Thailand Statistician
October 2022; 20(4): 791-811
http://statassoc.or.th
Contributed paper

Bayesian Inference for the Discrete Weibull Regression Model with Type-I Right Censored Data

Dusit Chaiprasithikul and Monthira Duangsaphon*
Department of Mathematics and Statistics, Faculty of Science and Technology,
Thammasat University, Pathumthani, Thailand.
*Corresponding author; e-mail: monthira@mathstat.sci.tu.ac.th

Received: 13 July 2021
Revised: 26 November 2021
Accepted: 2 December 2021

Abstract

This study purposes the use of Bayesian estimation for the discrete Weibull regression under type-I right censored data. Moreover, we compared the performance of the maximum likelihood estimation and the Bayesian estimation with uniform noninformative priors and informative priors using the random walk Metropolis algorithm. A simulation study was conducted to compare the performance of three different estimation methods using mean square error with three types of data: excessive zeros data, under-dispersion data, and over-dispersion data. A real dataset is analyzed to see how the model works in practice. The results from both the simulation study and a real data application showed that the maximum likelihood estimation and the Bayesian estimation with informative priors are both appropriate for the discrete Weibull regression under type-I right censored data in the cases of excessive zeros and under-dispersion. However, the Bayesian estimation with informative priors is more appropriate for the discrete Weibull regression under type-I right censored data than other methods in the case of over-dispersion.

Keywords: Bayesian estimation, random walk Metropolis algorithm, discrete Weibull regression, type-I right censored, over-dispersion.

1. Introduction

Count data refers to the number of times an event or an item occurs over a fixed period of time, which can take only the non-negative integer values. Examples include the number of times cardiac arrest happens over a fixed period of time, the number of times patients visit a doctor over a fixed period of time within a hospital, the number of epileptic seizures experienced over a fixed period of time, the number of claims in an insurance company over a fixed period of time, and the number of recurrent circuit breaker failures over a fixed period of time. This form of counts information is applied to many research areas such as medicine, actuarial science, biostatistics, demography, economics, engineering, political science, and sociology. Individual count data is called a count variable, which is treated as a random variable: the Poisson, negative binomial, and discrete Weibull distributions are widely used to represent its distribution.

Regression analysis for count data is used in realistic contexts when other variables have an effect on the count response variable. Modeling count data may present three types of dispersion: equi-dispersion, under-dispersion, and over-dispersion. The Poisson regression is most commonly used for modeling count data, e.g., Lovett and Flowerdew (1989) and Hutchinson and Holtman (2005). Despite the popularity of the Poisson regression, it is limited by its equi-dispersion, i.e., the assumption that the mean and variance are equal. Alternatively, the negative binomial regression has become the most widely used for modeling count data, e.g., Gardner et al. (1995), and Allison and Waterman (2002), as this regression is appropriate for modeling over-dispersion count data. With real data, the dispersion is mostly under-dispersed or over-dispersed, and the discrete Weibull regression can be adapted to both situations. Moreover, many datasets have multiple zeros response variables.

The discrete Weibull regression is an interesting subject for study and development, e.g., Kalktawi (2017), Kalktawi et al. (2018), Haselimashhadi et al. (2018), and Collins et al. (2020). Focusing on over-dispersion data wherein the variance is larger than the mean may limit the effectiveness of a standard model. When the value of an observation or measurement is only partly known we refer to this observation as being censored. In some cases, the response variable takes large values or outliers affecting its mean and variance, causing over-dispersion that potentially having a negative effect on the performance of a regression model. Thus, censoring the large values of this response can control this over-dispersion. Censored count data can appear in many applications where recording the count response variable is available for a limited range while the covariate values are always observed. Therefore, appropriate censoring is applied to solve problem with over-dispersion data. There are three types of right censoring: type-I right censoring that fixes a predetermined censoring value; type-II right censoring that fixes a predetermined number of uncensored data; and random censoring, extended from the type-I right censoring, in which the predetermined censoring value is random. The challenge faced by practitioners is the selection of censored data, e.g., Saffari et al. (2013), Kalktawi (2017), Yu (2018), and Saffari and Allen (2019).

Improvement can be made in the estimation of parameters for the discrete Weibull regression. Recently, Kalktawi (2017) introduced the maximum likelihood estimation for the discrete Weibull regression model with type-I right censoring, which is the classical inference that uses only empirical knowledge from the likelihood function. In the cases of over-dispersion, the performance of a censoring model is better than a standard model. Besides the classical inference, the Bayesian statistical inference differs from the maximum likelihood estimation that uses two sources of information, i.e., prior knowledge about the parameters from the prior probability distribution and empirical knowledge from the likelihood function. Moreover, Haselimashhadi et al. (2018) proposed the Bayesian estimation for the discrete Weibull regression model with an application to health data. When a prior distribution has no population basis – in other words, there is no specific prior knowledge about the parameters – the noninformative prior is used; for instance, Haselimashhadi et al. (2018) chose uniform noninformative priors on parameters. On the other hand, when there is prior knowledge of the parameters, the informative prior is provided in the Bayesian estimation.

In this paper, we focus on the Bayesian estimation for the discrete Weibull regression under type-I right censored data with uniform noninformative priors and informative priors. The main difficulty faced when dealing with the Bayesian estimation comes from the computation of the posterior probability distribution, in which case the Markov chain Monte Carlo (MCMC) method is used to draw a sample from a probability distribution. We choose one MCMC method, that is, the random walk Metropolis algorithm, for simulating a sample from a posterior probability distribution.

The remainder of this paper is organized as follows. In Section 2, we introduce the discrete Weibull distribution and the discrete Weibull regression, present the discrete Weibull regression model under type-I right censored data and the Bayesian estimation for the discrete Weibull regression model under type-I right censored data, and define the random walk Metropolis algorithm. In Section 3, we investigate the performance of the estimations through a simulation study and apply our computational methods to a real dataset. Finally, we conclude our findings in Section 4.

2. Materials and Methods

2.1. Discrete Weibull distribution

The discrete Weibull distribution was proposed by Nakagawa and Osaki (1975). They considered failure studies in which the time to failure is often measured in the number of cycles to failure and becomes a discrete random variable. In failure analysis, the failure data in failure studies are generally measured in discrete time such as cycles, blows, shocks, or revolutions. Moreover, the discrete Weibull distribution is useful to reliability engineers and theoreticians.

Let Y be a discrete random variable which follows the discrete Weibull distribution with the parameters q and β, denoted by $Y \sim DW(q,\beta)$. The cumulative distribution function and the probability mass function of a discrete random variable Y are given by

$$F_Y(y;q,\beta)=\begin{cases}1-q^{(y+1)^\beta}, & y=0,1,2,\dots\\ 0, & \text{otherwise,}\end{cases}$$

and

$$p_Y(y;q,\beta)=\begin{cases}q^{y^\beta}-q^{(y+1)^\beta}, & y=0,1,2,\dots\\ 0, & \text{otherwise,}\end{cases}$$

respectively, where $0<q<1$ and $\beta>0$ are the shape parameters (Nakagawa and Osaki, 1975). In addition, the parameter $q=1-p_Y(0;q,\beta)$ which is the probability of Y being more than zero.

The relationship to the continuous Weibull distribution and discrete Weibull distribution is shown through the shape parameter. The cumulative distribution function of a continuous random variable Y_C is given by

$$F_{Y_C}(y_C;\lambda,\beta)=\begin{cases}1-e^{-\lambda y_C^\beta}, & y_C\geq 0\\ 0, & \text{otherwise,}\end{cases}$$

where λ and β are the scale and shape parameter, respectively. The parameter β from the discrete Weibull distribution is equivalent to the shape parameter β from the continuous Weibull distribution. Moreover, the parameter q from the discrete Weibull distribution is equivalent to $e^{-\lambda}$ when λ is the scale parameter from the continuous Weibull distribution.

2.2. Discrete Weibull regression

Regression analysis for count data is a statistical technique to evaluate the relationship between a dependent variable that is a count variable and one or more explanatory variables; accordingly, it useful in real life when other variables have an effect on response variables. The discrete Weibull regression can link the independent variables through the shape parameters q and β. In this paper, we linked the independent variables only through the shape parameter q.

Lee and Wang (2003) assumed that the scale parameter λ in the continuous Weibull distribution is related to k covariates via the log link function; $\log(\lambda)=\mathbf{x\alpha}$. Similar to the continuous Weibull distribution, Kalktawi (2017) assumed that the parameter q in the discrete Weibull distribution is related to k covariates via the log-log link function; $\log(-\log(q))=\mathbf{x\alpha}$.

