To appear in the *Journal of Applied Statistics*
Vol. 00, No. 00, Month 20XX, 1–20

A novel Bayesian regression model for counts with an application to health data

H. Haselimashhadi[a], V. Vinciotti[a*] and K. Yu[a]

[a]*Department of Mathematics, Brunel University London, UB83PH UK*

(*Received 00 Month 20XX; accepted 00 Month 20XX*)

Discrete data are collected in many application areas and are often characterised by highly-skewed distributions. An example of this, which is considered in this paper, is the number of visits to a specialist, often taken as a measure of demand in healthcare. A discrete Weibull regression model was recently proposed for regression problems with a discrete response and it was shown to possess desirable properties. In this paper, we propose the first Bayesian implementation of this model. We consider a general parametrization, where both parameters of the discrete Weibull distribution can be conditioned on the predictors, and show theoretically how, under a uniform non-informative prior, the posterior distribution is proper with finite moments. In addition, we consider closely the case of Laplace priors for parameter shrinkage and variable selection. Parameter estimates and their credible intervals can be readily calculated from their full posterior distribution. A simulation study and the analysis of four real datasets of medical records show promises for the wide applicability of this approach to the analysis of count data. The method is implemented in the R package `BDWreg`.

Keywords: Discrete Weibull; Bayesian inference; Discrete response

1. Introduction

Data in the form of counts appear in many application areas, from medicine, social and natural sciences to econometrics, finance and industry [10]. In medicine, two examples of this are the length of stay in hospital, commonly used as an indicator of the quality of care and planning capacity within a hospital [4, 11], and the number of visits to a specialist [32], often taken as a measure of demand in healthcare. Other examples are high-throughput genomic data generated by next generation sequencing experiments [5, 38, 43] or lifetime data, such as the number of cycles before a machine breaks down [35].

Similarly to Weibull regression, which is widely used in lifetime data analysis and survival analysis for continuous response variables, [24] have recently proposed a regression model for a discrete response based on the discrete Weibull distribution. A number of studies have found a good fit of this distribution in comparison with other distributions for count data [9, 15, 29]. In the context of regression, [24] show two important features of a discrete Weibull distribution that make this a valuable alternative to the more traditional Poisson and Negative Binomial distributions and their extensions, such as Poisson mixtures [22], Poisson-Tweedie [16], zero-inflated semiparametric regression [30] and COMPoisson [44]: the ability to capture both over and under-dispersion and a closed-form analytical expression of the quantiles of the conditional distribution.

In [24], maximum likelihood is used for the estimation of the parameters. This is in

*Corresponding author. Email: veronica.vinciotti@brunel.ac.uk

general the most common approach for parameter estimation in regression analysis of counts, due to a lack of simple and efficient algorithms for posterior computation [49]. Among the contributions to Bayesian estimation of discrete regression models, [14] consider the case of Poisson regression, [49] provide an efficient Bayesian implementation of negative Binomial regression, [34] develop Bayesian estimation for a Poisson and negative Binomial regression with a conditional autoregressive correlation structure, whereas [2, 18, 31, 36] study zero-inflated Poisson regression. In this paper, we contribute to this literature, by providing the first Bayesian approach for parameter estimation in discrete Weibull regression. For the choice of prior distributions, we consider both the case of non-informative priors, such as Jeffreys and uniform, and the case of Laplace priors with a hyper penalty parameter. We prove that under a uniform non-informative prior, the posterior distribution is proper with finite moments. The choice of Laplace priors induces parameter shrinkage [28, 40], and, with the use of Bayesian credible intervals, leads to variable selection, similar to alternative approaches such as spike and slab priors [23].

The remainder of this paper is organized as follows. Section (2) describes the discrete Weibull regression model, with a more general parametrization than that presented in [24]. Section (3) describes Bayesian parameter estimation for a discrete Weibull regression model. Section (4) proves that the posterior distribution is proper under a uniform non-informative prior distribution on the parameters. Section (5) presents an extensive simulation study, whereas Section (6) shows the analysis of real data and a comparison with existing approaches. Finally, we draw some conclusions in Section (7).

2. Discrete Weibull regression

2.1 *Discrete Weibull distribution*

The discrete Weibull distribution was introduced by [35], as a discretized form of a continuous Weibull distribution, similarly to the geometric distribution, which is the discretized form of the exponential distribution, and the negative Binomial, which is the discrete alternative of a Gamma distribution. In some papers, this is referred to as a type I discrete Weibull, as two other distributions were subsequently defined. [9] review the three different distributions and point out the advantages of using the type I distribution: it has an unbounded support, in contrast to the type II distribution, and it has a more straightforward interpretation, in contrast to the type III distribution.

If a random variable Y follows a (type I) discrete Weibull distribution, then the cumulative distribution function of Y is given by

$$F(y; q, \beta) = \begin{cases} 1 - q^{(y+1)^\beta} & \text{if } y = 0, 1, 2, \ldots \text{(jump points)} \\ 0 & \text{if } y < 0 \end{cases} \tag{1}$$

with $0 < q < 1$ and $\beta > 0$ the shape parameters. A similar definition can be given on the support $1, 2, \ldots$. In this case, $F(y; q, \beta) = 1 - q^{y^\beta}$, for $y = 1, 2, \ldots$. Comparing this cdf with that of a continuous Weibull distribution with parameters α and β, which is given by

$$F_C(y; \alpha, \beta) = 1 - \exp\left[-\left(\frac{y}{\alpha}\right)^\beta\right] \quad \text{for } y > 0, \tag{2}$$

one can see that there is a direct correspondence between β, whereas q in the discrete case corresponds to $\exp\left[-\left(\frac{1}{\alpha}\right)^\beta\right]$ in the continuous case [26].

Given the form of the cumulative distribution function, the discrete Weibull distribution has the following probability mass function:

$$p(y; q, \beta) = q^{y^\beta} - q^{(y+1)^\beta}, \quad y = 0, 1, 2, ...$$

with q and β denoting the shape parameters. Throughout the paper, we will refer to this distribution as DW(q, β).

2.2 *Inference for Discrete Weibull: Existing Approaches*

[26] derive estimators of the parameters q and β using the method of moments and a new method which they call the method of proportions, and they find a good performance for the latter. Let $Y_1, \ldots, Y_n$ be a random sample from a DW(q, β) distribution and denote $Z = \sum_{i=1}^n I(Y_i = 0)$ and $U = \sum_{i=1}^n I(Y_i = 1)$. Using the method of proportions, the following estimators of q and β are proposed:

$$\hat{q} = 1 - \frac{Z}{n}$$
$$\hat{\beta} = \ln\left[\ln\left(1 - \frac{Z}{n} - \frac{U}{n}\right) / \ln\left(1 - \frac{Z}{n}\right)\right] / \ln(2).$$

These estimators use only the zeros and ones in the sample. [3] derive an improved estimator of β, by taking all observations into account. In particular, let d_m be the maximum observed value of Y and let $k = d_m - 1$. If $d_m > 2$, then the following improved estimator is proposed:

$$\hat{\beta} = \frac{1}{k} \sum_{d=1}^{k} \ln\left[\ln\left(1 - \hat{F}(d)\right) / \ln(\hat{q})\right] / \ln(d+1),$$

where $\hat{F}$ denotes the empirical cdf. When $d_m = 2$, this estimator is equivalent to the one from [26]. Note that in both cases, no estimates of β can be obtained when $\hat{q} = 1$, i.e. there are no zero counts in the observed data, or $\hat{q} = 0$, i.e. all counts are zero. However, in other cases, the estimators perform relatively well, particularly in the case of small sample sizes.

[27] considers maximum likelihood for the estimation of q and β. The likelihood function for a discrete Weibull sample is given by:

$$L(y|q, \beta) = \prod_{i=1}^{n} \left(q^{y_i^\beta} - q^{(y_i+1)^\beta}\right),$$

the maximum of which can be found numerically.

There is no explicit work in the literature for building confidence intervals for discrete Weibull parameters, although standard asymptotic likelihood and bootstrap approaches can be used. The Bayesian approach that we devise in this paper will lead naturally to credible intervals for the parameters.

2.3 *Regression via a discrete Weibull*

Let Y be the response variable with possible values $0, 1, \ldots,$ and let $X_1, \ldots, X_p$ be p covariates. We assume that the conditional distribution of Y given X follows a DW distribution with parameters q and β. Similar to the approach of [42], we introduce a regression model by linking each parameter, q and β, to the predictors. In particular, we propose the following link functions:

(1) q dependent on X via

$$\log(-\log(q)) = X\boldsymbol{\theta} \text{ or}$$
$$\log\left(\frac{q}{1-q}\right) = X\boldsymbol{\theta},$$

where $X = (1\, X_1 \ldots X_p)$ and $\boldsymbol{\theta} = (\theta_0 \ldots \theta_p)'$.

(2) β dependent on X via

$$\log(\beta) = X\boldsymbol{\gamma},$$

where $\boldsymbol{\gamma} = (\gamma_0\, \gamma_1 \ldots \gamma_p)'$.

The first parametrization was proposed by [24], in line with the link function used in continuous Weibull regression. As $q = 1 - P(Y = 0)$ for a DW distribution, we consider also a parametrization for q via a logit link function, which is commonly used in classification problems, for probabilities that are bounded between 0 and 1. In addition, we consider a log link for the parameter β, in order to capture more complex dependencies. In general, there are no identifiability issues in the model, as the part of the likelihood from zero observations depends only on q. However, in our simulation and real data analyses, we found the logit function to be only marginally superior to the log-log transformation, whereas the additional β parametrization leads to a significant increase in the number of parameters and is therefore often not selected.

