y|\theta)p(\theta)d(\theta)}.$$

In the Bayesian inference, the posterior distribution is proportional to the sum of the likelihood and the prior distribution because the denominator is normalization constant. The posterior distribution can be expressed as

$$p(\theta|y) \propto L(y|\theta)p(\theta).$$

Consequently, the posterior distribution is given by

$$p(\theta|y) \propto \prod_{i=1}^{n} \frac{6e^{\frac{-3y_i}{\theta}}(e^{\frac{y_i}{\theta}} - 1)(5\theta^2 + 6y_i)}{5\theta^2(\theta + 1)} \times \frac{2 \times 25}{\pi(\theta^2 + 25^2)}.$$

The Markov Chain Monte Carlo (MCMC) methods can be applied to obtain the posterior distribution. In this paper, we apply the Metropolis-Hastings-within Gibbs with 10,000 iterations in the LaplaceDemon function of the LaplaceDemon package (Statisticat, 2021) in the R programming language (R Core Team, 2022).

The LaplaceDemon function maximizes the logarithm of the joint posterior density, then

$$\log[p(\theta|y)] \propto \log[L(y|\theta)] + \log[p(\theta)].$$

The posterior distribution for the parameter of the weighted Bilal distribution is obtained by

$$\begin{aligned}\log[p(\theta|y)] \propto\ & \sum_{i=1}^{n} \log(6) - \frac{3}{\theta}\sum_{i=1}^{n} y_i + \sum_{i=1}^{n} \log(e^{\frac{y_i}{\theta}} - 1) + \sum_{i=1}^{n} \log(5\theta^2 + 6y_i) - \sum_{i=1}^{n} \log\left(5\theta^2(\theta + 1)\right) \\ & + \log\left(\frac{2 \times 25}{\pi(\theta^2 + 25^2)}\right).\end{aligned}$$

5. Simulation Study

The simulation study is conducted to assess the behavior of the Bayesian estimator in this section. The inversion method is used to generate random variates. The steps of random variate generation from the weighted Bilal distribution are as follows:

1. Generate $u_i, i = 1, \ldots, n$ from $U(0, 1)$.
2. Generate y_i from $F^{-1}(y_i) = u_i$ of the weigthed Bilal distribution.

The simulation study is carried out 1,000 times with $\theta = 0.5, 1, 2$ for the different sample sizes $n = 50, 100, 150, 200, 500$. The measures are based on: root mean square error (RMSE) and average bias, defined by

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{1,000} (\hat{\theta}_i - \theta)^2}{1,000}},$$

$$\text{Average bias} = \frac{\sum_{i=1}^{1,000} (\hat{\theta}_i - \theta)}{1,000}$$

Table 1: RMSE and average bias of the simulated estimate.

n	$\theta = 0.5$		$\theta = 1.0$		$\theta = 2.0$	
	RMSE	Average bias	RMSE	Average bias	RMSE	Average bias
50	0.0447	0.0086	0.0938	0.0256	0.1949	0.0436
100	0.0300	0.0016	0.0632	0.0122	0.1319	0.0186
150	0.0245	0.0012	0.0510	0.0043	0.1077	0.0114
200	0.0224	0.0026	0.0436	0.0041	0.0943	0.0119
500	0.0141	−0.0008	0.0283	0.0040	0.0574	0.0027

Table 1 displays RMSEs and average biases for the Bayesian estimator. The RMSEs decrease with increasing sample size and the average biases tend to zero for large sample size. Hence, the Bayesian estimator performs well for the parameter.

6. The Weighted Bilal Regression Model

In this section, we propose a regression model based on the weighted Bilal distribution with systematic structure. Let $\mathbf{x_i} = (x_{i1}, x_{i2}, \ldots, x_{ip})^T$ be the vector of covariates. The parameter θ_i is linked to the covariates by the log-linear structure, $\log(\theta_i) = \mathbf{x_i}^T\boldsymbol{\beta}$. Thus

$$\theta_i = e^{\mathbf{x_i}^T\boldsymbol{\beta}},$$

where $i = 1, 2, \ldots, n$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)$ be the vector of regression coefficients.

The pdf of the weighted Bilal regression model that the parameter θ depends on $\mathbf{x_i}$ can be defined by

$$f(y) = \frac{6e^{-3y/e^{\mathbf{x_i}^T\boldsymbol{\beta}}}\left(e^{y/e^{\mathbf{x_i}^T\boldsymbol{\beta}}} - 1\right)\left(5(e^{\mathbf{x_i}^T\boldsymbol{\beta}})^2 + 6y\right)}{5\left(e^{\mathbf{x_i}^T\boldsymbol{\beta}}\right)^2\left(e^{\mathbf{x_i}^T\boldsymbol{\beta}} + 1\right)}.$$

An uninformative prior distribution, normally-distributed with $\mu = 0$ and $\sigma^2 = 10000$ is assigned for each β.

$$\beta \sim N(0, 10000).$$

Here, the posterior densities of parameters are obtained by the LaplaceDemon function (Statisticat, 2021) in the R programming language (R Core Team 2022).