Let Y be a count response variable which takes only the non-negative integer values and let $x_1, x_2, \ldots, x_k$ be k explanatory variables. Assume that the conditional distribution of Y given $x_1, x_2, \ldots, x_k$ follows the discrete Weibull distribution with the parameters q and β, where the parameter q is related to k explanatory variables $x_1, x_2, \ldots, x_k$ via the log-log link function:

$$\log(-\log(q))=\mathbf{x\alpha}, \tag{1}$$

where $\mathbf{x}=\begin{pmatrix}1 & x_1 & \cdots & x_k\end{pmatrix}$ and $\boldsymbol{\alpha}=\begin{pmatrix}\alpha_0 & \alpha_1 & \cdots & \alpha_k\end{pmatrix}'$,

so
$$\log(-\log(q))=\alpha_0+\alpha_1 x_1+\ldots+\alpha_k x_k,$$

$$q\equiv q(\mathbf{x})=e^{-e^{\mathbf{x\alpha}}}. \tag{2}$$

The conditional probability mass function of Y given $x_1, x_2, \ldots, x_k$ can be written as

$$p_{Y|X_1,\ldots,X_k}(y|x_1,\ldots,x_k)=\begin{cases}\left(e^{-e^{\mathbf{x\alpha}}}\right)^{y^\beta}-\left(e^{-e^{\mathbf{x\alpha}}}\right)^{(y+1)^\beta}, & y=0,1,2,\ldots\\ 0, & \text{otherwise,}\end{cases} \tag{3}$$

where $\mathbf{x}=\begin{pmatrix}1 & x_1 & \cdots & x_k\end{pmatrix}$ and $\boldsymbol{\alpha}=\begin{pmatrix}\alpha_0 & \alpha_1 & \cdots & \alpha_k\end{pmatrix}'$.

Given n independent observations $y_1, y_2, \ldots, y_n$ and $x_{i1}, x_{i2}, \ldots, x_{ik}$, $i=1,2,\ldots,n$, from (3) for the count response variable $Y_1, Y_2, \ldots, Y_n$ and k explanatory variables $x_{i1}, x_{i2}, \ldots, x_{ik}$, the likelihood function of the discrete Weibull regression model is given by

$$L_{DW}=f(y_1,\ldots,y_n|\mathbf{x}_1,\ldots,\mathbf{x}_n)=\prod_{i=1}^{n}\left[\left(e^{-e^{\mathbf{x}_i\boldsymbol{\alpha}}}\right)^{y_i^\beta}-\left(e^{-e^{\mathbf{x}_i\boldsymbol{\alpha}}}\right)^{(y_i+1)^\beta}\right],$$

where $\mathbf{x}_i=\begin{pmatrix}1 & x_{i1} & \cdots & x_{ik}\end{pmatrix}$ and $\boldsymbol{\alpha}=\begin{pmatrix}\alpha_0 & \alpha_1 & \cdots & \alpha_k\end{pmatrix}'$.

The log-likelihood function of the discrete Weibull regression model is given by

$$l_{DW}=\sum_{i=1}^{n}\log\left[\left(e^{-e^{\mathbf{x}_i\boldsymbol{\alpha}}}\right)^{y_i^\beta}-\left(e^{-e^{\mathbf{x}_i\boldsymbol{\alpha}}}\right)^{(y_i+1)^\beta}\right].$$

2.3. Discrete Weibull regression under type-I right censored data

Right censoring is used when data is right skewed or has an outlier. In this paper, we choose type-I right censoring to solve the problem of over-dispersion data. Let $Y_1^*, Y_2^*, \ldots, Y_n^*$ be the count response variables which take only the non-negative integer values and let $x_{i1}, x_{i2}, \ldots, x_{ik}$, $i=1,2,\ldots,n$, be k explanatory variables. Assume that the conditional distribution of Y_i^* given $x_{i1}, x_{i2}, \ldots, x_{ik}$, $i=1,2,\ldots,n$, follows the discrete Weibull distribution with the parameters q_i and β, where the parameter q_i is related to k explanatory variables $x_{i1}, x_{i2}, \ldots, x_{ik}$, $i = 1, 2, \ldots, n$, via the log-log link function in (1). Given n independent observations $y_1^*, y_2^*, \ldots, y_n^*$ and $\mathbf{x}_i = \begin{pmatrix} 1 & x_{i1} & \cdots & x_{ik} \end{pmatrix}$, $i = 1, 2, \ldots, n$, from (3) censored from the right at a fixed censoring value C, the observed response variables $y_1, y_2, \ldots, y_n$ can be determined as

$$y_i = \min\left(y_i^*, C\right) = \begin{cases} y_i^*, & y_i^* < C, \\ C, & y_i^* \geq C. \end{cases} \tag{4}$$

Let δ_i, $i = 1, 2, \ldots, n$, be the censor indicator for type-I right censoring that can be specified as

$$\delta_i = I\left(y_i^* \geq C\right) = \begin{cases} 0, & y_i^* < C, \\ 1, & y_i^* \geq C. \end{cases} \tag{5}$$

The likelihood function of the discrete Weibull regression under type-I right censored data model is given by

$$L_{CDW} = f\left(y_1, \ldots, y_n \middle| \mathbf{x}_1, \ldots, \mathbf{x}_n\right) = \prod_{i=1}^{n} \left[\left(e^{-e^{\mathbf{x}_i \boldsymbol{\alpha}}}\right)^{y_i^\beta} - \left(e^{-e^{\mathbf{x}_i \boldsymbol{\alpha}}}\right)^{(y_i+1)^\beta} \right]^{1-\delta_i} \left[\left(e^{-e^{\mathbf{x}_i \boldsymbol{\alpha}}}\right)^{C^\beta} \right]^{\delta_i}.$$

The log-likelihood function of the discrete Weibull regression under type-I right censored data model is given by

$$\begin{aligned} l_{CDW} &= \sum_{i=1}^{n} (1-\delta_i) \log\left[\left(e^{-y_i^\beta e^{\mathbf{x}_i \boldsymbol{\alpha}}}\right) - \left(e^{-(y_i+1)^\beta e^{\mathbf{x}_i \boldsymbol{\alpha}}}\right) \right] - C^\beta \sum_{i=1}^{n} \delta_i e^{\mathbf{x}_i \boldsymbol{\alpha}}, \\ &= \sum_{i=1}^{n} (1-\delta_i) \log\left[w_i(\boldsymbol{\alpha}, \beta) \right] - C^\beta \sum_{i=1}^{n} \delta_i e^{\mathbf{x}_i \boldsymbol{\alpha}}, \end{aligned} \tag{6}$$

where $w_i(\boldsymbol{\alpha}, \beta) = \left(e^{-y_i^\beta e^{\mathbf{x}_i \boldsymbol{\alpha}}}\right) - \left(e^{-(y_i+1)^\beta e^{\mathbf{x}_i \boldsymbol{\alpha}}}\right)$.

In next section, we apply the inverse of the observed Fisher's information matrix to generate a random error vector in a random walk Metropolis algorithm. Let $I(\boldsymbol{\theta})$ be the observed Fisher's information matrix for the $(k+2) \times (k+2)$ unknown parameters with negative members of the second derivative of the log-likelihood function with respect to $\alpha_j, j = 0, 1, \ldots, k$ and β has the form:

$$I(\boldsymbol{\theta}) = \begin{bmatrix} -\frac{\partial^2}{\partial \alpha_0^2} l_{CDW} & -\frac{\partial^2}{\partial \alpha_0 \partial \alpha_1} l_{CDW} & \cdots & -\frac{\partial^2}{\partial \alpha_0 \partial \beta} l_{CDW} \\ -\frac{\partial^2}{\partial \alpha_0 \partial \alpha_1} l_{CDW} & -\frac{\partial^2}{\partial \alpha_1^2} l_{CDW} & \cdots & -\frac{\partial^2}{\partial \alpha_1 \partial \beta} l_{CDW} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{\partial^2}{\partial \alpha_0 \partial \beta} l_{CDW} & -\frac{\partial^2}{\partial \alpha_1 \partial \beta} l_{CDW} & \cdots & -\frac{\partial^2}{\partial \beta^2} l_{CDW} \end{bmatrix}, \tag{7}$$

where

$$\frac{\partial^2}{\partial \alpha_j \partial \alpha_{j'}} l_{CDW} = \sum_{i=1}^{n} \frac{1-\delta_i}{w_i} \left[\frac{\partial^2 w_i}{\partial \alpha_j \partial \alpha_{j'}} - \frac{1}{w_i} \frac{\partial w_i}{\partial \alpha_j} \frac{\partial w_i}{\partial \alpha_{j'}} \right] - C^\beta \sum_{i=1}^{n} \delta_i x_{ij} x_{ij'} e^{\mathbf{x}_i \boldsymbol{\alpha}},$$