3. Bayesian inference for discrete Weibull regression

In this section, we discuss Bayesian estimation of the regression parameters $\boldsymbol{\theta} = (\theta_0 \ldots \theta_p)'$ and $\boldsymbol{\gamma} = (\gamma_0 \ldots \gamma_p)'$. The advantage of choosing Bayesian approaches over classical maximum likelihood inference is two-fold. Firstly, the possibility of taking prior information into account, such as sparsity or information from historical data, and, secondly, the procedure returns automatically the distribution of all parameters, from which credible intervals can easily be obtained.

Given n observations y_i and $(x_{i1} \ldots x_{ip})$, $i = 1, \ldots, n$, for the response Y and the covariates X, respectively, and letting x_i be the row vector $x_i = (1\, x_{i1} \ldots x_{ip})$, the likelihood for the most general case (and using a logit(q) link) is given by

$$L(x, y|\boldsymbol{\theta}, \boldsymbol{\gamma}) = \prod_{i=1}^{n} \left(\left(\frac{e^{x_i\boldsymbol{\theta}}}{1 + e^{x_i\boldsymbol{\theta}}}\right)^{{y_i}^{exp(x_i\boldsymbol{\gamma})}} - \left(\frac{e^{x_i\boldsymbol{\theta}}}{1 + e^{x_i\boldsymbol{\theta}}}\right)^{(y_i+1)^{exp(x_i\boldsymbol{\gamma})}} \right).$$

We consider different prior distributions on $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$. Unfortunately, in the context of discrete Weibull regression, there are no conjugate priors. However, in the next section, we will show theoretically how a uniform non-informative prior leads to a posterior distribution which is proper with finite moments and, in the simulation and real data study, we show how this prior achieves an acceptable rate of mixing as well as comparable estimation to maximum likelihood. Similar results are obtained using an objective

Jeffreys prior, which we also discuss in the next sections. In addition, we consider closely a proper prior distribution on the regression coefficients that encourages sparsity, by setting a Laplace prior on θ and γ, i.e

$$\begin{aligned} p(\boldsymbol{\theta}|\lambda) &= \tfrac{\lambda}{2}e^{-\lambda|\boldsymbol{\theta}|}, \qquad \lambda > 0, \\ p(\boldsymbol{\gamma}|\tau) &= \tfrac{\tau}{2}e^{-\tau|\boldsymbol{\gamma}|}, \qquad \tau > 0. \end{aligned} \tag{3}$$

For a given choice of λ and τ, maximising the posterior probability under these priors corresponds to maximising the L_1 penalised log-likelihood

$$\log L(x, y|\boldsymbol{\theta}, \boldsymbol{\gamma}) - \lambda \sum_{j=1}^{p} |\theta_j| - \tau \sum_{k=1}^{p} |\gamma_k|,$$

as in the traditional lasso approach [40, 48]. We further assume an InverseGamma(a,b) hyper prior for both λ and τ, leading to the posterior distribution

$$p(\boldsymbol{\theta}, \boldsymbol{\gamma}|x, y) \propto L(x, y|\boldsymbol{\theta}, \boldsymbol{\gamma}) \times p(\boldsymbol{\theta}|\lambda) \times p(\boldsymbol{\gamma}|\tau) \times p(\lambda) \times p(\tau).$$

As conjugate priors are not available, we choose an adaptive Metropolis-Hastings sampling to draw samples from the full conditional posterior [20, 21]. MCMC samplers have been used before in the continuous Weibull regression context by [37], which utilizes a Reversible Jump MCMC, and [46] which uses a hybrid method consisting of Metropolis-Hastings and Gibbs sampler to estimate parameters in a three parameters continuous Weibull distribution. Moreover, [41] make use of a Metropolis-Hasting sampler to make inference for a continuous two-parameters Weibull distribution in a censoring framework. In line with this literature, we consider the following adaptive Metropolis-Hastings procedure:

Step 1. Initialize the algorithm with MLE estimation of the parameters or random values within the space of the parameters.
Step 2. Set a proposal distribution $g(.)$ on the full set of parameters $\pi = (\theta, \gamma)$. We choose a multivariate normal proposal with covariance matrix set from the Fisher information matrix, but other choices are possible.
Step 3. Draw a random sample from the proposal distribution, e.g. π_k at iteration k.
Step 4. Evaluate the acceptance probability

$$\alpha = \min\left(1, \frac{L(X, y|\pi_k)p(\pi_k)g(\pi_{k-1}|\pi_k)}{L(X, y|\pi_{k-1})p(\pi_{k-1})g(\pi_k|\pi_{k-1})}\right),$$

where $L(X, y|\pi_k)$ is the conditional DW likelihood given the proposal values and $p(.)$ is the prior.
Step 5. Accept the proposal π_k with probability α.
Step 6. Following the adaptive Metropolis-Hasting approach of [20], we update the covariance of the proposal by computing the sample covariance of the estimated chain.
Step 7. Remove a percentage of the chain for burn-in. We use 25%, in our simulations and real applications.

This algorithm is implemented in the R package `BDWreg`, which is freely available in CRAN at `https://cran.r-project.org/package=BDWreg`. From the posterior distribution, the mode of the marginal densities can be used as point estimate of the parameters, whereas the whole distribution is used for building credible intervals. In the case of Laplace priors, the inclusion or not of zero in the Highest Posterior Density (HPD) interval is used to guide variable selection as in [40], although Bayes factors should be used for model selection. In terms of computational complexity, DW and NB distributions have the same number of parameters, but there are fewer operations involved in the evaluation of the DW distribution than in the NB distribution, leading to an expected lower computational complexity for DW.

4. Some key theoretical results

Although a standard conjugate prior distribution is not available for the discrete Weibull regression model, MCMC methods can be used to draw samples from the posterior distributions, as described above. This, in principal, allows us to use virtually any prior distribution. However, in the case of non-informative priors, we should select only those that yield proper posteriors. In this section, we show some key theoretical results on this. In particular, we prove that the choice of uniform non-informative priors on the parameters, i.e. $p(\theta) \propto 1$ and $p(\gamma) \propto 1$, leads to a proper posterior distribution with finite moments.

Thus, as a first result, we show that, under uniform non-informative priors, the posterior is proper, that is

$$0 < \int_{\boldsymbol{\theta}} \int_{\boldsymbol{\gamma}} L(x, y|\boldsymbol{\theta}, \boldsymbol{\gamma}) d\boldsymbol{\gamma}\, d\boldsymbol{\theta} < \infty.$$

For simplicity, we consider the case where there is no regression model on β, i.e. $p(\beta) \propto 1$ for $\beta > 0$. In addition, we consider the logit link for q, although the proof will cover also the log-log case, as we explain later on.

Lemma 4.1 *Let*

$$f(y) = 1 - (e^{-a})^{y^\beta} - \Big(\frac{a}{1+a}\Big)^{y^\beta},$$

and assuming $y > 0$, $\beta > 0$ and $a > 0$, then $f(y)$ is an increasing function of y.

Proof. The derivative of f with respect to y is

$$\begin{aligned} \frac{d\, f(y)}{dy} &= -\beta y^{\beta-1} e^{-ay^\beta} \log(e^{-a}) - \beta y^{\beta-1} \Big(\frac{a}{1+a}\Big)^{y^\beta} \log\Big(\frac{a}{1+a}\Big) \\ &= \beta y^{\beta-1} \Big(a e^{-ay^\beta} - \Big(\frac{a}{1+a}\Big)^{y^\beta} \log\Big(\frac{a}{1+a}\Big)\Big). \end{aligned}$$

Since $a > 0$ and $\log(\frac{a}{1+a}) < 0$, the derivative is always positive.

■

Theorem 4.2 *Let y, X and $\boldsymbol{\theta} = (\theta_0 \dots \theta_p)'$ be vector of response, matrix of covariates and regression parameters, respectively. Let $S = \{i; y_i \neq 0, y_i \neq 1\} \neq \emptyset$ and assume X is of full rank. Under the DW regression model $Y|x \sim DW\Big(\frac{e^{x\boldsymbol{\theta}}}{1+e^{x\boldsymbol{\theta}}}, \beta\Big)$ and choosing uniform non-informative priors on $\boldsymbol{\theta}$ and β, i.e. $p(\theta) \propto 1$ and $p(\beta) \propto 1$ for $\beta > 0$, the* *posterior distribution is proper, i.e.*

$$\int_\beta \int_{\boldsymbol{\theta}} \prod_{i=1}^{n} \left\{ \left(\frac{e^{x_i\boldsymbol{\theta}}}{1+e^{x_i\boldsymbol{\theta}}} \right)^{y_i^\beta} - \left(\frac{e^{x_i\boldsymbol{\theta}}}{1+e^{x_i\boldsymbol{\theta}}} \right)^{(y_i+1)^\beta} \right\} d\boldsymbol{\theta}\, d\beta < \infty.$$