7. Applications

In this section, two real data sets are considered to illustrate the performance of the proposed model. The proposed model is compared with other models that have one parameter, namely the exponential model, the Lindley model (Lindley, 1958), and the Bilal model (Abd-Elrahman, 2013). Based on the Bayesian approach, the Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002) is applied to consider the best model. The lowest of the DIC value indicates the better model.

Following Spiegelhalter et al. (2002), the deviance can be given by

$$D(\Theta) = -2\log L(\Theta) + c,$$

where c is a constant that cancels out for comparing difference models.

Let $\bar{D} = E\left[D(\Theta)\right]$ is the posterior mean of the deviance and m_D is the effective number of parameters, defined by

$$m_D = \bar{D} - D(\hat{\Theta}), \tag{10}$$

where $D(\hat{\Theta})$ is the deviance evaluated at the posterior mean of $\hat{\Theta}$.

The DIC is calculated by

$$\text{DIC} = \bar{D} + m_D,$$

Rearranging Equation (10), we get $\bar{D} = m_D + D(\hat{\Theta})$. Hence, the DIC can be expressed as

$$\text{DIC} = D(\hat{\Theta}) + 2m_D.$$

7.1. Application 1

The first data set represents breaking stress of carbon fibres (in Gba), presented in the Adequacy (Nichols and Padgett, 2006). The observations are shown in Table 2.

Table 2: Breaking stress of carbon fibres.

3.7	2.74	2.73	2.5	3.6	3.11	3.27	2.87	1.47	3.11	4.42	2.41
3.19	3.22	1.69	3.28	3.09	1.87	3.15	4.9	3.75	2.43	2.95	2.97
3.39	2.96	2.53	2.67	2.93	3.22	3.39	2.81	4.20	3.33	2.55	3.31
3.31	2.85	2.56	3.56	3.15	2.35	2.55	2.59	2.38	2.81	2.77	2.17
2.83	1.92	1.41	3.68	2.97	1.36	0.98	2.76	4.91	3.68	1.84	1.59
3.19	1.57	0.81	5.56	1.73	1.59	2.00	1.22	1.12	1.71	2.17	1.17
5.08	2.48	1.18	3.51	2.17	1.69	1.25	4.38	1.84	0.39	3.68	2.48
0.85	1.61	2.79	4.7	2.03	1.8	1.57	1.8	2.03	1.61	2.12	1.89
2.88	2.82	2.05	3.65								

The total time on test (TTT) plot can be used to show the shape of the hazard rate function of the data set. Figure 4 shows the TTT plot for breaking stress of carbon fibres data. The plot reveals that the hazard rate function of the data set is an increasing shape; thus, the data set can been described by the weighted Bilal distribution.

Table 3 displays the posterior means and the DIC values for breaking stress of carbon fibres data of the exponential, Lindley, Bilal, and weighted Bilal distributions. The table indicates that the weighted Bilal distribution has the lowest

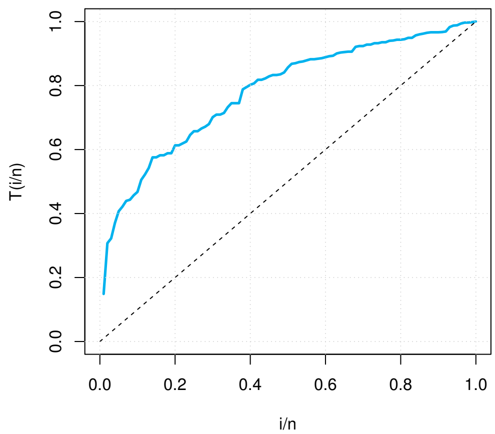

Figure 4: TTT plot for breaking stress of carbon fibres data.

DIC value; hence, it provides a better fit than other distributions.

Table 3: Posterior means and the DIC values for breaking stress of carbon fibres data.

Distribution	Posterior mean	DIC
Exponential	0.005	526.391
Lindley	0.009	417.427
Bilal	1.150	337.312
Weighted Bilal	1.012	334.242

7.2. Application 2

The second data set is considered to compare the exponential, Lindley, Bilal, and weighted Bilal regression models. The data set contains 99 observations of U.S. oil field with 3 variables (Baltagi, 2011). The variables are as follows:

- y, crude prices (USD/barrel)
- x_1, sulphur (in %)
- x_2, gravity (degree API)

Table 4 shows posterior means of the regression parameters and the DIC values of the fitted regression models. We can conclude that the weighted Bilal regression model is the best model among the competing regression models because it has the smallest DIC value.