$$\frac{\partial^2}{\partial\alpha_j\partial\beta}l_{CDW}=\sum_{i=1}^{n}\frac{1-\delta_i}{w_i}\left[\frac{\partial^2 w_i}{\partial\alpha_j\partial\beta}-\frac{1}{w_i}\frac{\partial w_i}{\partial\alpha_j}\frac{\partial w_i}{\partial\beta}\right]-C^{\beta}\log(C)\sum_{i=1}^{n}\delta_i x_{ij}e^{\mathbf{x}_i\boldsymbol{\alpha}},$$

$$\frac{\partial^2}{\partial\beta^2}l_{CDW}=\sum_{i=1}^{n}\frac{1-\delta_i}{w_i}\left[\frac{\partial^2 w_i}{\partial\beta^2}-\frac{1}{w_i}\left(\frac{\partial w_i}{\partial\beta}\right)^2\right]-C^{\beta}\left[\log(C)\right]^2\sum_{i=1}^{n}\delta_i e^{\mathbf{x}_i\boldsymbol{\alpha}},$$

$$\frac{\partial w_i}{\partial\alpha_j}=-x_{ij}e^{\mathbf{x}_i\boldsymbol{\alpha}}\left[e^{-y_i^{\beta}e^{\mathbf{x}_i\boldsymbol{\alpha}}}y_i^{\beta}-e^{-(y_i+1)^{\beta}e^{\mathbf{x}_i\boldsymbol{\alpha}}}(y_i+1)^{\beta}\right],$$

$$\frac{\partial w_i}{\partial\beta}=-e^{\mathbf{x}_i\boldsymbol{\alpha}}\left[e^{-y_i^{\beta}e^{\mathbf{x}_i\boldsymbol{\alpha}}}y_i^{\beta}\log(y_i)-e^{-(y_i+1)^{\beta}e^{\mathbf{x}_i\boldsymbol{\alpha}}}(y_i+1)^{\beta}\log(y_i+1)\right],$$

$$\frac{\partial^2 w_i}{\partial\alpha_j\partial\alpha_{j'}}=-x_{ij}x_{ij'}e^{\mathbf{x}_i\boldsymbol{\alpha}}\begin{bmatrix}e^{-y_i^{\beta}e^{\mathbf{x}_i\boldsymbol{\alpha}}}y_i^{\beta}\left\{1-y_i^{\beta}e^{\mathbf{x}_i\boldsymbol{\alpha}}\right\}\\-e^{-(y_i+1)^{\beta}e^{\mathbf{x}_i\boldsymbol{\alpha}}}(y_i+1)^{\beta}\left\{1-(y_i+1)^{\beta}e^{\mathbf{x}_i\boldsymbol{\alpha}}\right\}\end{bmatrix},$$

$$\frac{\partial^2 w_i}{\partial\alpha_j\partial\beta}=-x_{ij}e^{\mathbf{x}_i\boldsymbol{\alpha}}\begin{bmatrix}e^{-y_i^{\beta}e^{\mathbf{x}_i\boldsymbol{\alpha}}}y_i^{\beta}\log(y_i)\left\{1-y_i^{\beta}e^{\mathbf{x}_i\boldsymbol{\alpha}}\right\}\\-e^{-(y_i+1)^{\beta}e^{\mathbf{x}_i\boldsymbol{\alpha}}}(y_i+1)^{\beta}\log(y_i+1)\left\{1-(y_i+1)^{\beta}e^{\mathbf{x}_i\boldsymbol{\alpha}}\right\}\end{bmatrix},$$

and

$$\frac{\partial^2 w_i}{\partial\beta^2}=-e^{\mathbf{x}_i\boldsymbol{\alpha}}\begin{bmatrix}e^{-y_i^{\beta}e^{\mathbf{x}_i\boldsymbol{\alpha}}}y_i^{\beta}\left(\log(y_i)\right)^2\left\{1-y_i^{\beta}e^{\mathbf{x}_i\boldsymbol{\alpha}}\right\}\\-e^{-(y_i+1)^{\beta}e^{\mathbf{x}_i\boldsymbol{\alpha}}}(y_i+1)^{\beta}\left(\log(y_i+1)\right)^2\left\{1-(y_i+1)^{\beta}e^{\mathbf{x}_i\boldsymbol{\alpha}}\right\}\end{bmatrix}.$$

2.4. Bayesian estimation

In this section, we present the Bayesian inference for the discrete Weibull regression model under type-I right censored data. We focus on the three covariates, so the parameters $\boldsymbol{\alpha}$ and β are considered. We investigate the performance of the estimation through both noninformative and informative prior distributions. Firstly, we perform the estimators using the uniform noninformative prior distribution proposed by Haselimashhadi et al. (2018). In the context of discrete Weibull regression, there are no conjugate priors. It is often more natural to express prior information directly in terms of the parameters $\boldsymbol{\alpha}$. According to α_j is a regression coefficient that can be a real number; $\alpha_j \in \mathbb{R}, j=0,1,2,3$ and the possible values of a normal distribution is a real number. Then, we select the prior distribution of α_j as a normal distribution (Gelman et al. 2008, Fu 2016, Chanialidis et. al., 2018). Moreover, a normal prior distribution on α is particularly convenient with the computational methods. According to the parameter β from the discrete Weibull distribution is equivalent to the shape parameter β from the continuous Weibull distribution that $\beta>0$ and the possible values of a Gamma distribution is a positive real number. Then, we select the prior distribution of β as a Gamma distribution (Aslam et al., 2014; Chacko and Mohan 2019). The joint prior distribution of the parameters $\boldsymbol{\alpha}$ and β under the independence assumption is

$$\pi(\boldsymbol{\theta})=\pi(\boldsymbol{\alpha},\beta)=\pi(\alpha_0)\pi(\alpha_1)\pi(\alpha_2)\pi(\alpha_3)\pi(\beta).$$

According to the informative prior distributions of the parameters $\boldsymbol{\alpha}$ and β, we assume the hyperparameters of α_i are (μ_i,σ_i^2), $i=0,1,2,3$ and the hyperparameters of β are (a,b).

The choice of the hyperparameters' values will generally depend upon the applications of real data. At this moment, we leave them unspecified.

The joint posterior density function of the parameters $\boldsymbol{\alpha}$ and β can be written as

$$p\left(\boldsymbol{\theta}\middle|\mathbf{y},\mathbf{x},\boldsymbol{\delta}\right)=\frac{L\left(\boldsymbol{\theta}\middle|\mathbf{y},\mathbf{x},\boldsymbol{\delta}\right)\pi(\boldsymbol{\theta})}{\int_0^\infty\int_{-\infty}^\infty\cdots\int_{-\infty}^\infty L\left(\boldsymbol{\theta}\middle|\mathbf{y},\mathbf{x},\boldsymbol{\delta}\right)\pi(\boldsymbol{\theta})d\alpha_0\cdots d\alpha_3 d\beta},$$

$$\propto L\left(\boldsymbol{\theta}\middle|\mathbf{y},\mathbf{x},\boldsymbol{\delta}\right)\pi(\boldsymbol{\theta}). \tag{8}$$

The Bayes estimator of function $h(\boldsymbol{\theta})$ of the parameters $\boldsymbol{\alpha}$ and β under squared error loss function is the expected value of function $h(\boldsymbol{\theta})$ under the joint posterior density function. Therefore, the Bayes estimator of function $h(\boldsymbol{\theta})$ is given by

$$\hat{h}(\boldsymbol{\theta})=\int_0^\infty\int_{-\infty}^\infty\cdots\int_{-\infty}^\infty h(\boldsymbol{\theta})p\left(\boldsymbol{\theta}\middle|\mathbf{y},\mathbf{x},\boldsymbol{\delta}\right)d\alpha_0\cdots d\alpha_3 d\beta. \tag{9}$$

Since the integral in (9) does not have a closed form, we choose the Metropolis-Hastings algorithm to estimate the Bayes estimators.

The Metropolis-Hastings algorithm is a Markov chain Monte Carlo (MCMC) method for simulating a sample from a probability distribution that is the target distribution from which direct sampling is difficult (Hastings 1970). This algorithm is similar to acceptance-rejection method the proposal (candidate) value can be generated from the proposal distribution. Then, the proposal value is accepted with an acceptance probability. Moreover, the Metropolis-Hastings algorithm is converging to the target distribution itself. In this paper, we choose a random walk Metropolis algorithm, which is a special case of a Metropolis-Hastings algorithm.