Proof. Since $S \neq \emptyset$,

$$\prod_{i=1}^{n} \left\{ \left(\frac{e^{x_i\boldsymbol{\theta}}}{1+e^{x_i\boldsymbol{\theta}}} \right)^{y_i^\beta} - \left(\frac{e^{x_i\boldsymbol{\theta}}}{1+e^{x_i\boldsymbol{\theta}}} \right)^{(y_i+1)^\beta} \right\} \leq \prod_{i\in S} \left\{ \left(\frac{e^{x_i\boldsymbol{\theta}}}{1+e^{x_i\boldsymbol{\theta}}} \right)^{y_i^\beta} - \left(\frac{e^{x_i\boldsymbol{\theta}}}{1+e^{x_i\boldsymbol{\theta}}} \right)^{(y_i+1)^\beta} \right\}.$$

Choosing any $k \in S$, such that $\min |x_{kj}| \neq 0, j = 1, \ldots, p$, results in

$$\begin{aligned}
\prod_{i\in S} \left\{ \left(\frac{e^{x_i\boldsymbol{\theta}}}{1+e^{x_i\boldsymbol{\theta}}} \right)^{y_i^\beta} - \left(\frac{e^{x_i\boldsymbol{\theta}}}{1+e^{x_i\boldsymbol{\theta}}} \right)^{(y_i+1)^\beta} \right\} &\leq \left(\frac{e^{|x_k|\boldsymbol{\theta}}}{1+e^{|x_k|\boldsymbol{\theta}}} \right)^{y_k^\beta} - \left(\frac{e^{|x_k|\boldsymbol{\theta}}}{1+e^{|x_k|\boldsymbol{\theta}}} \right)^{(y_k+1)^\beta} \\
&= \left(\frac{e^{|x_k|\boldsymbol{\theta}}}{1+e^{|x_k|\boldsymbol{\theta}}} \right)^{y_k^\beta} \left(1 - \left(\frac{e^{|x_k|\boldsymbol{\theta}}}{1+e^{|x_k|\boldsymbol{\theta}}} \right)^{(y_k+1)^\beta - y_k^\beta} \right) \\
&\leq \left(\frac{e^{|x_k|\boldsymbol{\theta}}}{1+e^{|x_k|\boldsymbol{\theta}}} \right)^{y_k^\beta}.
\end{aligned}$$

Without loss of generality, we assume $p = 1$, so $\boldsymbol{\theta} = (\theta_0\ \theta_1)'$ and $x_k = (1\ x_{k1})$. Then we consider the four cases where the θs are both positive, negative or of different signs, respectively.
Assuming $\theta_j \leq 0, j = 0, 1$ we get,

$$\left(\frac{e^{|x_k|\boldsymbol{\theta}}}{1+e^{|x_k|\boldsymbol{\theta}}} \right)^{y_k^\beta} \leq (e^{|x_k|\boldsymbol{\theta}})^{y_k^\beta},$$

where the integral over $\boldsymbol{\theta}$ and β is bounded $(\leq \dfrac{1}{2|x_{k1}|\log(y_k)})$.

Similarly, for $\theta_0 \leq 0$ and $\theta_1 > 0$ we have,

$$\left(\frac{e^{|x_k|\boldsymbol{\theta}}}{1+e^{|x_k|\boldsymbol{\theta}}} \right)^{y_k^\beta} \leq (e^{\theta_0})^{y_k^\beta} \left(1 + e^{-|x_{k1}|\theta_1} \right)^{-\beta},$$

and

$$\begin{aligned}
\int_0^\infty \int_{\theta_1=0}^\infty \int_{\theta_0=-\infty}^0 (e^{\theta_0})^{y_k^\beta} \left(1 + e^{-|x_{k1}|\theta_1} \right)^{-\beta} d\theta_0 d\theta_1 d\beta \\
&= \int_{\theta_1=0}^\infty \int_{\beta=0}^\infty \frac{1}{y_k^\beta} \left(1 + e^{-|x_{k1}|\theta_1} \right)^{-\beta} d\beta\, d\theta_1 \\
&= \int_{\theta_1=0}^\infty \frac{1}{\log\left(y_k(1+e^{-|x_{k1}|\theta_1}) \right)} d\theta_1.
\end{aligned}$$

The function $\log^{-1}\left(y_k(1+e^{-|x_{k1}|\theta_1}) \right)$ is continuous and bounded over the domain of

θ_1, provided $y_k > 1$. Thus, the integral is bounded. A similar derivation would hold for the case $\theta_0 > 0$, $\theta_1 \leq 0$.

For the final case, $\theta_j > 0, j = 0, 1$, we have

$$\prod_{i=1}^{n}\left\{\left(\frac{e^{x_i\boldsymbol{\theta}}}{1+e^{x_i\boldsymbol{\theta}}}\right)^{y_i^\beta}-\left(\frac{e^{x_i\boldsymbol{\theta}}}{1+e^{x_i\boldsymbol{\theta}}}\right)^{(y_i+1)^\beta}\right\} \leq \left(1+e^{-|x_k|\boldsymbol{\theta}}\right)^{-y_k^\beta} - \left(1+e^{-|x_k|\boldsymbol{\theta}}\right)^{-(y_k+1)^\beta}$$
$$\leq e^{-e^{|x_k|\boldsymbol{\theta}}(y_k+1)^\beta} - e^{-e^{|x_k|\boldsymbol{\theta}}y_k^\beta}$$
$$= e^{-e^{\theta_0}(y_k+1)^\beta}e^{-e^{|x_{k1}|\theta_1}(y_k+1)^\beta} - e^{-e^{\theta_0}y_k^\beta}e^{-e^{|x_{k1}|\theta_1}y_k^\beta}$$
$$\leq e^{-\theta_0(y_k+1)^\beta}e^{-|x_{k1}|\theta_1(y_k+1)^\beta} - e^{-\theta_0 y_k^\beta}e^{-|x_{k1}|\theta_1 y_k^\beta},$$

where the last term is a direct result of Lemma (4.1) with $a = e^{|x_{kj}|\theta_j}, j = 0, 1$. Thus,

$$\int_{\theta_1=0}^{\infty}\int_{\theta_0=0}^{\infty} e^{-e^{\theta_0}(y_k+1)^\beta}e^{-e^{|x_{k1}|\theta_1}(y_k+1)^\beta} - e^{-e^{\theta_0}y_k^\beta}e^{-e^{|x_{k1}|\theta_1}y_k^\beta}\,d\theta_0\,d\theta_1$$
$$\leq \int_{\theta_1=0}^{\infty}\int_{\theta_0=0}^{\infty} e^{-\theta_0(y_k+1)^\beta}e^{-|x_{k1}|\theta_1(y_k+1)^\beta} - e^{-\theta_0 y_k^\beta}e^{-|x_{k1}|\theta_1 y_k^\beta}\,d\theta_0\,d\theta_1$$
$$= \frac{1}{|x_{k1}|}\left(\frac{1}{(y_k+1)^{2\beta}} - \frac{1}{y_k^{2\beta}}\right),$$

and

$$\int_{\beta=0}^{\infty}\frac{1}{|x_{k1}|}\left(\frac{1}{(y_k+1)^{2\beta}} - \frac{1}{y_k^{2\beta}}\right)d\beta = \frac{1}{|x_{k1}|}\left(\frac{-(y_k+1)^{-2\beta}}{2\log(y_k+1)} + \frac{y_k^{-2\beta}}{2\log(y_k)}\right)\Bigg|_{\beta=0}^{\beta=\infty}$$
$$= \frac{1}{2|x_{k1}|}\frac{\log(\frac{y_k+1}{y_k})}{\log(y_k)\log(y_k+1)} < \infty,$$

which completes the proof. Similar derivations can be carried out in the general case of $p > 1$ by finding the bounds for all different combinations of signs of the parameters, following a similar approach to that used in this proof. ■

Having proved that the posterior is proper, in the following theorem we show that the posterior moments exist and are finite.

THEOREM 4.3 *Under the same conditions of Theorem (4.2), the posterior distribution of $(\boldsymbol{\theta}, \beta)$ has finite $(m_0, m_1, \ldots, m_p, m_\beta)$ moments, that is*

$$\int_\beta\int_{\boldsymbol{\theta}}\prod_{i=1}^{n}\left\{\left(\frac{e^{x_i\boldsymbol{\theta}}}{1+e^{x_i\boldsymbol{\theta}}}\right)^{y_i^\beta} - \left(\frac{e^{x_i\boldsymbol{\theta}}}{1+e^{x_i\boldsymbol{\theta}}}\right)^{(y_i+1)^\beta}\right\}\theta_0^{m_0}\ldots\theta_p^{m_p}\beta^{m_\beta}\,d\boldsymbol{\theta}d\beta < \infty.$$

Proof. This proof is similar to the proof of theorem (4.2).

