Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.7, No.9, 2017
www.iiste.org

Bayesian Regression Model for Counts in Scholarship

Adesina, O.S [1*], Agunbiade, D.A [1], Osundina, S.A [2]

1. Department of Mathematical Sciences, Olabisi Onabanjo University, Ogun State, Nigeria
2. Department of Basic Sciences, Babcock University, Ogun State, Nigeria
*Email of the corresponding author: olumidestats@gmail.com

Abstract

Discrete Weibul (DW) is considered to have the ability to capture under and over-dispersion simultaneously and also have a closed-form analytical expression of the quantiles of the conditional distribution. There is a need to further investigate how effective the model is, as compared to other competing models in the context of classical and Bayesian technique. In this study, the strength of DW is investigated, for both on frequentist and Bayesian technique. The Bayesian DW adopts parameterization, which makes both parameters of the discrete Weibull distribution to be dependent on the predictors. Bayesian Generalized linear mixed model is also implemented and is compared with the BDW, since Bayesian generalized linear mixed model is known to be robust in handling over-dispersion in count data. A simulation study and real life data was carried out for over and under-dispersed count data. The empirical analysis shows the superiority of Bayesian Generalized linear mixed model over Bayesian DW in the case of over-dispersed data as identified in the simulated data and real life data, but not for under-dispersed data as in the case of simulated study.

Keywords: Discrete Weibull, Bayesian Statistics, MCMCGlmm, over-dispersion, under-dispersion

1. Introduction

There are quite a number of models for fitting count data, Poisson model is found to be first of such models, but Poisson model is found deficient because of the problem of over and under-dispersion, some other models for count data have been developed which are improvement on the Poisson model and they include but not limited to negative binomial, COM-Poisson, Zero inflated family, Discrete Weibull, and hurdle model.

In estimating the parameters of count models, Quasi-maximum likelihood estimate or Poisson maximum likelihood is considered the most popular method for estimating count data with the following reasons as; it is us known to give convenient or satisfactory results, it is computational simple and easily found in many software packages, it has reasonable robust properties, also, it is recommended when doubt exist about the form of the variance function, (Cameron and Trivedi 2005).

Recent studies have adopted Bayesian techniques for fitting count data and it is found efficient in estimating count data. Bayesian approaches deals with complex models that lack analytically tractable likelihood functions, and the procedures are flexible to be adapted to produce estimates that are excellent and perfect substitutes of maximum likelihood estimates (Cameron and Trivedi 2005). The likelihood, frequentist or classical approach requires probabilistic model specification of prior beliefs about unknown parameters, provided a model has been initially specified. The frequentist inferences about the parameter require probabilities calculated from the sampling distribution of the data, given the fixed but unknown parameter. With Bayesian approach, the prior information (probabilistically specified information before the current data are analysed or based on received information) will be combined using Bayes' theorem, the outcome which gives posterior distribution of the parameters, say θ. Bayesian statistics gives a complete inference on the posterior distribution of the parameter given the actual data that occurred, hence, Bayesian estimate is calculated from the posterior distribution, and therefore estimate or the credible interval depends on the data that actually occurred (Bolstad 2007).

In the study carried out by Haselimashhadi *et. al.*, (2016), the author recognise that using discrete Weibull distribution for modelling count gives a promising results as compared to traditional Poisson and Negative Binomial distributions and their extensions, such as Poisson mixtures, Tweedie, zero-inflated regression and COM-Poisson distribution by (Sellers and Shmueli 2010). Discrete Weibull distribution can capture over and under-dispersion simultaneously and a give closed-form analytical expression of the quantiles of the conditional distribution. In describing COM-Poisson model, (Chanialidis 2015) states that it is flexible enough to handle any kind of dispersion but the key reason why the COM-Poisson distribution not practically used as much is that its normalization constant is not available in a closed from, therefore making approximation to it either computationally inefficient or not sufficiently exact.