Let the joint posterior density function of the parameters $\boldsymbol{\alpha}$ and β, $p\left(\boldsymbol{\theta}\middle|\mathbf{y},\mathbf{x},\boldsymbol{\delta}\right)$, in (8) be the target distribution, $\boldsymbol{\theta}$ be the current state value and $\boldsymbol{\theta}^*$ be the proposal value generated from the proposal distribution $q\left(\boldsymbol{\theta}^*\middle|\boldsymbol{\theta}\right)$. Then, the proposal value $\boldsymbol{\theta}^*$ is accepted with the probability $p=\min\left(1,R_{\boldsymbol{\theta}}\right)$, where

$$R_{\boldsymbol{\theta}}=\frac{L\left(\boldsymbol{\theta}^*\middle|\mathbf{y},\mathbf{x},\boldsymbol{\delta}\right)\pi(\boldsymbol{\theta}^*)}{L\left(\boldsymbol{\theta}\middle|\mathbf{y},\mathbf{x},\boldsymbol{\delta}\right)\pi(\boldsymbol{\theta})}\times\frac{q\left(\boldsymbol{\theta}\middle|\boldsymbol{\theta}^*\right)}{q\left(\boldsymbol{\theta}^*\middle|\boldsymbol{\theta}\right)}.$$

In the random walk Metropolis algorithm, the proposal distribution is symmetrical, depending only on the distance between the current state value and the proposal value. Then, the proposal value $\boldsymbol{\theta}^*$ is accepted with probability $p=\min\left(1,R_{\boldsymbol{\theta}}\right)$, where

$$R_{\boldsymbol{\theta}}=L\left(\boldsymbol{\theta}^*\middle|\mathbf{y},\mathbf{x},\boldsymbol{\delta}\right)\pi\left(\boldsymbol{\theta}^*\right)\Big/L\left(\boldsymbol{\theta}\middle|\mathbf{y},\mathbf{x},\boldsymbol{\delta}\right)\pi\left(\boldsymbol{\theta}\right).$$

The iterative steps of the random walk Metropolis algorithm can be described as follows:

Step 1: Initialize the parameters $\boldsymbol{\theta}^{(0)}=\left(\alpha_0^{(0)},\alpha_1^{(0)},\alpha_2^{(0)},\alpha_3^{(0)},\beta^{(0)}\right)$ for the algorithm using the maximum likelihood estimation (MLE) of the parameters $\boldsymbol{\theta}=\left(\alpha_0,\alpha_1,\alpha_2,\alpha_3,\beta\right)$.

Step 2: For $l=1,2,\ldots,L$ repeat the following steps;

a. Generate random error vector $\boldsymbol{\varepsilon}$ from a multivariate normal distribution with a zero-mean vector and variance-covariance matrix as a diagonal matrix in which the diagonal elements are the diagonal of the inverse of the observed Fisher's information matrix in (7); $\boldsymbol{\varepsilon} \sim \mathcal{N}\left(\boldsymbol{\mu}=\mathbf{0}, \boldsymbol{\Sigma}=diag\left(I^{-1}(\boldsymbol{\theta})\right)\right)$. Then, set $\boldsymbol{\theta}^{*}=\boldsymbol{\theta}^{(l-1)}+\boldsymbol{\varepsilon}$.

b. Calculate $p=\min(1, R_{\boldsymbol{\theta}})$ where $R_{\boldsymbol{\theta}}=L\left(\boldsymbol{\theta}^{*} \mid \mathbf{y}, \mathbf{x}, \boldsymbol{\delta}\right) \pi\left(\boldsymbol{\theta}^{*}\right) / L(\boldsymbol{\theta} \mid \mathbf{y}, \mathbf{x}, \boldsymbol{\delta}) \pi(\boldsymbol{\theta})$.

c. Generate u from a uniform distribution; $u \sim U(0,1)$.

If $u \leq p$, accept $\boldsymbol{\theta}^{*}$ and set $\boldsymbol{\theta}^{(l)}=\boldsymbol{\theta}^{*}$ with probability p.

If $u>p$, reject $\boldsymbol{\theta}^{*}$ and set $\boldsymbol{\theta}^{(l)}=\boldsymbol{\theta}^{(l-1)}$ with probability $1-p$.

Step 3: Remove B of the chain for burn-in.

Step 4: Calculate the estimated values of the Bayes estimators of the parameters $\boldsymbol{\alpha}$ and β under the squared error loss function from the average of the generated values given by

$$\hat{\theta}_{Bayes}=\frac{1}{L-B} \sum_{l=B+1}^{L} \theta^{(l)}, \tag{10}$$

where θ is a parameter in vector $\boldsymbol{\theta}=(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \beta)$.

3. Results and Discussion

3.1. Simulation study

In this section, a Monte Carlo simulation is conducted to assess and compare the performance of the maximum likelihood estimation and the Bayesian estimation for the discrete Weibull regression with type-I right censored data and various selected sample sizes $n=60$, 90, 120 and 150. The three explanatories are $x_{i1} \sim N(0,1)$, $x_{i2} \sim U(-0.3,0.3)$, and $x_{i3} \sim Ber(0.4)$. In particular, we select $(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \beta)=(0.1,-0.2,1.6,0.2,0.9)$ for the excessive zeros case, $(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \beta)=(-2.8,0.01,0.4,-0.2,2.5)$ for the under-dispersion case, and $(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \beta)=(-2.8,0.01,0.4,-0.2,0.9)$ for the over-dispersion case. We compute q_i for each type of data from the log-log link function in (2). We then generate the count response variables $Y_1^*, Y_2^*, \ldots, Y_n^*$ from (3) using function `rdw( )` from package `DWreg` in R. We censor the data using type-I right censoring with different percentages of censored data: censoring at $C=2,3$ for the excessive zeros case, censoring at $C=3,4,5$ for the under-dispersion case, and censoring at $C=29,34,40,49$ for the over-dispersion case. Then we get the response variables $y_1, y_2, \ldots, y_n$ as observed data from (4) and the indicator $\delta_1, \delta_2, \ldots, \delta_n$ is the censor indicator from (5).

Next, we calculate the maximum likelihood estimators of the parameters $\boldsymbol{\alpha}$ and β by minimizing the negative log-likelihood function of the discrete Weibull regression under the type-I right censored data model in (6). Then, we get $\hat{\boldsymbol{\theta}}_{ML}^{(m)}$ using function `optim( )` from package stats in R. We calculate the Bayes estimators of the parameters $\boldsymbol{\alpha}$ and β with uniform noninformative priors under the squared error loss function using the random walk Metropolis algorithm with $L=10{,}000$ replicates and 10% of the chain for burn-in; $B=1{,}000$. Then, we get $\hat{\boldsymbol{\theta}}_{Bayes(U)}^{(m)}$ from (10) where $R_{\boldsymbol{\theta}}$ in Step 2(b) is $R_{\boldsymbol{\theta}}=L\left(\boldsymbol{\theta}^{*} \mid \mathbf{y}, \mathbf{x}, \boldsymbol{\delta}\right) / L(\boldsymbol{\theta} \mid \mathbf{y}, \mathbf{x}, \boldsymbol{\delta})$. We calculate the Bayes estimators of the parameters $\boldsymbol{\alpha}$ and β with informative priors under the squared error loss function using the random walk Metropolis algorithm with $L = 10,000$ replicates and 10% of the chain for *burn-in*; $B = 1,000.$ Then, we get $\hat{\boldsymbol{\theta}}_{Bayes}^{(m)}$ from (10).