Without loss of generality, we consider $p = 1$. Then, for example in the last case, assuming $\theta_j > 0, j = 0, 1$,

$$\int_\beta \int_{\boldsymbol{\theta}=(\theta_0,\theta_1)>0} \prod_{i=1}^{n} \left\{ \left(\frac{e^{x_i\boldsymbol{\theta}}}{1+e^{x_i\boldsymbol{\theta}}} \right)^{y_i^\beta} - \left(\frac{e^{x_i\boldsymbol{\theta}}}{1+e^{x_i\boldsymbol{\theta}}} \right)^{(y_i+1)^\beta} \right\} \theta_0^{m_0}\theta_1^{m_1}\beta^{m_\beta}\, d\boldsymbol{\theta} d\beta \leq$$

$$\left(\prod_{i=0}^{1} \frac{\Gamma(m_i+1)}{|x_{ki}^{m_i+1)}|} \right) \frac{1}{(m_0+m_1+2)^{m_\beta+1}} \left(\frac{1}{(\log(y_k))^{m_\beta+1}} - \frac{1}{(\log(y_k+1))^{m_\beta+1}} \right).$$

In general, for $p+2$ parameters $(\theta_0, \dots, \theta_p, \beta)$ and corresponding moments $(m_0, \dots, m_p, m_\beta)$ we have,

$$\int_\beta \int_{\boldsymbol{\theta}=(\theta_0,\dots,\theta_p)>0} \prod_{i=1}^{n} \left\{ \left(\frac{e^{x_i\boldsymbol{\theta}}}{1+e^{x_i\boldsymbol{\theta}}} \right)^{y_i^\beta} - \left(\frac{e^{x_i\boldsymbol{\theta}}}{1+e^{x_i\boldsymbol{\theta}}} \right)^{(y_i+1)^\beta} \right\} \theta_0^{m_0}\dots\theta_p^{m_p}\beta^{m_\beta}\, d\boldsymbol{\theta} d\beta \leq$$

$$\left(\prod_{i=0}^{p} \frac{\Gamma(m_i+1)}{|x_{ki}^{m_i+1)}|} \right) \frac{1}{(\sum_{i=0}^{p} m_i+1)^{m_\beta+1}} \left(\frac{1}{(\log(y_k))^{m_\beta+1}} - \frac{1}{(\log(y_k+1))^{m_\beta+1}} \right),$$

which completes the proof. ■

Theorem (4.2) and Theorem (4.3) refer to the model with logit link on q and constant β. In fact, the results apply also to the case of log-log link, given Lemma (4.1). In the next sections, we consider empirical results on simulated and real data using non-informative priors. Of course, any proper prior distribution can also be used when prior information is available. In particular, in the next sections, we consider the case of sparsity and variable selection. In this case, we use Laplace priors as defined in Equation 3.

Given that the parameters q and β of a discrete Weibull distribution are non-location parameters, we also consider the use of a Jeffreys prior [7], as this prior has the property of being invariant under model reparametrization. Taking a $DW(q, \beta)$ distribution, this prior is defined by $p(q, \beta) = \sqrt{|I(q, \beta)|}$, with $I(q, \beta)$ being the Fisher information matrix. This prior has already been studied in the context of the continuous Weibull distribution by [45], who also study its theoretical properties. By exploiting the link between the parameters of the discrete Weibull and those of the continuous Weibull distribution, and considering that the Jeffreys prior for the continuous Weibull distribution in equation (2) is proportional to $\frac{1}{\alpha\beta}$, the corresponding prior for the $DW(q, \beta)$ distribution is given by

$$p(q, \beta) \propto \frac{[-\log(q)]^{1/\beta}}{\beta}.$$

This leads to a natural choice of prior for the regression case with the log-log link, which is also used in continuous Weibull regression models.

5. Simulations study

In this section, we perform a simulation study where we show the effectiveness of the Bayesian estimation procedure, both in the case of data drawn from a DW regression model and in the case of model misspecification, where the generating model is that of Poisson or Negative Binomial (NB). Finally, we test the use of Laplace priors in a variable selection scenario.

5.1 *Simulation from a DW regression model*

Table (1) shows six configurations of parameters used in the simulation, where we consider the two link functions for q and the link function for β described in Section (2), i.e. imposing a linear model on logit(q) or log-log(q), and on log(β). We choose the regression and distribution parameters in such a way to obtain different shapes of the distribution. For cases 2 to 6, we generate the two predictors uniformly in the interval $[-1.5, 1.5]$

Table 1. The configuration of DW regression models used in the simulations.

Case	Model	True Parameters	
1	$DW(q, \beta)$	$q = .41$	$\beta = .75$
2	$DW(q,\ \log : reg\beta)$	$q = .8$	$\gamma_0 = .1$, $\gamma_1 = -.15$, $\gamma_2 = .5$
3	$DW(\text{logit} : regQ, \beta)$	$\theta_0 = .4$, $\theta_1 = -.1$, $\theta_2 = .34$	$\beta = .7$
4	$DW(\text{logit} : regQ,\ \log : reg\beta)$	$\theta_0 = .4$, $\theta_1 = -.1$, $\theta_2 = .34$	$\gamma_0 = .1$, $\gamma_1 = -.15$, $\gamma_2 = .5$
5	$DW(\text{log-log} : regQ, \beta)$	$\theta_0 = .4$, $\theta_1 = -.1$, $\theta_2 = .34$	$\beta = .7$
6	$DW(\text{log-log} : regQ,\ \log : reg\beta)$	$\theta_0 = .4$, $\theta_1 = -.1$, $\theta_2 = .34$	$\gamma_0 = .1$, $\gamma_1 = -.15$, $\gamma_2 = .5$

and we simulate 500 observations. For the Bayesian estimation of the parameters, we use non-informative priors and make use of an adaptive Metropolis-Hastings algorithm with an independent Gaussian proposal to draw samples from the posterior. The scale of the proposal is adjusted so that a recommended acceptance rate lies in the interval $(22, 25)\%$ [6]. We consider 25,000 iterations of the sampler and use the first 25% of the data as burn-in.

Figure (1) shows the posterior distribution of the parameters and the chain convergence in the first case, when no exogenous variables are present. Similar plots are obtained for the other cases. Table 2 reports also the coverage of 95% confidence and credible intervals. For the Bayesian approach, we consider both the uniform and Jeffreys non-informative priors. Figure (2) shows the marginal densities of the parameters and the 95% HPD interval for all six cases, as well as the maximum likelihood point estimate and the true value of the parameters. Overall, the plots show convergence of the chain and accurate estimation of the parameters. Table 2 shows how the coverage obtained when using a Jeffreys prior is closer to the nominal level than when using a uniform prior, and is in line with the coverage of the frequentist approach.

Table 2. Parameter estimates for case 1 from likelihood and Bayesian approaches, under uniform and Jeffreys priors. The last row reports the coverage of 95% confidence and credible intervals.

	Frequentist		**Bayesian (Uniform)**		**Bayesian (Jeffreys)**	
	$q = .41$	$\beta = 0.75$	$q = .41$	$\beta = 0.75$	$q = .41$	$\beta = 0.75$
Estimate	0.411	0.761	0.412	0.758	0.418	0.772
Standard Error	0.042	0.081	0.042	0.080	0.046	0.083
95% Coverage	98%	96%	100%	96%	98%	97%

5.2 *Simulation from a Poisson and NB regression model*

The aim of this section is to test the fitting of a DW regression model to data generated from a Poisson and NB regression. To this end, we design two experiments using two explanatory variables, $X = (X_1, X_2)$, and $n = 500$ data points. We simulate data for the predictors from uniform distributions, namely $X_1 \sim U(0, 1)$ and $X_2 \sim U(0, 1.5)$. In the first experiment, we assume that the conditional distribution of Y given X is Poisson($e^{X\alpha}$), whereas in the second experiment, we assume it to be a NB distribution with mean $\mu = e^{X\alpha}$ and variance $\mu + \mu^2/\theta$ with $\theta = 4.5$. We fix the intercept and the regression parameters to $\alpha = (-0.5, 4.3, -2.2)$, with values chosen to cover a wide

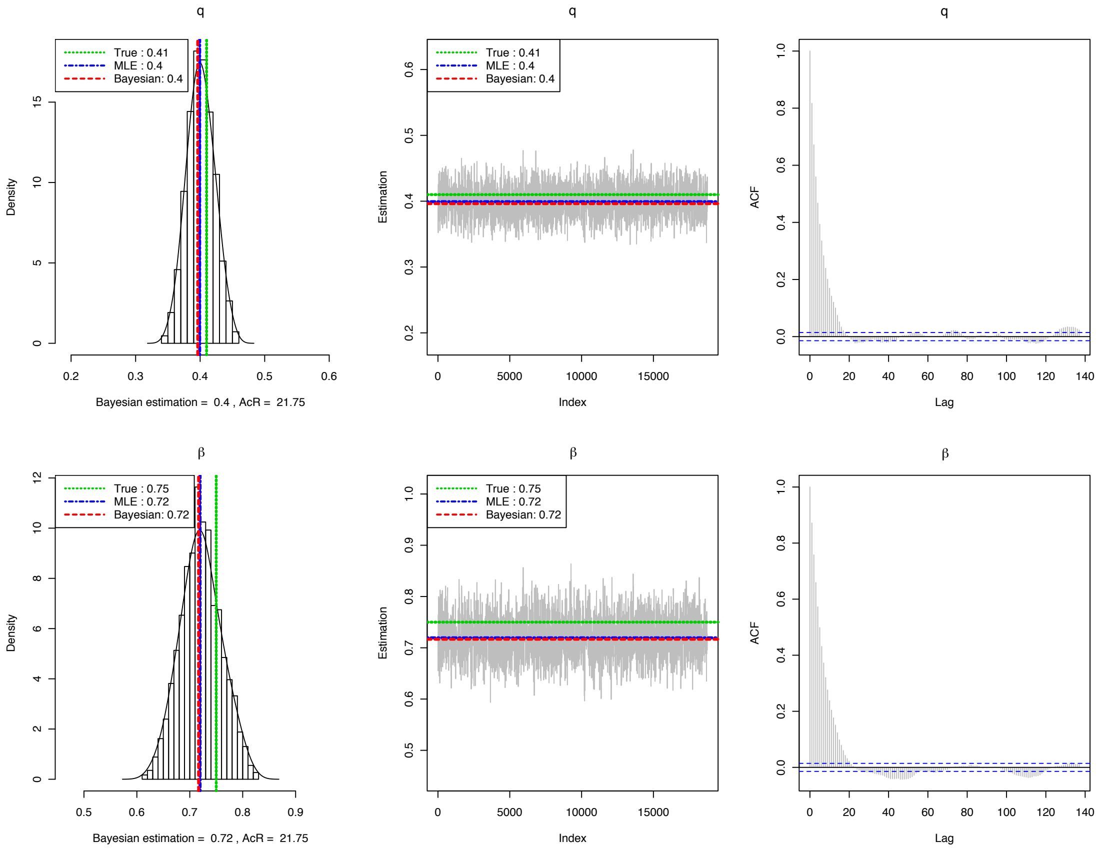

Figure 1. Marginal densities and chain convergence for q (top) and β (bottom), for case 1 where there are no exogenous variables in model.