Saengthong, *et. al.,* (2015) proposed the zero inflated negative binomial-Crack (ZINB-CR) distribution which is a mixture of Bernoulli distribution and negative binomial-Crack (NB-CR) distribution, as an alternative distribution for the excessive zero counts and over-dispersion. Klein et. al., (2015) proposed a general class of Bayesian generalized additive models for zero-inflated and over-dispersed data within the framework of GAMLSS, the authors developed Bayesian inference based on MCMC simulation technique and it was applied to claim frequencies in car insurances. Murat *et. al.,* (2015) carried out a study to compare Bayesian approach for zero-inflated and the classic zero-inflated Poisson (ZIP) models, based on parameter estimation and information criteria. The result showed that the Bayesian ZIP model suggests a much more improved fit over classic ZIP model. Haselimashhadi *et. al.,* (2016) proposed Bayesian implementation on discrete Weibull regression model by implementing the model on parameterization, where both parameters of the discrete Weibull distribution can be made dependent on the predictors. The authors introduced a logit link for dicrete Weilbul distribution, and drew comparison between discrete Weibul log link and logit link on a number of information criteria.

This study seeks to add to literature by providing Bayesian approach for parameter estimation in discrete Weibull regression along with Bayesian Multivariate Generalised Linear Mixed Model of Poisson distribution. For the choice of prior distributions, non-informative prior is considered, a recent study where non-informative prior is used is Chandra and Rathaur (2017) based on (Jeffery and uniform prior distribution). For the Bayesian Multivariate Generalised Linear Mixed Model based on Poisson distribution, inverse-Wishart family of prior distributions is used, which is a multivariate generalization of the scaled inverse-Chi-square (Gelman 2013).

This study aims first at identifying how Bayesian models fit data well as compared to frequentist models for count; in the case of under-dispersed, over-dispersed and excess zeros data. Second is to compare Bayesian and Frequentist using number of scholarly journal articles published by University lecturers as response variable for both over-dispersed and excess zeros. Third and lastly, is to predict the relationship covariates have with the response variable (numbers of publications).

The remaining part of this paper is organized as follows. Section two describes the inference on Bayesian discrete Weibull regression model. Section three describes the inference on Bayesian Multivariate Generalised Linear Mixed Model. In Section four, a simulation study for both over and under-dispersion is examined, whereas Section five shows the analysis of real data via Bayesian regression models and a comparison was made with existing approaches. Finally, in section six conclusions were drawn based on results obtained.

2. Inference for Discrete Weibull Regression

2.1 Discrete Weibull Distribution

Nagakawa and Osaki (1975) introduced the discrete Weibull distribution as a discretized form of a continuous Weibull distribution, just like the geometric distribution is the discretized form of the exponential distribution. The discrete Weibull distribution in the context of this study, is referred to as a type I discrete Weibull. Bracquemond and Gaudoin (2003) outlined the advantages of type I over the type II and III, type I has an unbounded support as compared to type II, also type I has a more straightforward interpretation as compared to type III.

If a random variable Y follows a discrete Weibull distribution of type I, then the cumulative distribution function of Y is defined as:

$$\begin{cases} G(y;q,\beta) = 1 - q^{(1+y)^{\beta}}, & y = 0,1,2....... \\ 0 & \text{if} \quad y < 0 \end{cases} \tag{1}$$

The probability mass function is given as

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

2.2 Discrete Weibull regression

Let Y be the response variable with possible values $0,1,2........$ and let $X_1,......X_p$ be p covariates. We assume that the conditional distribution of Y given X follows a DW distribution with parameters q and β. There are a number of possible choices to link the parameters q and β to the covariates, proposing a logit and log link follows that:

i. q is dependent on X as follows:

$$\log(q/1-q) = \theta_0 + \theta_1 X_1 + ... + \theta_p X_p \tag{3}$$

or

$$\log(-\log(q)) = \theta_0 + \theta_1 X_1 + ... + \theta_p X_p$$

ii. β is dependent on X in this manner

$$\log(\beta) = \tau_0 + \tau_1 X_1 + ...\tau_p X_p \tag{4}$$

$$\log(\beta) = X\tau \text{, Where } \tau = (\tau_0, \tau_1, ...\tau_1)'$$

A discrete Weibull regression with a discrete is considered to have the ability to capture over and under-dispersion simultaneously and a closed-form analytical expression of the quantiles of the conditional distribution. Haselimashhadi *et. al.*, (2016) proposed one additional parameterization for q through a logit link function, and has shown to be rather effective for statistical inference, and a link also between the second parameter β and the covariates, in order to capture more complex dependencies.