For each sample sizes $n,$ we repeat the previous steps for a $M = 1,000$ times. The parameter estimates and the mean squared error (MSE) of estimators based on $M = 1,000$ from the MLE and the Bayesian estimation with uniform noninformative priors (Bayes(Uniform)) and informative priors (Bayes(Informative)) are reported in Tables 1 and 2 for the excessive zeros case when censoring at $C = 3$ and $C = 2,$ Tables 3 to 5 for the under-dispersion case when censoring at $C = 5,$ $C = 4$ and $C = 3,$ and Tables 6 to 9 for the over-dispersion case when censoring at $C = 49,$ $C = 40,$ $C = 34$ and $C = 29.$

Table 1 Parameter estimates (Est.) and MSE for the excessive zeros case $\boldsymbol{\theta} = (0.1, -0.2, 1.6, 0.2, 0.9)$ at $C = 3$

n (% censored)	parameter	MLE		Bayes (Uniform)		Bayes (Informative)	
		Est.	MSE	Est.	MSE	Est.	MSE
60 (5.66%)	α_0	0.0947	**0.0494**	0.0494	0.0563	0.0901	0.0498
	α_1	–0.2339	**0.0347**	–0.2506	0.0415	–0.2365	0.0356
	α_2	1.7953	**1.2122**	1.9203	1.4732	1.8147	1.2437
	α_3	0.2449	**0.1235**	0.2544	0.1421	0.2419	0.1248
	β	1.0056	0.0634	1.0472	0.0789	0.9872	**0.0599**
90 (5.63%)	α_0	0.1057	**0.0317**	0.0759	0.0335	0.1031	0.0318
	α_1	–0.2161	**0.0184**	–0.2247	0.0206	–0.2172	0.0187
	α_2	1.7150	**0.6491**	1.7831	0.7326	1.7245	0.6582
	α_3	0.2118	**0.0660**	0.2157	0.0702	0.2098	0.0662
	β	0.9565	0.0373	0.9804	0.0417	0.9433	**0.0359**
120 (5.68%)	α_0	0.0981	**0.0230**	0.0752	0.0246	0.0957	0.0231
	α_1	–0.2128	**0.0134**	–0.2193	0.0146	–0.2138	0.0136
	α_2	1.7028	**0.4564**	1.7501	0.4921	1.7101	0.4578
	α_3	0.2230	**0.0535**	0.2261	0.0568	0.2215	0.0537
	β	0.9426	0.0256	0.9596	0.0277	0.9327	**0.0247**
150 (5.64%)	α_0	0.0993	**0.0201**	0.0816	0.0209	0.0976	0.0202
	α_1	–0.2118	**0.0100**	–0.2165	0.0106	–0.2125	0.0101
	α_2	1.6522	**0.3202**	1.6875	0.3389	1.6579	0.3212
	α_3	0.2144	**0.0420**	0.2169	0.0436	0.2133	**0.0420**
	β	0.9362	0.0194	0.9492	0.0209	0.9285	**0.0189**

Note: the boldface identifies the smallest MSE for each case.

Table 2 Parameter estimates (Est.) and MSE for the excessive zeros case $\boldsymbol{\theta} = (0.1, -0.2, 1.6, 0.2, 0.9)$ at $C = 2$

n (% censored)	parameter	MLE Est.	MLE MSE	Bayes (Uniform) Est.	Bayes (Uniform) MSE	Bayes (Informative) Est.	Bayes (Informative) MSE
60 (12.31%)	α_0	0.0967	**0.0504**	0.0439	0.0653	0.0915	0.0508
	α_1	−0.2351	**0.0354**	−0.2529	0.0453	−0.2382	0.0367
	α_2	1.7942	**1.2739**	1.9304	1.5840	1.8194	1.3095
	α_3	0.2440	**0.1296**	0.2558	0.1527	0.2412	0.1320
	β	1.0073	0.0898	1.0658	0.1801	0.9879	**0.0859**
90 (12.37%)	α_0	0.1070	**0.0324**	0.0734	0.0347	0.1032	0.0326
	α_1	−0.2159	**0.0200**	−0.2265	0.0228	−0.2177	0.0204
	α_2	1.6975	**0.6971**	1.7721	0.7883	1.7096	0.7088
	α_3	0.2142	**0.0686**	0.2186	0.0738	0.2113	0.0688
	β	0.9500	0.0517	0.9902	0.0594	0.9368	**0.0505**
120 (12.49%)	α_0	0.1001	**0.0241**	0.0744	0.0254	0.0970	**0.0241**
	α_1	−0.2142	**0.0143**	−0.2212	0.0157	−0.2152	0.0145
	α_2	1.6936	**0.4780**	1.7469	0.5207	1.7027	0.4832
	α_3	0.2234	**0.0575**	0.2270	0.0611	0.2215	0.0579
	β	0.9332	0.0339	0.9623	0.0373	0.9230	**0.0330**
150 (12.35%)	α_0	0.1001	**0.0202**	0.0798	0.0212	0.0977	0.0203
	α_1	−0.2108	**0.0108**	−0.2161	0.0116	−0.2115	0.0109
	α_2	1.6476	**0.3365**	1.6859	0.3615	1.6542	0.3388
	α_3	0.2158	**0.0432**	0.2186	0.0452	0.2146	0.0433
	β	0.9353	0.0279	0.9566	0.0299	0.9261	**0.0270**

Note: the boldface identifies the smallest MSE for each case.

Table 3 Parameter estimates (Est.) and MSE for the under-dispersion case $\boldsymbol{\theta} = (-2.8, 0.01, 0.4, -0.2, 2.5)$ at $C = 5$

n (% censored)	parameter	MLE Est.	MLE MSE	Bayes (Uniform) Est.	Bayes (Uniform) MSE	Bayes (Informative) Est.	Bayes (Informative) MSE
60 (4.56%)	α_0	–2.9553	**0.2158**	–3.0408	0.2578	–2.9588	0.2169
	α_1	0.0086	**0.0269**	0.0080	0.0280	0.0082	**0.0269**
	α_2	0.4309	**0.8177**	0.4359	0.8663	0.4300	0.8188
	α_3	–0.2099	0.1029	–0.2204	0.1072	–0.2122	**0.1024**
	β	2.6466	0.1234	2.6888	0.1433	2.6334	**0.1202**
90 (4.53%)	α_0	–2.8593	**0.1192**	–2.9147	0.1333	–2.8618	0.1198
	α_1	0.0157	**0.0144**	0.0161	0.0149	0.0157	**0.0144**
	α_2	0.4225	**0.4722**	0.4218	0.4870	0.4201	0.4727
	α_3	–0.2094	0.0614	–0.2159	0.0637	–0.2107	**0.0613**
	β	2.5705	0.0619	2.5971	0.0689	2.5617	**0.0611**
120 (4.53%)	α_0	–2.8528	**0.0937**	–2.8943	0.1021	–2.8542	0.0938
	α_1	0.0069	**0.0109**	0.0063	0.0110	0.0068	0.0108
	α_2	0.3998	**0.3274**	0.3984	0.3394	0.3961	0.3275
	α_3	–0.2149	0.0427	-0.2206	0.0430	–0.2165	**0.0422**
	β	2.5599	0.0511	2.5798	0.0556	2.5530	**0.0504**
150 (4.51%)	α_0	–2.8706	0.0752	–2.9027	0.0818	–2.8714	**0.0749**
	α_1	0.0106	**0.0081**	0.0106	0.0082	0.0105	**0.0081**
	α_2	0.4245	0.2613	0.4263	0.2657	0.4241	**0.2602**
	α_3	–0.1910	0.0322	–0.1952	0.0325	–0.1920	**0.0321**
	β	2.5613	0.0421	2.5762	0.0448	2.5553	**0.0412**

Note: the boldface identifies the smallest MSE for each case.

Table 4 Parameter estimates (Est.) and MSE for the under-dispersion case $\boldsymbol{\theta}=(-2.8, 0.01, 0.4, -0.2, 2.5)$ at $C=4$

n (% censored)	parameter	MLE Est.	MLE MSE	Bayes (Uniform) Est.	Bayes (Uniform) MSE	Bayes (Informative) Est.	Bayes (Informative) MSE
60 (17.06%)	α_0	–2.9354	**0.2310**	-3.0346	0.2765	–2.9381	0.2313
	α_1	0.0082	**0.0302**	0.0084	0.0322	0.0081	0.0305
	α_2	0.4408	**0.9009**	0.4535	0.9623	0.4438	0.9056
	α_3	–0.2104	**0.1136**	-0.2238	0.1200	–0.2138	0.1138
	β	2.6224	0.1372	2.6738	0.1586	2.6047	**0.1335**
90 (16.55%)	α_0	–2.8519	0.1355	–2.9141	0.1499	–2.8543	**0.1354**
	α_1	0.0158	**0.0162**	0.0159	0.0167	0.0157	**0.0162**
	α_2	0.4181	**0.5259**	0.4162	0.5477	0.4167	0.5310
	α_3	–0.2136	0.0703	–0.2217	0.0726	–0.2152	**0.0702**
	β	2.5626	0.0752	2.5939	0.0823	2.5514	**0.0738**
120 (16.72%)	α_0	–2.8452	0.1062	–2.8913	0.1148	–2.8458	**0.1057**
	α_1	0.0059	**0.0115**	0.0061	0.0117	0.0060	**0.0115**
	α_2	0.3796	0.3745	0.3793	0.3830	0.3776	**0.3742**
	α_3	–0.2147	**0.0471**	–0.2218	0.0486	–0.2164	0.0475
	β	2.5502	0.0631	2.5736	0.0673	2.5408	**0.0619**
150 (16.75%)	α_0	–2.8674	0.0839	–2.9058	0.0920	–2.8682	**0.0834**
	α_1	0.0105	**0.0089**	0.0108	0.0091	0.0107	**0.0089**
	α_2	0.4161	**0.3100**	0.4187	0.3197	0.4162	0.3111
	α_3	–0.1882	0.0369	–0.1944	0.0372	–0.1904	**0.0368**
	β	2.5558	0.0507	2.5756	0.0543	2.5488	**0.0498**

Note: the boldface identifies the smallest MSE for each case.