range of shapes for the target distribution. Figures (3) shows the conditional posterior distribution fitted by $DW(regQ, \beta)$ for a fixed value of $x_1 = 0.5$ and sliding values of x_2 in the $[0, 0.7]$ interval. The figure shows overall a good fitting, both for Bayesian and frequentist approaches, and a better fit when the logit link is used compared to the log-log link in both Poisson and NB experiments. For the frequentist estimation, we use the R package `DWreg` [24].

5.3 *Simulation on Variable Selection*

In this simulation, we show the performance of DW regression for variable selection. To this end, we consider a simulation with 50 predictors and assume that 75% of the parameters, 37 out of 50, are zero. We generate the remaining non-zero parameters uniformly in the $[-0.5, 0.5]$ interval. We simulate 500 observations for each predictor from a $U(0, 1.5)$ distribution, and the response variable from a DW distribution using a logit link for q or the log link for β. Similar results are obtained with the log-log link function. For parameter estimation, the number of iterations is set to 25000 iterations and an InverseGamma(2,1) hyper prior is chosen for the penalty parameters. This prior allows to cover a large range of penalties, in the (0.02,70) interval, with a tendency to small penalties (i.e. sparsity) due to a mean of 1 and a median of 0.5. Variable selection is performed by considering the 95% HPD interval for each parameter.

Table (3) shows the performance of the method in terms of selection of variables under six different generating models. In particular, the table reports the True Negative Rate

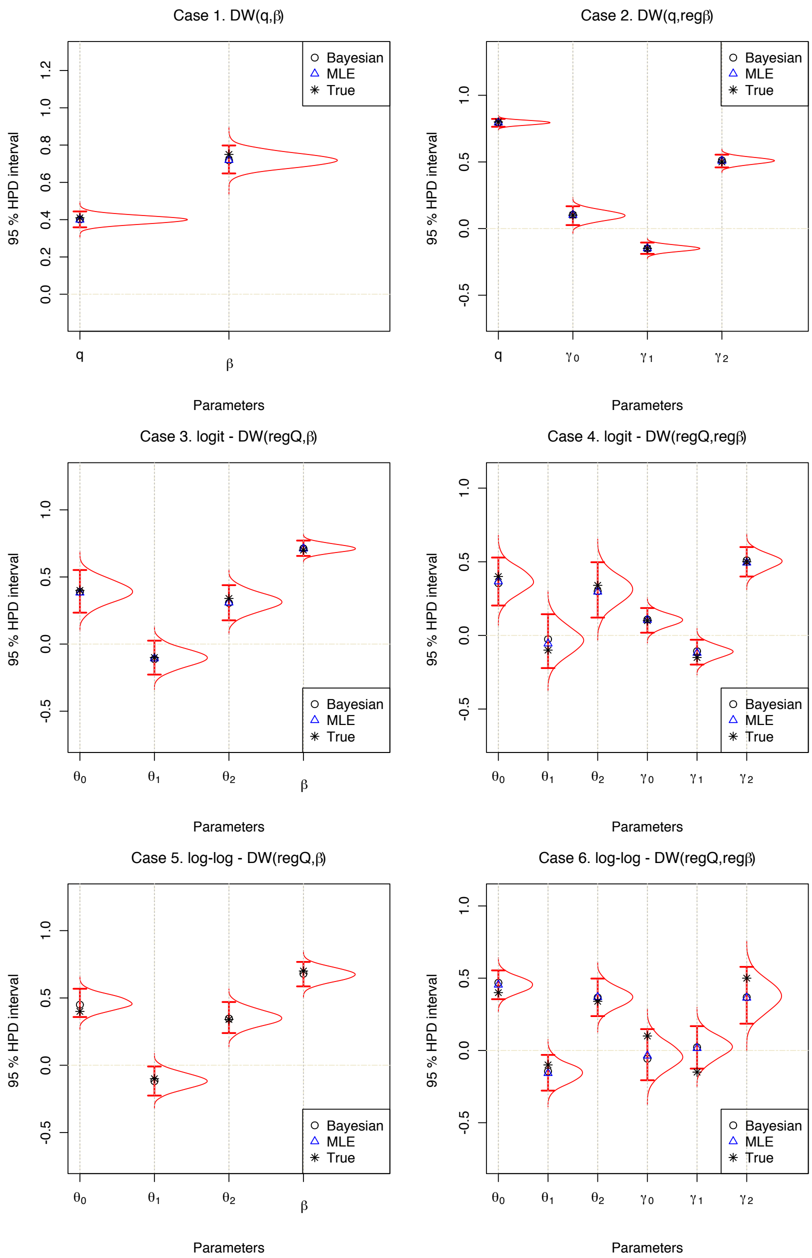

Figure 2. Marginal densities and 95% high probability density interval for cases 1-6 in Table (1).

(TNR), Recall $\left(\dfrac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}\right)$, Precision $\left(\dfrac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}\right)$ and F_1 score $\left(\dfrac{2\mathrm{TP}}{2\mathrm{TP}+\mathrm{FN}+\mathrm{FP}}\right)$, averaged over 20 simulations. The table shows a good performance overall, particularly for the $BDW(regQ, \beta)$ models. The model with the $\log(\beta)$ link does not perform very well when q decreases, i.e. when the number of zeros in the sample increases. In these cases, the models show a low recall, that is a high false negative rate.

6. Analysis of counts in medicine

In this section, we show the performance of the Bayesian discrete Weibull regression model on real datasets from the medical domain. We compare the proposed model with the Bayesian Poisson (BPoisson) and Bayesian Negative Binomial (BNB) mod-