2.3 Bayesian Inference for Discrete Weibull Regression

Bayesian estimation of regression parameters $\theta = (\theta_0.........\theta_p)'$ and $\tau = (\tau_0.........\tau_p)'$ is discussed in this sub-section. Given a number of observation n, y_i and $(x_{i1}..........x_{ip})'$, $i = 1,........., n$ for the response variable Y, and the covariates X, respectively, and making x_i to be the row vector, x_i can then be written as $x_i = (x_{i1}..........x_{ip})$. From (2) and (3) above, the likelihood is given by:

$$\ell(\theta,\tau \mid X,Y) = \prod_{i=1}^{n}\left(\left(\frac{e^{x_i\theta}}{1+e^{x_i\theta}}\right)^{y^{x_i\tau}} - \left(\frac{e^{x_i\theta}}{1+e^{x_i\theta}}\right)^{(1+y)^{x_i\tau}}\right) \tag{5}$$

Since from (3), $q = e^{x_i\theta}/1+e^{x_i\theta}$

Samples are drawn with Metropolis-Hastings sampling (Hastings, 1970) from the full conditional posterior and implementation is provided in the R package BDWreg using non-informative prior with an independent Gaussian proposal to draw samples from the posterior. From the posterior distribution, the mode of the marginal densities is used as point estimate of the parameters, while the whole distribution is used for building credible intervals. The idea about using non-informative prior distributions is basically 'to let the data speak for themselves,' so that inferences are not affected by information external to the current data (Gelman 2013). The improper prior is another way of describing the non-informative prior and there is a possibility for improper prior distributions to lead to proper posterior distributions.

Jeffreys introduced the approach that is sometimes used to define non-informative prior distributions, based on one-to-one transformations of the parameter: $\phi = g(\theta)$, expressing prior on ϕ gives:

$$p(\phi) = p(\theta)\left|\frac{d\theta}{d\phi}\right| = p(\theta)\left|g'(\theta)\right|^{-1} \tag{6}$$

The assumption of Jeffreys' general principle is that $p(\theta)$ should yield an equivalent result if applied to the transformed parameter.

The non-informative prior density can be defined following Jeffreys' principle as $p(\theta) \propto \left[I(\theta)\right]^{1/2}$, where $I(\theta)$ is the Fisher information of θ.

$$I(\theta) = \left(\left(\frac{d \log p(y|\theta)}{d\theta} \right)^2 | \theta \right) = -E\left(\frac{d^2 \log p(y|\theta)}{d\theta^2} | \theta \right) \tag{7}$$

Evaluating $I(\theta)$ at $\theta = g^{-1}(\phi)$ shows that Jefferys' prior model is invariant to parameterization.

$$\begin{aligned} I(\phi) &= -E\left(\frac{d^2 \log p(y|\phi)}{d\theta^2} \right) \\ &= -E\left(\frac{d^2 \log p(y|\theta = g^{-1}(\phi)}{d\theta^2} \left| \frac{d\theta}{d\phi} \right|^2 \right) \\ &= I(\theta) \left| \frac{d\theta}{d\phi} \right|^2 \end{aligned}$$

Hence,
$$I(\phi)^{1/2} = I(\theta) \left| \frac{d\theta}{d\phi} \right| \tag{8}$$

3. Bayesian Inference for GLMMs

GLMMs is an extension of generalized linear models (for example Poisson regression) to include both fixed and random effects (mixed models). The general form of the model (in matrix notation) is:

The general form of GLMMs model in matrix notation is

$$y = X\beta + Z\tau + e \tag{9}$$

Where y is a $N \times 1$ column vector, X and Z relates are design matrices to the fixed and random predictors to the data respectively. These predictors have associated parameter vectors β and τ, and e is a vector of residuals. In MCMCglmm, over-dispersion is always dealt with, in the data after accounting for fixed and random sources of variation. MCMCglmm does not use a multiplicative model, but an additive model.