Table 5 Parameter estimates (Est.) and MSE for the under-dispersion case $\boldsymbol{\theta} = (-2.8, 0.01, 0.4, -0.2, 2.5)$ at $C = 3$

n (% censored)	parameter	MLE		Bayes (Uniform)		Bayes (Informative)	
		Est.	MSE	Est.	MSE	Est.	MSE
60 (42.10%)	α_0	–2.9373	0.2878	–3.0847	0.3746	–2.9434	**0.2860**
	α_1	0.0009	**0.0412**	0.0015	0.0437	0.0013	**0.0412**
	α_2	0.4241	**1.2051**	0.4271	1.3230	0.4226	1.2227
	α_3	–0.2180	**0.1615**	–0.2369	0.1768	–0.2237	0.1637
	β	2.6171	0.2215	2.7058	0.2653	2.5890	**0.2127**
90 (41.28%)	α_0	–2.8587	0.1752	–2.9518	0.2068	–2.8616	**0.1745**
	α_1	0.0181	**0.0227**	0.0187	0.0233	0.0181	**0.0227**
	α_2	0.3914	**0.7241**	0.3938	0.7695	0.3905	0.7322
	α_3	–0.2185	0.0992	–0.2312	0.1032	–0.2218	**0.0990**
	β	2.5649	0.1297	2.6214	0.1479	2.5463	**0.1266**
120 (41.48%)	α_0	–2.8490	0.1312	–2.9187	0.1483	–2.8517	**0.1311**
	α_1	0.0068	**0.0157**	0.0066	0.0160	0.0069	**0.0157**
	α_2	0.3658	**0.4952**	0.3632	0.5134	0.3620	0.4962
	α_3	–0.2157	0.0694	–0.2266	0.0711	–0.2189	**0.0693**
	β	2.5519	0.1017	2.5947	0.1115	2.5389	**0.1000**
150 (41.55%)	α_0	–2.8827	0.1113	–2.9376	0.1256	–2.8843	**0.1101**
	α_1	0.0100	0.0124	0.0103	0.0126	0.0103	**0.0123**
	α_2	0.4184	**0.4125**	0.4184	0.4327	0.4161	0.4146
	α_3	–0.1918	0.0526	–0.1996	0.0532	–0.1942	**0.0524**
	β	2.5723	0.0874	2.6049	0.0949	2.5605	**0.0850**

Note: the boldface identifies the smallest MSE for each case.

Table 6 Parameter estimates (Est.) and MSE for the over-dispersion case $\boldsymbol{\theta} = (-2.8, 0.01, 0.4, -0.2, 0.9)$ at $C = 49$

n (% censored)	parameter	MLE		Bayes (Uniform)		Bayes (Informative)	
		Est.	MSE	Est.	MSE	Est.	MSE
60 (15.95%)	α_0	–2.8892	0.2132	–3.0102	0.2417	–2.9071	**0.2041**
	α_1	0.0090	0.0322	0.0090	**0.0297**	0.0083	0.0301
	α_2	0.5447	1.0049	0.4456	0.8921	0.4934	**0.8125**
	α_3	–0.2221	0.1246	–0.2220	0.1136	–0.2207	**0.1130**
	β	0.9308	0.0166	0.9543	0.0174	0.9287	**0.0152**
90 (15.53%)	α_0	–2.8097	0.1314	–2.8942	0.1305	–2.8253	**0.1218**
	α_1	0.0167	0.0174	0.0160	**0.0158**	0.0162	0.0159
	α_2	0.4768	0.6416	0.4145	0.5159	0.4445	**0.5060**
	α_3	–0.2116	0.0767	–0.2191	0.0701	–0.2147	**0.0700**
	β	0.9089	0.0093	0.9272	0.0091	0.9096	**0.0084**
120 (15.58%)	α_0	–2.8115	0.1097	–2.8863	0.1049	–2.8272	**0.0980**
	α_1	0.0110	0.0122	0.0080	**0.0109**	0.0095	0.0112
	α_2	0.4606	0.5229	0.3787	**0.3664**	0.4179	0.3671
	α_3	–0.2133	0.0565	–0.2182	**0.0457**	–0.2152	0.0478
	β	0.9077	0.0083	0.9244	0.0078	0.9093	**0.0073**
150 (15.68%)	α_0	–2.8266	0.0878	–2.8938	0.0789	–2.8430	**0.0760**
	α_1	0.0151	0.0098	0.0122	**0.0084**	0.0137	0.0086
	α_2	0.4496	0.4691	0.4169	**0.2970**	0.4310	0.3107
	α_3	–0.1942	0.0469	–0.1923	**0.0363**	–0.1937	0.0386
	β	0.9082	0.0069	0.9233	0.0060	0.9106	**0.0059**

Note: the boldface identifies the smallest MSE for each case.

Table 7 Parameter estimates (Est.) and MSE for the over-dispersion case $\boldsymbol{\theta} = (-2.8, 0.01, 0.4, -0.2, 0.9)$ at $C = 40$

n (% censored)	parameter	MLE Est.	MLE MSE	Bayes (Uniform) Est.	Bayes (Uniform) MSE	Bayes (Informative) Est.	Bayes (Informative) MSE
60 (21.67%)	α_0	–2.8772	0.2196	–3.0158	0.2541	–2.8993	**0.2083**
	α_1	0.0103	0.0326	0.0085	0.0317	0.0094	**0.0310**
	α_2	0.5615	1.0951	0.4381	0.9470	0.4984	**0.8806**
	α_3	–0.2227	0.1347	–0.2268	0.1242	–0.2235	**0.1222**
	β	0.9252	0.0174	0.9549	0.0189	0.9248	**0.0159**
90 (21.05%)	α_0	–2.8052	0.1465	–2.8989	0.1380	–2.8209	**0.1301**
	α_1	0.0183	0.0181	0.0165	**0.0165**	0.0174	**0.0165**
	α_2	0.4826	0.7263	0.4068	0.5683	0.4426	**0.5566**
	α_3	–0.2231	0.0883	-0.2205	**0.0754**	-0.2211	0.0764
	β	0.9084	0.0109	0.9282	0.0100	0.9081	**0.0095**
120 (21.12%)	α_0	–2.8117	0.1223	–2.8908	0.1103	–2.8285	**0.1071**
	α_1	0.0114	0.0132	0.0078	**0.0117**	0.0095	0.0118
	α_2	0.4254	0.6258	0.3782	**0.3977**	0.3988	0.4166
	α_3	–0.2079	0.0621	–0.2154	**0.0493**	–0.2111	0.0517
	β	0.9060	0.0101	0.9252	0.0086	0.9084	**0.0084**
150 (21.23%)	α_0	–2.8216	0.0972	–2.8996	0.0855	–2.8407	**0.0817**
	α_1	0.0149	0.0103	0.0124	**0.0087**	0.0137	0.0090
	α_2	0.4632	0.5204	0.4134	**0.3188**	0.4370	0.3330
	α_3	–0.2003	0.0474	–0.1942	**0.0382**	–0.1979	0.0397
	β	0.9073	0.0078	0.9254	0.0068	0.9102	**0.0066**

Note: the boldface identifies the smallest MSE for each case.