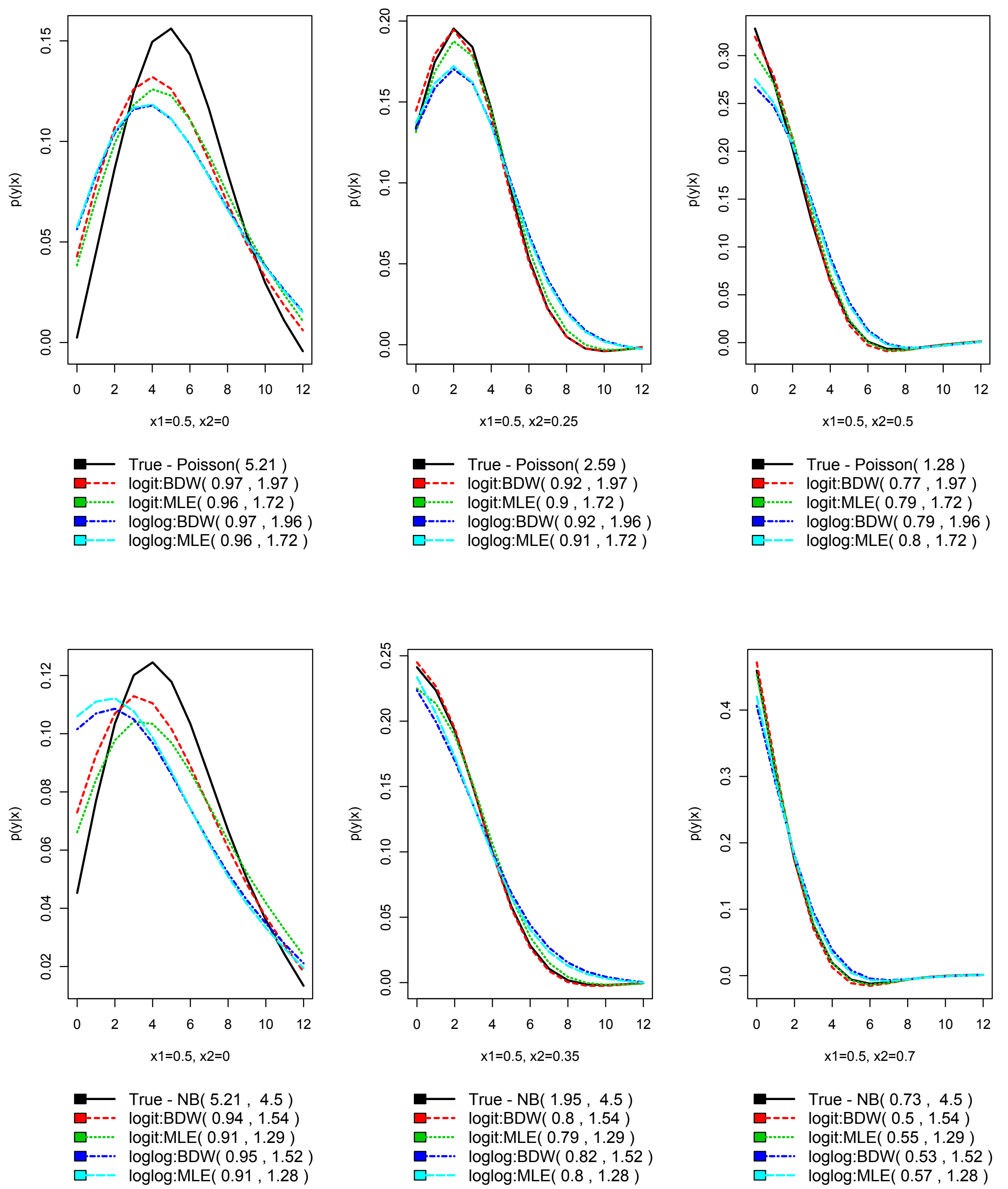

Figure 3. Fitting Poisson (top) and NB (bottom) simulated data by $DW(regQ, \beta)$ for a range of values of x_2 and fixed $x_1 = 0.5$. The plots show the true conditional pmf (black) together with the conditional pmf fitted by the Bayesian DW model proposed in this paper, with the logit(q) (red) and log-log(q) (blue) links, and by the corresponding frequentist approaches (green and light blue, respectively).

els on the basis of a number of commonly used criteria: Bayesian Information Criteria (BIC) [12], Akaike Information Criteria (AIC) [12], Deviance Information Criterion (DIC) [47], Quasi-likelihood Information Criteria (QIC) [39], Consistent AIC (CAIC) [8], Bayesian Predictive Information Criterion (BPIC) [1] and the Prior Predictive Density (PPD) used in the Bayes factor [25].

Table 3. Performance of BDW with Laplace priors. Variables are selected based on the 95% HPD interval and the selection is compared with the truth on the basis of True Negative Rate (TNR), recall, precision and F_1 score.

Model	TNR	Recall	Precision	F_1
$BDW(regQ, \beta = .1)$	93%	90%	93%	91%
$BDW(regQ, \beta = .8)$	95%	89%	95%	92%
$BDW(regQ, \beta = 1.6)$	93%	91%	93%	92%
$BDW(regQ, reg\beta)$	97%	68%	96%	79%
$BDW(q = .85, reg\beta)$	90%	92%	91%	91%
$BDW(q = .50, reg\beta)$	93%	37%	84%	52%

6.1 *Comparison with Bayesian generalised linear models*

[24] show how the discrete Weibull distribution is able to capture different levels of dispersion in the data, as varying q and β can lead both to cases of over and under-dispersion. In this section, we show the ability of BDW to estimate parameters in three different scenarios, represented by the following three medical datasets:

(1) The data on inhaler usage from [19], with 5209 observations. The response is the daily counts of inhalers usage, whereas the covariates are humidity, barometric pressure, daily temperature, air particles level. The sample mean and variance of the data are 1.3 and 0.8 respectively, so this is a case of under-dispersion relative to Poisson [24].
(2) The German health survey dataset available in the R package `COUNT` under the name `badhealth`, with 1127 observations. The response is the number of visits to doctors during the year 1988 and the predictors are whether the patient claims to be in bad health or not, and the age of the patient. The response variable ranges from 0 to 40 visits and has a sample mean of 2.4 and variance of 12, suggesting over-dispersion relative to Poisson.
(3) The German health registry dataset available in the R package `COUNT` under the name `rwm`, with 27326 observations. The response is the number of visits to doctors for the years 1984-1988 and the predictors are age, years of education and household yearly income. The response variable, number of visits, has about 37% of zeros, a sample mean of 3.2 and a variance of 32.4, pointing again to a case of over-dispersion and excessive zeros.

We fit a BDW model with a non-informative prior on the regression parameters, 25,000 iterations for the adaptive Metropolis-Hastings sampler and we check that the acceptance rate lies in the recommended $(20, 30)$% interval [6]. For the case of BPoisson, BNB regression, zero-inflated Poisson (ZIP) and zero-inflated Negative Binomial (ZINB) regression, we make use of the `MCMCpack` R package [33] using the same configurations as with our approach. Table (4) shows a comparison of the models on the three datasets. We only report the results of the BDW(regQ,β) models, which show superior performance to the other BDW models on these datasets. Of the two links on q, the logit(q) link performs better than the log-log(q) link. As for a comparison with the other models, Poisson has the worst performance for all cases, while NB has a performance comparable to the logit-BDW model in the overdispersed scenario, while it does not perform well in the underdispersed and excessive zero scenario. In the latter case, zero-inflated negative Binomial has a performance comparable to DW. This is promising and it points to a future extension of DW to a zero-inflated DW model.

6.2 *Comparison with Bayesian penalised regression*

In this section, we compare the performance of BDW to BPoisson and BNB regression for variable selection on a dataset with several variables. In particular, we consider the multivariate data of [32]. The data consist of 5096 observations from the 1985 wave of the German Socioeconomic Panel. As in [32], we measure the demand in healthcare by the number of visits to a specialist (except gynecology or pedriatics) in the last quarter. The 20 covariates are listed in full in Table 5 and they are the same considered in [32]. This is an extreme example of excessive zeros as the response variable contains 67.82% of zeros.

We fit a BDW model with a Laplace prior on the regression parameters and an InverseGamma(2,1) hyper-prior on the shrinkage parameters. We consider 175000 iterations for the MCMC routine and similar configurations for the Bayesian Poisson and NB models. We also extend the comparison by including Bayesian zero-inflated models and frequentist L_1 regularized models. For the latter, we use the `glmnet` package [17] to fit regularized Poisson regression and the `glm.nb` R function to fit regularized negative Binomial regression. In both cases, the penalty parameter is chosen by BIC. According to the results in Table (6), $DW(regQ, \beta)$ with the log-log link achieves overall the best performance compared with the others BDW models and with NB and Poisson models.

Figure (4) shows the marginal densities of the parameters for the $DW(regQ, \beta)$ with the log-log link. Highlighted in red are those variables that are found to be significant based on the 95% HPD interval. The selection is overall in accordance with the results obtained by [32] using a jittering approach, with variables such as gender, chronic complaints, sick leave and disability found to be significant, and other variables like unemployment, private insurance and those related to job characteristics, such as heavy labor, stress, variety on job, self-determined and control found not to be significant. Figure 5 shows the effect of the variable chronic complaints on the conditional distribution, suggesting that the probability of a large number of visits is higher for the case of chronic complaints than for the case of no complaints. Table (7) further compares the selection of variables with those selected by Poisson and NB regression models. Overall, there is high agreement between DW and NB, with the exception of the variable control which is found significant by NB (both in the Bayesian and frequentist estimation) but not by DW. Poisson and BPoisson tend to select many more variables.

7. Conclusion

In this paper we have proposed a novel Bayesian regression model for count data, by assuming a discrete Weibull conditional distribution. A discrete Weibull regression model was originally prosed by [24] in a frequentist context and a number of desirable features of the model compared to existing ones were highlighted. The Bayesian implementation in this paper is based on a more general model, where both parameters can be linked to the predictors. We have experimented with different link functions and have found the models with the link on q and constant β to work particularly well, with the logit(q) link displaying superior performance than the log-log(q) link in the simulations and in three of the four applications considered in this paper. Including a link to both q and β was found to be too complex for the applications considered, but it may be useful for other applications showing more complex dependencies and where more data are available. In terms of the Bayesian inferential approach, we have shown theoretically how the posterior is proper and with finite moments under a uniform non-informative prior on the parameters.

We have shown the applicability of the Bayesian discrete Weibull model to count data from the medical domain. In particular, we have analysed datasets on the number of visits to doctors/specialists, a quantity that is often used as an indicator of healthcare demand. The response variable in the examples considered is discrete and is characterized

Table 4. Comparison of Bayesian DW, Poisson and Negative Binomial on three datasets and under a number of information criteria. (*) denotes the minimum value. For the dataset with excessive zeros, the comparison includes also zero-inflated models.

Model	AIC	BIC	CAIC	QIC	DIC	BPIC	log(PPD)	df
Inhaler Use (under-dispersed)								
log-log : BDW	13497.22	13536.57	13542.57	2.59*	13487.63	13493.88	-6745.93	6
logit : BDW	13494.19*	13533.54*	13539.54*	2.59*	13484.92*	13490.49*	-6739.41*	6
BPoisson	14009.01	14041.80	14046.80	2.69	13822.54	13734.31	-6960.64	5
BNB	13952.85	13992.33	13998.20	2.68	13771.0	13686.47	-6960.81	6
German Health Survey (over-dispersed)								
log-log : BDW	4478.9	4499.0	4502.0	3.98	4474.60	4478.33	-2245.75	4
logit : BDW	4475.2*	4495.3*	4449.3*	3.97*	4474.16*	4477.70*	-2242.23*	4
BPoisson	5638.9	5654.02	5656.10	5.01	5638.14	5641.18	-2826.88	3
BNB	4475.9	4495.9	4499.97	3.97*	4474.66	4478.10	-2243.87	4
German Health Registry (excessive zeros)								
log-log : BDW	120340.1	120381.2	120386.2	4.4*	120334.6	120339.2	-60187.6	5
logit : BDW	120339.2*	120380.3*	120385.3*	4.4*	120327.0*	120331.9*	-60181.8*	5
BPoisson	209636.4	209669.2	209673.2	7.7	209635.8	209639.6	-104836.7	4
BNB	120658.7	120708.0	120714.0	4.4*	129125.8	133365.3	-60344.0	5
BZIP	169417.7	169450.6	169454.6	6.2	169402.1	169398.3	-83522.3	5
BZINB	120649.5	120682.4	120686.4	4.4*	120629.1	120622.9	-60245.2	6

Table 5. List of the variables and descriptions in *the number of visits to a specialist* dataset [32].