The inverse-Wishart distribution, a multivariate generalization of the scaled inverse-χ^2, is used to describe the prior distribution of the matrix Σ. The conjugate prior distribution for (μ, Σ), the normal-inverse-Wishart, is conveniently parameterized in terms of hyperparameters $(\mu_0, \Lambda_0/\kappa_0; \nu_0, \Lambda_0)$, matrix Σ is expressed as

$$\begin{aligned} &\Sigma \sim Inv-Wishart_{\nu_0}(\Lambda_0^{-1}) \\ &\mu | \Sigma \sim N(\mu_0, \Sigma/\kappa_0) \end{aligned} \tag{10}$$

Which correspond to the joint prior density

$$p(\mu, \Sigma) \propto |\Sigma|^{-((\nu_0+d)/2+1)} \exp(-\frac{1}{2} tr(\Lambda_0 \Sigma^{-1} - \frac{\kappa_0}{2}(\mu - \mu_0)^T \Sigma^{-1}(\mu - \mu_0)) \tag{11}$$

The parameters ν_0 and Λ_0 describe the degrees of freedom and the scale matrix for the inverse-Wishart distribution on Σ. The remaining parameters are the prior mean, μ_0, and the number of prior measurements, κ_0, on the Σ scale.

Multiplying the prior density by the normal likelihood results in a posterior density result to

$$\mu_n = \frac{\kappa_0}{\kappa_0 + n}\mu_0 - \frac{n}{\kappa_0 + n}\bar{y}$$
$$\kappa_n = \kappa_0 + n$$
$$\nu_n = \nu_0 + n$$
$$\Lambda_n = \Lambda_0 + S + \frac{\kappa_0 n}{\kappa_0 + n}(\bar{y} - \mu_0)(\bar{y} - \mu_0)^T \quad (12)$$

where S is the sum of squares matrix about the sample mean,

$$S = \sum_{i=1}^{n}(y_i - \bar{y})(y_i - \bar{y})^T \quad (13)$$

Samples from the joint posterior distribution of (μ, Σ) are easily obtained using the following procedure: first, draw $\Sigma \mid y \sim Inv-Wishart_{\nu_n}(\Lambda_n^{-1})$, then draw $\mu \mid \Sigma, y \sim N(\mu_n, \Sigma/\kappa_n)$.

3.2 Variance Structures Model Parameters of GLMM

Given that (β, τ) are residuals, (e) are assumed to come from a multivariate normal distribution as given below:

$$\begin{pmatrix} \beta \\ \tau \\ e \end{pmatrix} \sim N\left(\begin{bmatrix} \phi_0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} B & 0 & 0 \\ 0 & G & 0 \\ 0 & 0 & R \end{bmatrix} \right) \quad (14)$$

Where β_0 are the prior means for fixed effects with prior co-variances, matrix B, along with G and R are the expected co-variances of the random effects and the residual respectively, (Hadfied 2010). MCMC GLMM is modelled in R-and G-structure, R structure represents the Random structure; where the latent variables are assumed to have the multivariate normal distribution. For the G-structure, the residual model is indicated in a way that should allow each linear predictor to have a unique residual. Conjugate priors the variance structures (R and G) follows an inverse-Wishart prior distribution and can be Gibbs sampled in a single block in many cases (Gelman 2006).

MCMCglmm allows variance structures of the form:

$$G = (V_1 \otimes A_1) \oplus (V_2 \otimes A_2) \oplus ....(V_n \otimes A_n) \quad (15)$$

and the inverse structure has the form

$$G^{-1} = (V_1^{-1} \otimes A_1^{-1}) \oplus (V_2^{-1} \otimes A_2^{-1}) \oplus ....(V_n^{-1} \otimes A_n^{-1}) \quad (16)$$

Where (V) and (A) are matrices, (V) is estimated, while (A) are usually high dimensional and treated as known. Each component term, however, is formed through the Kronecker product $\otimes$ which allows for possible dependence between random effects within a component term, while $\oplus$ is the direct sum. Expanded form give:

$$G = \begin{bmatrix} V_1 \otimes A_1 & 0 \\ 0 & (V_2 \otimes A_2) \end{bmatrix} \quad (17)$$