Table 8 Parameter estimates (Est.) and MSE for the over-dispersion case $\boldsymbol{\theta} = (-2.8, 0.01, 0.4, -0.2, 0.9)$ at $C = 34$

n (% censored)	parameter	MLE		Bayes (Uniform)		Bayes (Informative)	
		Est.	MSE	Est.	MSE	Est.	MSE
60 (26.57%)	α_0	–2.8899	0.2355	–3.0260	0.2724	–2.9076	**0.2235**
	α_1	0.0077	0.0342	0.0071	0.0337	0.0075	**0.0328**
	α_2	0.4884	1.1340	0.4309	1.0156	0.4589	**0.9562**
	α_3	–0.2147	0.1338	–0.2270	0.1322	–0.2190	**0.1253**
	β	0.9281	0.0200	0.9574	0.0211	0.9256	**0.0181**
90 (26.02%)	α_0	–2.8053	0.1454	–2.9020	0.1473	–2.8218	**0.1353**
	α_1	0.0196	0.0186	0.0187	0.0177	0.0189	**0.0175**
	α_2	0.4695	0.7872	0.4068	0.6168	0.4350	**0.6037**
	α_3	–0.2160	0.0913	–0.2236	0.0815	–0.2183	**0.0797**
	β	0.9067	0.0118	0.9284	0.0116	0.9067	**0.0107**
120 (26.20%)	α_0	–2.8055	0.1251	–2.8902	0.1140	–2.8231	**0.1092**
	α_1	0.0123	0.0147	0.0077	**0.0127**	0.0101	0.0130
	α_2	0.4413	0.6286	0.3751	**0.4203**	0.4051	0.4244
	α_3	–0.2092	0.0661	–0.2165	**0.0533**	–0.2124	0.0551
	β	0.9029	0.0110	0.9239	0.0093	0.9054	**0.0092**
150 (26.15%)	α_0	–2.8316	0.1071	–2.9040	0.0903	–2.8472	**0.0899**
	α_1	0.0165	0.0108	0.0138	**0.0093**	0.0151	0.0096
	α_2	0.4579	0.5436	0.4146	**0.3262**	0.4337	0.3545
	α_3	–0.2006	0.0556	–0.1955	**0.0420**	–0.1978	0.0450
	β	0.9094	0.0089	0.9268	0.0076	0.9116	**0.0074**

Note: the boldface identifies the smallest MSE for each case.

Table 9 Parameter estimates (Est.) and MSE for the over-dispersion case
$\boldsymbol{\theta} = (-2.8, 0.01, 0.4, -0.2, 0.9)$ at $C = 29$

n (% censored)	parameter	MLE Est.	MLE MSE	Bayes (Uniform) Est.	Bayes (Uniform) MSE	Bayes (Informative) Est.	Bayes (Informative) MSE
60 (31.91%)	α_0	–2.8875	0.2318	–3.0291	0.2835	–2.9059	**0.2259**
	α_1	0.0068	0.0366	0.0058	0.0364	0.0063	**0.0354**
	α_2	0.5112	1.2369	0.4345	1.1243	0.4685	**1.0654**
	α_3	–0.2193	0.1578	–0.2322	0.1490	–0.2232	**0.1454**
	β	0.9255	0.0203	0.9563	0.0229	0.9221	**0.0192**
90 (30.95%)	α_0	–2.8063	0.1522	–2.9094	0.1561	–2.8235	**0.1403**
	α_1	0.0212	0.0199	0.0179	0.0191	0.0193	**0.0187**
	α_2	0.4339	0.8027	0.4031	0.6372	0.4177	**0.6227**
	α_3	–0.2240	0.0957	–0.2276	0.0857	–0.2243	**0.0842**
	β	0.9076	0.0128	0.9312	0.0130	0.9074	**0.0118**
120 (31.23%)	α_0	–2.8209	0.1191	–2.8943	0.1175	–2.8324	**0.1088**
	α_1	0.0101	0.0144	0.0076	0.0135	0.0089	**0.0133**
	α_2	0.4467	0.6259	0.3685	**0.4408**	0.4050	0.4485
	α_3	–0.2075	0.0642	–0.2174	0.0583	–0.2108	**0.0569**
	β	0.9072	0.0107	0.9251	0.0102	0.9073	**0.0095**
150 (31.22%)	α_0	–2.8331	0.1007	–2.9117	0.0962	–2.8512	**0.0889**
	α_1	0.0174	0.0117	0.0140	**0.0104**	0.0157	**0.0104**
	α_2	0.4798	0.6005	0.4134	**0.3530**	0.4442	0.3887
	α_3	–0.1894	0.0618	–0.1946	**0.0459**	–0.1924	0.0497
	β	0.9082	0.0093	0.9288	0.0086	0.9117	**0.0079**

Note: the boldface identifies the smallest MSE for each case.

From the numerical results for fixed censoring time C, as the sample size (n) increases, the MSE of the estimators decreases for all methods. In the cases of excessive zeros and under-dispersion, the performance of the MLE and the Bayes estimators with informative priors for parameter $\boldsymbol{\alpha}$ is quite similar and better than the Bayes estimators with uniform noninformative priors in terms of the MSE of the estimators. In the cases of over-dispersion, the performance of the Bayes estimators for the parameters $\boldsymbol{\alpha}$ and β is better than the MLE in terms of the MSE of the estimators. Moreover, the Bayes estimators with informative priors for parameter β shows the best performance for all cases in terms of the MSE of the estimators. Besides, note that the MSE of the estimators of α_2 are generally too big that may cause from we define a strong effect on x_2 or the high variance of the estimators of α_2. In addition, the MSE of estimators is depends on the explanatory variable. However, when n is large enough the MSE of estimators will decrease and not too big anymore.

Finally, we can summarize the best performance method for the discrete Weibull regression under type-I right censored data in Table 10.

Table 10 The best performance method for the discrete Weibull regression under type-I right censored data

n	parameter	Excessive zeros case	Under-dispersion case	Over-dispersion case
60, 90	$\boldsymbol{\alpha}$	MLE, Bayes (Informative)	MLE, Bayes (Informative)	Bayes (Informative)
	β	Bayes (Informative)	Bayes (Informative)	Bayes (Informative)
120, 150	$\boldsymbol{\alpha}$	MLE, Bayes (Informative)	MLE, Bayes (Informative)	Bayes (Uniform), Bayes (Informative)
	β	Bayes (Informative)	Bayes (Informative)	Bayes (Informative)

3.1 Real data application

In this section, a real dataset is presented to show the performance of the proposed methodology. The dataset is the German health registry for the years 1984-1988 available in the R package COUNT under the name *rwm*, with 27,326 observations. The response variable is the number of visits to doctors during the year and the three explanatory variables are age, years of formal education, and household yearly income. The response variable has 37.09% of zeros, the sample mean is 3.18, and the sample variance is 32.37, that is in the case of excessive zeros and over-dispersion data. Moreover, this dataset was found to be suitable for the Bayesian discrete Weibull regression model by Haselimashhadi et al. (2018). To demonstrate the approach with this dataset, we random sample sizes of $n = 60, 90, 120$ and 150 from this dataset that correspond with the main dataset, which has excessive zeros and is over-dispersion data using the bootstrap technique to estimate the standard errors of the different parameter estimates with the following process:

Step 1: Create a bootstrap sample of size n, $\mathbf{z}_1^*, \mathbf{z}_2^*, \ldots, \mathbf{z}_n^*$, from the original data $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n$; $\mathbf{z}_i = \left(y_i^*, \mathbf{x}_i\right)$, $i = 1, 2, \ldots, n$, with replacement giving $1/n$ probability for each $\mathbf{z}_i^*$, $i = 1, 2, \ldots, n$. Thus, we obtain the following: $\mathbf{z}_i^* = \left(y_i^{**}, \mathbf{x}_i^*\right)$, $i = 1, 2, \ldots, n$.

Step 2: Censor the bootstrap sample from Step 1 using type-I right censoring at $C = 3, 4, 5$. Thus, we obtain the censored data $y_1, y_2, \ldots, y_n$ from (4) and the censor indicator $\delta_1, \delta_2, \ldots, \delta_n$ from (5).

Step 3: Estimate the maximum likelihood estimators of the parameters $\boldsymbol{\alpha}$ and β by minimizing the negative log-likelihood function of the discrete Weibull regression under the type-I right censored data model in (6). Then, get $\hat{\boldsymbol{\theta}}_{ML}^{*(b)}$ using function `optim( )` from package stats in R.

Step 4: Estimate the Bayes estimators of the parameters $\boldsymbol{\alpha}$ and β with uniform non-informative priors under the squared error loss function using the random walk Metropolis algorithm with $L = 10,000$ replicates and 10% of the chain for *burn-in*; $B = 1,000$. Then, get $\hat{\boldsymbol{\theta}}_{Bayes(U)}^{*(b)}$ from (10) where $R_{\boldsymbol{\theta}}$ in Step 2(b.) is $R_{\boldsymbol{\theta}} = L\left(\boldsymbol{\theta}^* \middle| \mathbf{y}, \mathbf{x}, \boldsymbol{\delta}\right) / L\left(\boldsymbol{\theta} \middle| \mathbf{y}, \mathbf{x}, \boldsymbol{\delta}\right)$.