Variable	Description
Age	Age in decades
Chronic complaints	1 if has chronic complaints for at least 1 year
Control	1 if has a job where work performance is strictly controlled
Degree of disability $>$ 20%	1 if the degree of disability is greater than 20%
Education	Number of years in education after age 16
HH-income	Net monthly household income
Hospitalized $>$ 7 days	1 if was more than 7 days hospitalized in the previous year
Marital Status	1 if single
Month of unemployment	Number of months of unemployment in the previous year
Physically heavy labour	1 if has a job in which physically heavy labour is required
Physician density	Number of physicians per 100,000 inhabitants in the place of residence
Population $<$ 5000	1 if place of residence has less than 5,000 inhabitants
Population 20000-100000	1 if place of residence has between 20,000 and 100,000 inhabitants
Population 5000-20000	1 if place of residence has between 5,000 and 20,000 inhabitants
Private insurance	1 if had private medical insurance in the previous year
Self-determined	1 if has a job where the individual can plan and carry out job tasks
Sex	1 if female
Sick leave $>$ 14 days	1 if missed more than 14 work days due to illness in the previous year
Stress	1 if has a job with high level of stress
Variety on job	1 if job offers a lot of variety

Table 6. Comparison of BDW with Bayesian and regularized NB and Poisson on the number of visits to a specialist dataset of [32]. ($*$) denotes the minimum value, whereas df is the number of non-zero coefficients. For the Bayesian models, these are based on the 95% HPD interval.

Model	AIC	BIC	CAIC	QIC	DIC	BPIC	log(PPD)	df
logit:BDW(regQ,β)	12720.4	12864.2	2.5*	12886.2	12710.8	12731.5	-6392.3	11
log-log:BDW(regQ,β)	12698.5*	12842.3*	2.5*	12864.3*	12693.3*	12713.6*	-6383.3*	11
BDW(q,regβ)	13256.0	13399.8	2.6	13421.8	13250.4	13270.3	-6665.8	6
logit:BDW(regQ,regβ)	12750.3	12920.7	2.5*	12951.7	12713.0	12744.9	-6516.3	19
log-log:BDW(regQ,regβ)	12748.9	12924.1	2.5*	12955.2	12715.4	12741.3	-6519.1	19
BPoisson	21588.2	21705.8	4.2	21723.8	21594.6	21615.8	-10832.6	17
BNB	12867.3	12939.2	2.5*	12950.2	12838.3	12834.8	-6452.3	11
BZIP	16677.1	16760.1	3.2	16760.0	16698.7	16720.6	-8385.6	15
BZINB	12850.0	12921.1	2.5*	12932.2	12872.0	12894.1	-6456.8	13
Poisson (glmnet)	21571.1	21706.1	4.2	21724.1	-	-	-	17
NB (glm.nb)	12839.3	12911.2	2.5*	12922.6	-	-	-	12

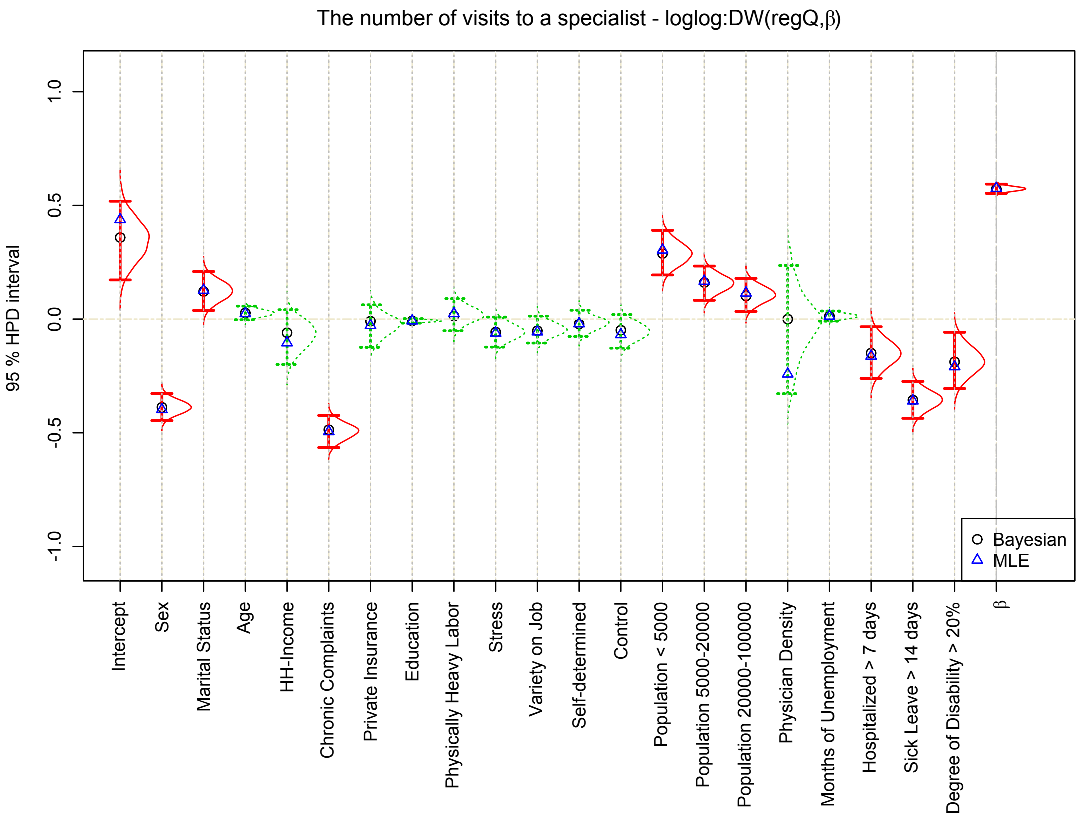

Figure 4. Marginal densities of the parameters for the $BDW(regQ,\beta)$ model with log-log(q) link on the number of visits to a specialist dataset. The red lines are for the cases where the 95% HDP interval does not contain zero (significant variable). Green dotted lines for the opposite.

Table 7. Significant covariates that are selected by $BDW(regQ,\beta)$ with log-log link, Bayesian and regularized NB and Poisson regression models, and Bayesian zero-inflated Poisson and NB, for the number of visits to a specialist dataset. An $(*)$ indicates a non-zero coefficient.

Variable	**BDW(regQ,β)**	**NB**	**BNB**	**BZINB**	**Poisson**	**BPoisson**	**BZIP**
Sex	*	*	*	*	*	*	*
Marital status	*	*	*		*	*	*
Age					*		
HH-income					*	*	
Chronic complaints	*	*	*	*	*	*	*
Private insurance							
Education					*		*
Physically heavy labour					*	*	*
Stress					*	*	*
Variety on job					*	*	*
Self-determined							
Control		*	*	*	*	*	*
Population $<$ 5000	*	*	*	*	*	*	*
Population 5000-20000	*	*	*	*	*	*	
Population 20000-100000	*	*		*	*	*	*
Physician density						*	
Months of unemployment			*	*		*	
Hospitalized $>$ 7 days	*	*	*	*	*	*	*
Sick Leave $>$ 14 days	*	*	*	*	*	*	*
Degree of disability $>$ 20	*	*	*	*	*	*	

by a skewed distribution, making the whole conditional distribution of interest and not only the conditional mean. We have tested the inference procedure on simulated and real

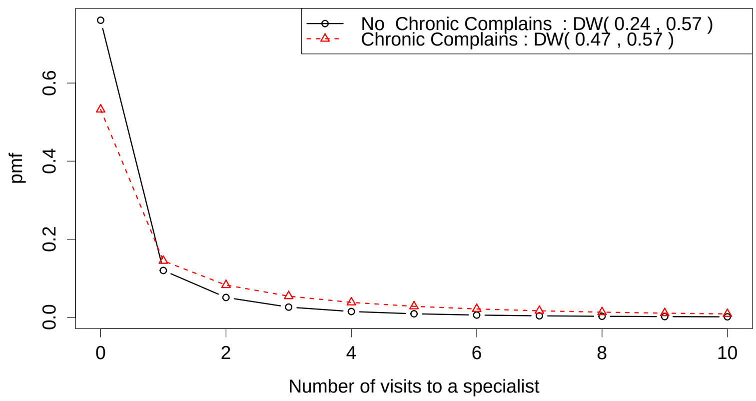

Figure 5. Effect of the variable Chronic Complaints on the conditional distribution for the healthcare data, when all other variables are held constant.

data with various characteristics, such as under-dispersion, over-dispersion and excess of zeros. Overall, we have found a good performance of the method in comparison with Poisson and NB regression models, on the basis of a number of information criteria and of the selection of influential variables.