The zero off-diagonals represent the independence between component terms.The simplest form is expressed in form of Identity matrices

$$(V_1 \otimes A_1) = \sigma_1^2 I \quad (18)$$

which assumes that random effects within a component term are independent but have a common variance and appropriate G component may have the form:

$$V_1 \otimes A_1 = \begin{bmatrix} \sigma_{u_1}^2 & \sigma_{u_1,u_2} \\ \sigma_{u_2,u_1} & \sigma_{u_2}^2 \end{bmatrix} \otimes I \quad (19)$$

In practical sense, for each component of the variance structure take the arguments V , n and fix which specify the expected (co)variance matrix at the limit, the degree of freedom parameter, and the partition to condition on. When $fix = 1$, the whole matrix is fixed.

4 Simulation Study

In this session, simulation from Discrete Weibull distribution is performed in the case of over-dispersion and under-dispersion count response variable. To simulate over-dispersed response variable from DW, the value of β should be specified such that $0 < \beta \leq 1$, irrespectively of the value of q, and $\beta \geq 2$ in case of under-dispersion, irrespective of the value of q (Kalktawi *et.al.,* 2016). Bayesian and non-Bayesian estimation procedure is performed in both over and under-dispersed simulated data with R package DWreg by Vinciotti (2016). Two predictors are uniformly in interval (0, 1) and (0, 2) respectively, and simulate 1000 observations is simulated using for Bayesian technique, non-informative prior and Metropolis-Hastings algorithm with an independent Gaussian proposal is used to draw samples from the posterior, where the correlation among chains is considered in the proposal and 30,000 iterations was performed. Formation of parameters for simulation is as stated in table 1. Implementations are carried in R software by R Core team (2017).

Table 1: Formation of parameters for Simulating DW Regression Models

Model	True Parameters	Estimation type
Over-dispersion		
Model	True Parameters	Estimation type
$DW(q,\beta)$	$q = 0.8$ $\beta = 0.9$	Frequentist
Logit: $DW(q,\beta)$	$\theta_0 = 0.45, \theta_1 = 0.2, \theta_2 = 0.4$, $\beta = 0.9$	Bayesian
$\log : DW(q,\beta)$	$\theta_0 = 0.45, \theta_1 = 0.2, \theta_2 = 0.4$, $\beta = 0.9$	Bayesian
Under-dispersion		
Model	True Parameters	Estimation type
$DW(q,\beta)$	$q = 0.8$ $\beta = 0.9$	Frequentist
Logit: $DW(q,\beta)$	$\theta_0 = 0.45, \theta_1 = 0.2, \theta_2 = 0.4$, $\beta = 2.9$	Bayesian
$\log : DW(q,\beta)$	$\theta_0 = 0.45, \theta_1 = 0.2, \theta_2 = 0.4$, $\beta = 2.9$	Bayesian

Table 2: Frequentist Model Selection for Under and Over-dispersed Simulated data

Model	Inf. Criteria	Under-dispersed	(loglik)	Over-dispersion	(loglik)
Poison	AIC	1773.60		5810.3	
	BIC	1793.22	(-882.80)	5829.9	(-2901.13)
	CAIC	1787.29		5823.97	
Negbin	AIC	2074.46		4559.63*	
	BIC	2094.45	(-1033.41)	4579.25	(-2275.80)
	CAIC	2088.53		4573.30	
ZIP	AIC	2078.82		5587.90	
	BIC	2070.20	(-1033)	5580.20	(-2788.00)
	CAIC	2087.70		5597.70	
ZINB	AIC	2080.00		4563.13	
	BIC	2070.20	(-1033)	4553.32	(-2274.56)
	CAIC	1777.84		4570.82	
Hurdle Pois	AIC	1766.14		5587.65	
	BIC	1756.34	(-876.07)	5579.96	(-2787.88)
	CAIC	1777.84		5597.48	
Hurdle NB	AIC	1766.14		4563.00	
	BIC	1738.40*	(-867.1)	4553.01*	(-2274.50)
	CAIC	2088.52*		4570.70	
CMP	AIC	2073.83		6598.66	
	BIC	2072.83	(-1033.41)	6613.9	(-3296.33)
	CAIC	1764.66		6614.37	
D-WB	AIC	1756.96*		4565.84	
	BIC	1747.14	(-871.49)	4579.32	(-2275.92)
	CAIC	1764.66		-2275.92	

In the case of Bayesian Glmm, multivariate normal proposal distribution is used, which is determined during burin-in phase by adaptive methods. The (R and G) variance structure follows an inverse-Wishart distribution prior using Gibb's sampling approach to draw sample from the posterior distribution.