Table 5 displays the lower bound (LB) and the upper bound (UB) of the 95% probability interval for β_j. The 95% probability intervals for β_1 of the weighted Bilal regression model is $(-0.086, 0.013)$; therefore, β_1 is not significant at the 0.05 level. A new weighted Bilal regression model is constructed by considering only x_2. The result of the new weighted Bilal regression model is shown in Table 6. The table shows that the new weighted Bilal regression model provides the lower DIC value than the weighted Bilal regression model in Table 4.

Table 4: Posterior means and the DIC values for crude prices data.

Distribution	Posterior mean			DIC
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	
Exponential	−313.416	20.314	−296.266	3036.780
Lindley	−449.122	−133.872	−225.276	2622.003
Bilal	1.257	−0.042	−0.005	673.742
Weighted Bilal	1.180	−0.035	−0.003	671.575

Table 5: LB and UB of the 95% probability intervals for regression parameters.

Distribution	(LB, UB)		
	β_0	β_1	β_2
Exponential	(−419.145, −169.516)	(−46.803, 103.688)	(−421.320, −213.902)
Lindley	(−554.150, −307.252)	(−201.686, −49.480)	(−350.330, −142.912)
Bilal	(1.168, 1.364)	(−0.070, −0.007)	(−0.009, −0.004)
Weighted Bilal	(1.036, 1.280)	(−0.086, 0.013)	(−0.004, −0.001)

Table 6: Posterior means (LB, UB) for β_0, β_2 and the DIC value.

Distribution	Posterior mean (LB, UB)		DIC
	$\hat{\beta}_0$	$\hat{\beta}_2$	
Weighted Bilal	0.820 (0.721, 0.891)	0.010 (0.008, 0.012)	670.960

8. Conclusions

In this paper, we have proposed the weighted Bilal distribution. The distribution is the mixture of the Bilal distribution and length-biased Bilal distribution. Various statistical properties of the proposed distribution (including survival function, hazard rate function, mean residual life function, moments, moment generating function, Bonferroni curve, Lorenz curve and order statistic) have been provided. The Bayesian approach has been used to estimate its parameter. The simulation study has been conducted to assess the performance of the Bayesian estimator in terms of root mean square error and bias. The Bayesian approach provides good performance for the parameter. Moreover, we have proposed a regression model based on the weighted Bilal distribution. Two real data sets have been analyzed to show the usefulness of the weighted Bilal distribution and the weighted Bilal regression model. They provide better fits than other competitive models.

Acknowledgements

The author is grateful to the Editor and three anonymous reviewers for their valuable comments and suggestion. The author would like to thank Rajamangala University of Technology Phra Nakhon.

References

1. Abd-Elrahman, A. M. (2013). Utilizing ordered statistics in lifetime distributions production: a new lifetime distribution and applications. *Journal of Probability and Statistical Science*, 11(2):153–164.

2. Abd-Elrahman, A. M. (2017). A new two-parameter lifetime distribution with decreasing, increasing or upside-down bathtub-shaped failure rate. *Communications in Statistics-Theory and Methods*, 46(18):8865–8880.
3. Baltagi, B. H. (2011). *Econometrics*. Springer-Verlag, Berlin, Heidelberg, Germany.
4. Gelman, A. and Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, New York, USA.
5. Glaser, R. E. (1980). Bathtub and related failure rate characterizations. *Journal of the American Statistical Association*, 75(371):667–672, doi:10.1080/01621459.1980.10477530.
6. Lindley, D. V. (1958). Fiducial distributions and Bayes' theorem. *Journal of the Royal Statistical Society. Series B (Methodological)*, 20(1):102–107, doi:10.1111/j.2517–6161.1958.tb00278.x.
7. Nichols, M. D. and Padgett, W. (2006). A bootstrap control chart for Weibull percentiles. *Quality and Reliability Engineering International*, 22(2):141–151, doi:10.1002/qre.691.
8. Patil, G. P. and Ord, J. (1976). On size-biased sampling and related form-invariant weighted distributions. *Sankhyā: The Indian Journal of Statistics, Series B*, 38(1):48–61.
9. Peel, D. and MacLahlan, G. (2000). *Finite Mixture Models*. John & Sons, New York, USA.
10. R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
11. Riad, F. H., Alruwaili, B., Gemeay, A. M., and Hussam, E. (2022). Statistical modeling for COVID-19 virus spread in Kingdom of Saudi Arabia and Netherlands. *Alexandria Engineering Journal*, 61(12):9849–9866.
12. Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639, doi:10.1111/1467–9868.00353.
13. Statisticat, L. (2021). *LaplacesDemon: Complete Environment for Bayesian Inference*. R package version 16.1.6.