Step 5: Estimate the Bayes estimators of the parameters $\boldsymbol{\alpha}$ and β with informative priors under the squared error loss function using the random walk Metropolis algorithm with $L = 10,000$ replicates and 10% of the chain for *burn-in*; $B = 1,000$. Then, get $\hat{\boldsymbol{\theta}}_{Bayes}^{*(b)}$ from (10).

Step 6: For each sample size $n,$ repeat Step 1 to Step 5 a $B^{*}=1,000$ times.

Step 7: Calculate the parameter estimates and standard error (SE) of estimators based on $B^{*}=1,000.$

Figure 1 shows the SE of estimators plots for each sample size using methods only censoring at $C=5\ (\approx 12.78-16.40\%)$. Moreover, the results of censoring at $C=4\ (\approx 16.15-19.73\%)$ and $C=3\ (\approx 32.49-34.68\%)$ are similar to censoring at $C=5.$ For fixed censoring time C, as the sample size (n) increases the SE of the estimators decreases for all methods. In addition, the performance of the MLE and the Bayes estimators with informative priors for the parameters $\boldsymbol{\alpha}$ and β is quite similar and better than the Bayes estimators with uniform noninformative priors in terms of the parameter estimates and the SE of the estimators. The result of the application to health data from the German health registry for the years 1984-1988 is close to the simulation result of the excessive zeros case when censoring at $C=2(\approx 12.31-12.49\%)$.

4. Conclusions

In this article, we have considered the classical and Bayesian inference for the discrete Weibull regression under type I right censored data. The results of the simulation showed that as the sample size increases the MSE of the estimators decreases for all methods, indicating that the estimators are consistent. Moreover, the results of an application to the real data revealed that as the sample size increases the SE of the estimators decreases for all methods, indicating that the estimators are precise. In the cases of excessive zeros and under-dispersion, the MLE and the Bayes estimators with informative priors for the parameters $\boldsymbol{\alpha}$ and β are both appropriate for the discrete Weibull regression under type-I right censored data in terms of the MSE of the estimators. In addition, the MSE of the MLE and the Bayes estimators with informative priors are not difference at the third decimal. However, in the case of over-dispersion, the Bayes estimators with informative priors for the parameters $\boldsymbol{\alpha}$ and β are more appropriate for the discrete Weibull regression under type I right censored data than other methods in terms of the MSE of the estimators.

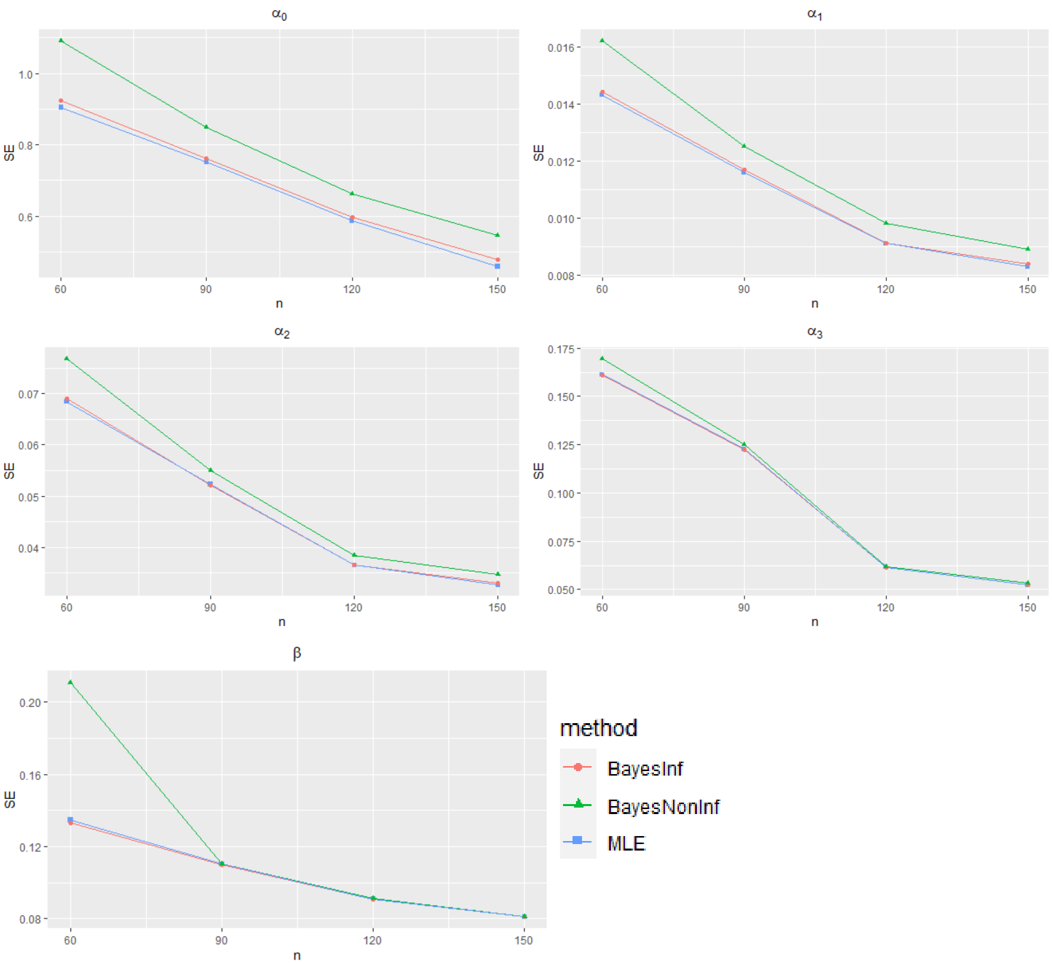

Figure 1 SE plots for each sample size by method when censoring at $C = 5$

Acknowledgements

The research of Dr. Duangsaphon is currently supported by the Thammasat University Research Unit in Theoretical and Computational Statistics. We gratefully acknowledge financial support from Thammasat University.

References

Allison PD, Waterman RP. Fixed-effects negative binomial regression models. Sociol Methodol. 2002; 32(1): 247-265.

Aslam M, Kazmi SMA, Ahmad I, Shah SH. Bayesian estimation for parameters of the Weibull distribution. Science International (Lahore). 2014; 26(5): 1915-1920.

Chacko M, Mohan R. Bayesian analysis of Weibull distribution based on progressive type-II censored competing risks data with binomial removals. Computation Stat. 2019; 34(4): 233-252.

Chanialidis C, Evers L, Neocleous T, Nobile A. Efficient Bayesian inference for COM-Poisson regression models. Stat Comput. 2018; 28(1): 595-608.

Collins K, Waititu A, Wanjoya A. Discrete Weibull and artificial neural network models in modelling over-dispersed count data. Int J Data Sci Anal. 2020; 6(5):153-162.

Fu S. Hierarchical Bayesian LASSO for a negative binomial regression. J Stat Comput Sim. 2016; 86(11): 2182-2203.

Gardner W, Mulvey EP, Shaw EC. Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. Psychol Bull. 1995; 118(3): 392-404.

Gelman A, Jakulin A, Pittau MG, Su Y. A weakly informative default prior distribution for logistic and other regression models. Ann Appl Stat. 2008; 2(4): 1360-1383.

Haselimashhadi H, Vinciotti V, Yu K. A novel Bayesian regression model for counts with an application to health data. J Appl Stat. 2018; 45(6): 1085-1105.

Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. Biometrika. 1970; 57(1):97-109.

Hutchinson MK, Holtman MC. Analysis of count data using Poisson regression. Res Nurs Health. 2005; 28(5): 408-418.

Kalktawi HS. Discrete Weibull regression model for count data. PhD [dissertation]. London: Brunel University London; 2017.

Kalktawi HS, Vinciotti V, Yu K. A simple and adaptive dispersion regression model for count data. Entropy. 2018; 20(2): 1-15.

Lee ET, Wang JW. Statistical methods for survival data analysis. Hoboken: John Wiley & Sons; 2003.

Lovett A, Flowerdew R. Analysis of count data using Poisson regression. Prof Geogr. 1989; 41(2): 190-198.

Nakagawa T, Osaki S. The discrete Weibull distribution. IEEE T Reliab. 1975; R-24(5): 300-301.

Saffari SE, Adnan R, Greene W, Ahmad MH. A Poisson regression model for analysis of censored count data with excess zeroes. Jurnal Teknologi. 2013; 63(2):71-74.

Saffari SE, Allen JC. Bivariate negative binomial regression model with excess zeros and right censoring: an application to Indonesian data. J Appl Stat. 2019; 47(10): 1901-1914.

Yu X. Censored count data analysis – statistical techniques and applications. PhD [dissertation]. Houston: The University of Texas School of Public Health; 2018.