For the BDW, we consider 30,000 iterations of the sampler and use the first 25% of the data as burn-in. The acceptance rate for the scale proposal is found to be (63.6; 64.32) % in the case of over-dispersed data, while (72.15; 73.42) % respectively for under-dispersed. Figure (1) shows the posterior distribution of the parameters and the chain convergence for under-dispersed count data under the logit link. Similar plots are obtained for the other cases. Figure (2) shows the marginal densities of the parameters and the 95% HPD.

Table 3: Bayesian Model Selection for Under-dispersed and Over-dispersed simulated data

Model	AIC	BIC	CAIC	QIC	DIC	PBIC
$\log it_U : BDW(q,\beta)$	1752.942	1772.57	1776.57	1.75603	1751.12*	1755.l52*
$\log_U : BDW(q,\beta)$	1751.84*	1771.4*	1775.5*	1.75493*	1751.21	1755.345
$\log it_O : BDW(q,\beta)$	-	-	-	-	2070.82	-
$\log_O : BDW(q,\beta)$	4561.804	4581.43	4585.43	4.56489	4559.66	4563.63
$BPglmm_U$	4560.406*	4580.03*	4584.00*	4.56341*	4559.65	4563.55*
$BPglmm_O$	-		-	-	3910.63*	-

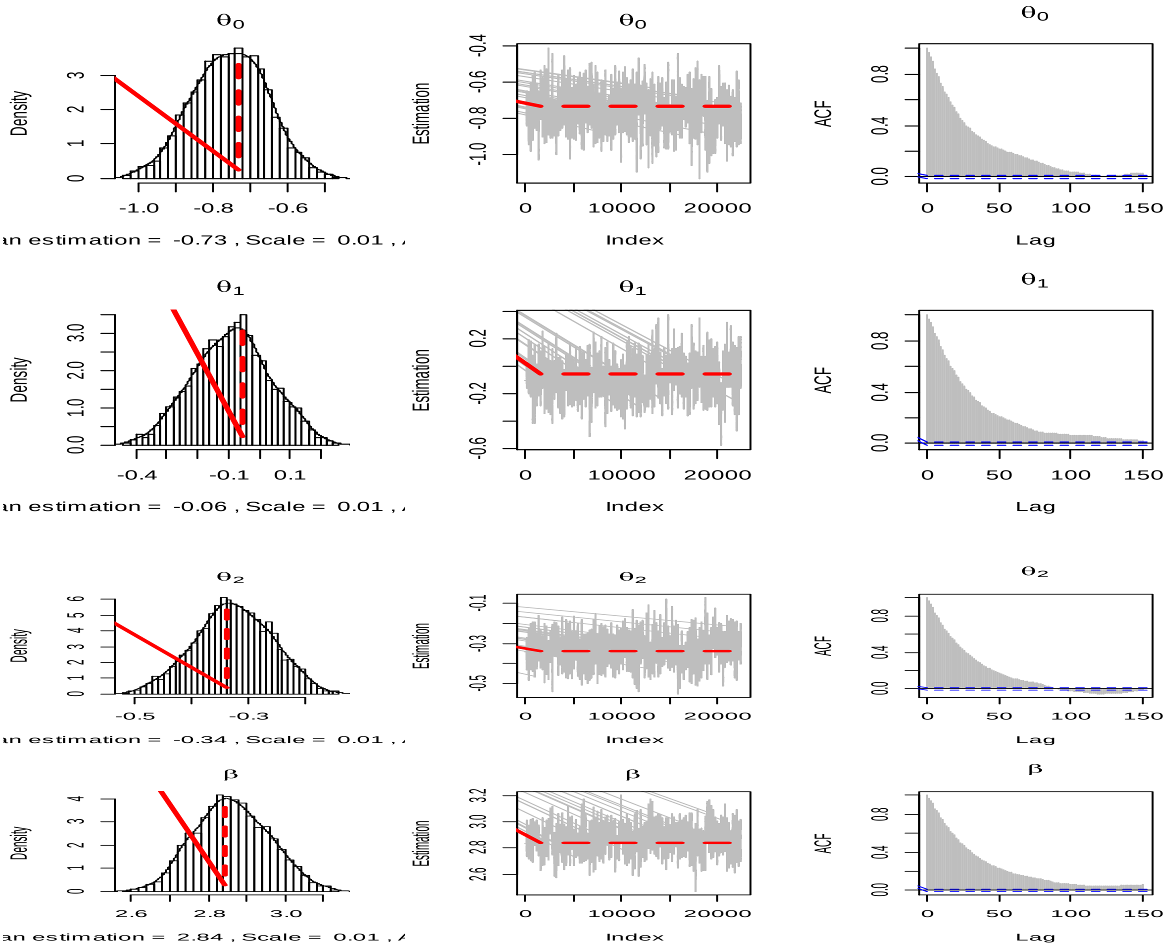
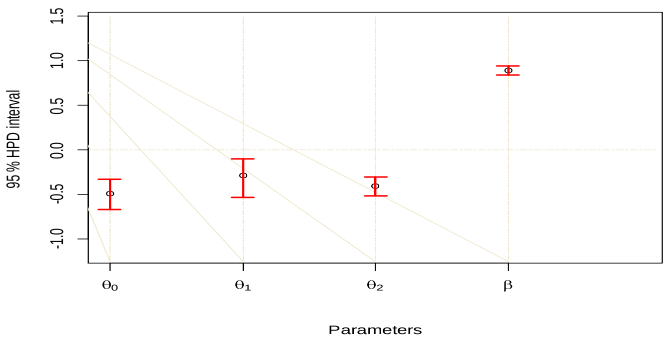
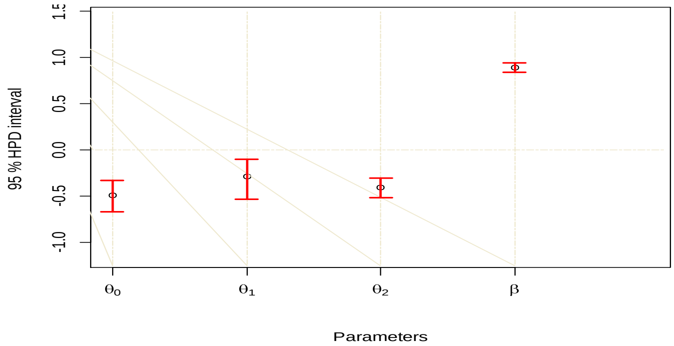


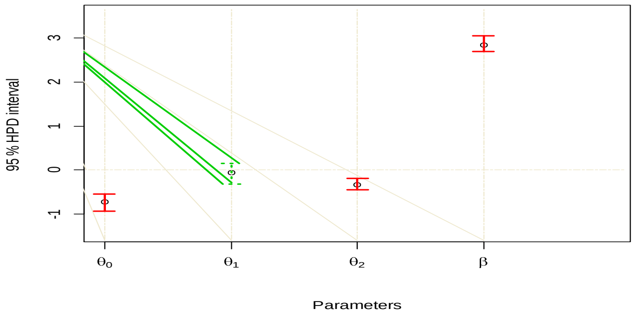
Figure 1: Posterior distribution of the Parameters and the Chain Convergence for Simulated Data.