The method is implemented in the R package `BDWreg`, which is available in CRAN. Future work will explore an extension of the approach proposed in this paper to more flexible DW regression models, such as zero-inflated, multilevel and mixture DW models, in a similar spirit to the existing models for continuous responses [13].

References

[1] T. Ando, *Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models*, Biometrika 94 (2007), pp. 443–458.

[2] J.F. Angers and A. Biswas, *A Bayesian analysis of zero-inflated generalized Poisson model*, Computational Statistics and Data Analysis 42 (2003), pp. 37–46.

[3] P. Araújo Santos and M.I. Fraga Alves, *Improved shape parameter estimation in a discrete Weibull model*, in *Recent Developments in Modeling and Applications in Statistics . Studies in Theoretical and Applied Statistics.*, Springer-Verlag, 2013, pp. 71–80.

[4] N. Atienza, J. Garcia-Heras, J.M. Munoz-Pichardo, and R. Villa, *An application of mixture distributions in modelization of length of hospital stay*, Statistics in Medicine 27 (2008), pp. 1403–1420.

[5] Y. Bao, V. Vinciotti, E. Wit, and P. 't Hoen, *Joint modeling of ChIP-seq data via a Markov random field model*, Biostatistics 15 (2014), pp. 296–310.

[6] M. Bedard, *Optimal acceptance rates for metropolis algorithms: Moving beyond 0.234*, Stochastic Processes and their Applications 118 (2008), pp. 2198–2222.

[7] J.O. Berger, J.M. Bernardo, and D. Sun, *Overall objective priors*, Bayesian Analysis 10 (2015), pp. 189–221.

[8] H. Bozdogan, *Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions*, Psychometrika 52 (1987), pp. 345–370.

[9] C. Bracquemond and O. Gaudoin, *A survey on discrete lifetime distributions*, International Journal of Reliability, Quality and Safety Engineering 10 (2003), pp. 69–98.

[10] A.C. Cameron and P.K. Trivedi, *Regression analysis of count data*, Cambridge university press, 2013.

[11] E. Carter and H. Potts, *Predicting length of stay from an electronic patient record system: a primary total knee replacement example*, BMC Medical Informatics and Decision Making 14 (2014).

[12] C.M. Dayton, *Model comparisons using information measures*, Journal of Modern Applied Statistical Methods 2 (2003), p. 2.

[13] D.B. Dunson, N. Pillai, and J.H. Park, *Bayesian density regression*, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69 (2007), pp. 163–183.

[14] G. El-Sayyad, *Bayesian and classical analysis of Poisson regression*, Journal of the Royal Statistical Society. Series B (Methodological) (1973), pp. 445–451.

[15] J.D. Englehardt and R. Li, *The discrete Weibull distribution: An alternative for correlated counts with confirmation for microbial counts in water distributions*, Risk Analysis 31 (2011), pp. 370–381.

[16] M. Esnaola, P. Puig, D. Gonzalez, R. Castelo, and J.R. Gonzalez, *A flexible count data model to fit the wide diversity of expression profiles arising from extensively replicated RNA-seq experiments*, BMC Bioinformatics 14 (2013), p. 254.

[17] J. Friedman, T. Hastie, and R. Tibshirani, *Regularization paths for generalized linear models via coordinate descent*, Journal of Statistical Software 33 (2010), p. 1.

[18] S.K. Ghosh, P. Mukhopadhyay, and J.C.J. Lu, *Bayesian analysis of zero-inflated regression models*, Journal of Statistical planning and Inference 136 (2006), pp. 1360–1375.

[19] G.K. Grunwald, S.L. Bruce, L. Jiang, M. Strand, and N. Rabinovitch, *A statistical model for under- or overdispersed clustered and longitudinal count data*, Biometrical Journal 53 (2011), pp. 578–594.

[20] H. Haario, E. Saksman, and J. Tamminen, *An adaptive metropolis algorithm*, Bernoulli (2001), pp. 223–242.

[21] W.K. Hastings, *Monte Carlo sampling methods using Markov chains and their applications*, Biometrika 57 (1970), pp. 97–109.

[22] P. Hougaard, M.T. Lee, and G.A. Whitmore, *Analysis of overdispersed count data by mixtures of Poisson variables and Poisson processes*, Biometrics 53 (1997), pp. 1225–1238.

[23] H. Ishwaran and J.S. Rao, *Spike and slab variable selection: frequentist and Bayesian strategies*, Annals of Statistics (2005), pp. 730–773.

[24] H.S. Kalktawi, V. Vinciotti, and K. Yu, *A Simple and Adaptive Dispersion Regression Model for Count Data*, arXiv:1511.00634 (2016).

[25] R.E. Kass, *Bayes factors in practice*, The Statistician (1993), pp. 551–560.

[26] M.S.A. Khan, A. Khalique, and A.M. Abouammoth, *On estimating parameters in a discrete Weibull distribution*, IEEE transactions on Reliability 38 (1989), pp. 348–350.

[27] K.B. Kulasekera, *Approximate MLE's of the parameters of a discrete Weibull distribution with type I censored data*, Microelectronics Reliability 34 (1994), pp. 1185–1188.

[28] M. Kyung, J. Gill, M. Ghosh, G. Casella, *et al.*, *Penalized regression, standard errors, and Bayesian lassos*, Bayesian Analysis 5 (2010), pp. 369–411.

[29] C.D. Lai, *Issues concerning constructions of discrete lifetime models*, Qualitative technology of quantitative managment 10 (2013), pp. 251–262.

[30] K.F. Lam, H. Xue, and Y. Bun Cheung, *Semiparametric analysis of zero-inflated count data*, Biometrics 62 (2006), pp. 996–1003.

[31] H. Liu and D.A. Powers, *Bayesian inference for zero-inflated Poisson regression models*, Journal of Statistics: Advances in Theory and Applications 7 (2012), pp. 155–188.

[32] J. Machado and M. Santos Silva, *Quantiles for counts*, JASA 100 (2005), pp. 1226–1237.

[33] A.D. Martin, K.M. Quinn, J.H. Park, *et al.*, *MCMCpack: Markov chain Monte Carlo in R*, Journal of Statistical Software 42 (2011), pp. 1–21.

[34] M. Mohebbi, R. Wolfe, and A. Forbes, *Disease mapping and regression with count data in the presence of overdispersion and spatial autocorrelation: a Bayesian model averaging approach*, International journal of environmental research and public health 11 (2014), pp. 883–902.

[35] T. Nagakawa and S. Osaki, *The discrete Weibull distribution*, IEEE transactions on reliability R-24 (1975).

[36] B.H. Neelon, A.J. O'Malley, and S.L.T. Normand, *A Bayesian model for repeated measures zero-inflated count data with application to outpatient psychiatric service use*, Statistical Modelling 10 (2010), pp. 421–439.

[37] P. Newcombe, H.R. Ali, F. Blows, E. Provenzano, P. Pharoah, C. Caldas, and S. Richardson, *Weibull regression with Bayesian variable selection to identify prognostic tumour markers of breast cancer survival*, Statistical methods in medical research (2014), p. 0962280214548748.

[38] F. Ozsolak and P.M. Milos, *RNA sequencing: advances, challenges and opportunities.*, Nature Review Genetics 12 (2011), pp. 87–98.

[39] W. Pan, *Akaike's information criterion in generalized estimating equations*, Biometrics 57 (2001), pp. 120–125.

[40] T. Park and G. Casella, *The Bayesian lasso*, Journal of the American Statistical Association 103 (2008), pp. 681–686.

[41] A. Polpo, M. Coque Jr, and C. Pereira, *Statistical analysis for Weibull distributions in presence of right and left censoring*, in *Reliability, Maintainability and Safety, 2009. ICRMS 2009. 8th International Conference on*, 2009, pp. 219–223.

[42] R.A. Rigby and D.M. Stasinopoulos, *Generalized additive models for location, scale and shape,(with discussion)*, Applied Statistics 54 (2005), pp. 507–554.

[43] M.D. Robinson and G.K. Smyth, *Small-sample estimation of negative Binomial dispersion, with applications to SAGE data.*, Biostatistics 9 (2008), pp. 321–332.
[44] K.F. Sellers and G. Shmueli, *A flexible regression model for count data*, The Annals of Applied Statistics 4 (2010), pp. 943–961.
[45] S.K. Sinha and J.A. Sloan, *Bayes estimation of the parameters and reliability function of the 3-parameter Weibull distribution*, IEEE Transactions on Reliability 37 (1988), pp. 364–369.
[46] A.A. Soliman, A.H. Abd-Ellah, N.A. Abou-Elheggag, and E.A. Ahmed, *Modified Weibull model: A Bayes study using MCMC approach based on progressive censoring data*, Reliability Engineering & System Safety 100 (2012), pp. 48–57.
[47] D.J. Spiegelhalter, N.G. Best, B.P. Carlin, and A. Van Der Linde, *Bayesian measures of model complexity and fit*, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64 (2002), pp. 583–639.
[48] R. Tibshirani, *Regression shrinkage and selection via the lasso*, Journal of the Royal Statistical Society. Series B (Methodological) (1996), pp. 267–288.
[49] M. Zhou, L. Li, D. Dunson, and L. Carin, *Lognormal and gamma mixed negative binomial regression*, in *Proceedings of the 29th International Conference on Machine Learning*, Vol. 2012, 2012, p. 1343.