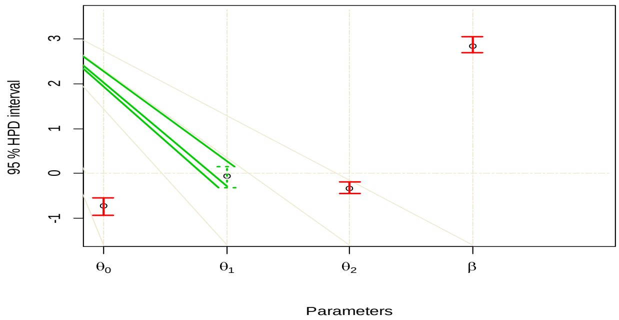
Log: over-dispersed



Logit over-dispersed



Log: under-dispersed



Logit: under-dispersed

Figure 2: Marginal Densities of Parameters and 95% HPD

The frequentist approach and Bayesian techniques is carried out on both under-dispersed and over-dispersed simulated dataset and compared. The Poisson, Negative Binomial, Zero-inflated Poisson, Zero-inflated NB, Hurdle models, COM-Poisson and DW was carried out using frequentist estimation. Also BDW logit-link, BDW log-link and MCMCglmm was equally implemented and the models are compared on the basis of criteria such as; Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC), logLik, Consistent AIC (CAIC) for frequentist. For Bayesian estimation, Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC), Deviance Information Criterion (DIC), Quasi-likelihood Information Criteria (QIC), Bayesian Predictive Information Criterion (BPIC), and Consistent AIC (CAIC).

Figure (2) shows that no parameter is significant in the case of over-dispersion (broken green lines), while only parameter one parameter (θ_1) is significant for both log link and logit link. In tables 2 and 3, where the frequentist models is represented in table (2) and Bayesian models in table (3). The (*) indicate the minimum value of the information criteria for over and under-dispersed data for simulated data, the lower the better. $\log it_U : BDW(q,\beta)$ represents under dispersed with logit link, $\log_U : BDW(q,\beta)$ represents under dispersed with log link, $\log it_O : BDW(q,\beta)$ represents over-dispersed with logit link, $\log_O : BDW(q,\beta)$ represents over-dispersed with log link, $BPglmm_U$ represents under dispersed MCMC Glmm based on Poisson distribution, and $BPglmm_O$ represents over dispersed MCMC Glmm based on Poisson distribution.

In the class of frequentist estimation technique, discrete Weibull (DW) outperforms others, in the case of under-dispersed count data while negative Binomial outperformed other models in the case for over-dispersed count data.

In the class of Bayesian model for under-dispersed count data, $\log_U : BDW(q,\beta)$ and $\log it_U : BDW(q,\beta)$ outperforms $BPglmm_U$. While $BPglmm_O$ outperforms $\log_O : BDW(q,\beta)$ and $\log it_O : BDW(q,\beta)$ in the case of over-dispersed data.

5. Analysis of Counts in Scholarship

In this section, the performance of the Bayesian discrete Weibull regression model is examined in the case of real-life datasets from the academic domain. Comparison is drawn between Bayesian discrete Weibull regression model and Bayesian MCMCglmm, and eight other regression models based on frequentist (Poisson, Neg Bin, ZIP, ZINB, Hurdle Poisson, Hurdle Negbin, Discrete Weibul, and COM-Poisson) based on the criteria stated above.

BDW model is fitted with non-informative prior on the regression parameters, carrying out 30,000 iterations for the Metropolis-Hastings algorithm and an acceptance rate in the (44.8; 45.93) % interval for over-dispersed data, implemented in BDWreg package, "MCMCglmm" package was used for Bayesian Multivariate Generalised Linear Mixed Model based on Poisson regression. Frequentist regression was carried out with the package COUNT in R by Hilbe (2016). Empirical analysis shows that logit(q) link of BDW outperformed the log(q) link in the case of over-dispersion. In the class of frequentist estimation technique, Hurdle Negbin outperforms other models, while Bayesian Poisson (BPglmm) outperforms log $BDW(q,\beta)$ and logit $BDW(q,\beta)$ in the case of Bayesian. For over-dispersed excess zeros, using frequentist approach, ZIP and Hurdle Negbin outperforms other models. For the Bayesian technique, Poisson (BPglmm) outperforms log $BDW(q,\beta)$ and logit $BDW(q,\beta)$.

Figure 4 shows the 95%Highest Posterior Distribution interval (HPD) for Bayesian discrete Weibull, while figure 3 shows the trace plot and density of Bayesian Glmm for over-dispersed count data. Figure 4 shows the parameters that are significant (green and doted), while the red and thick lines are parameters that are not significant.

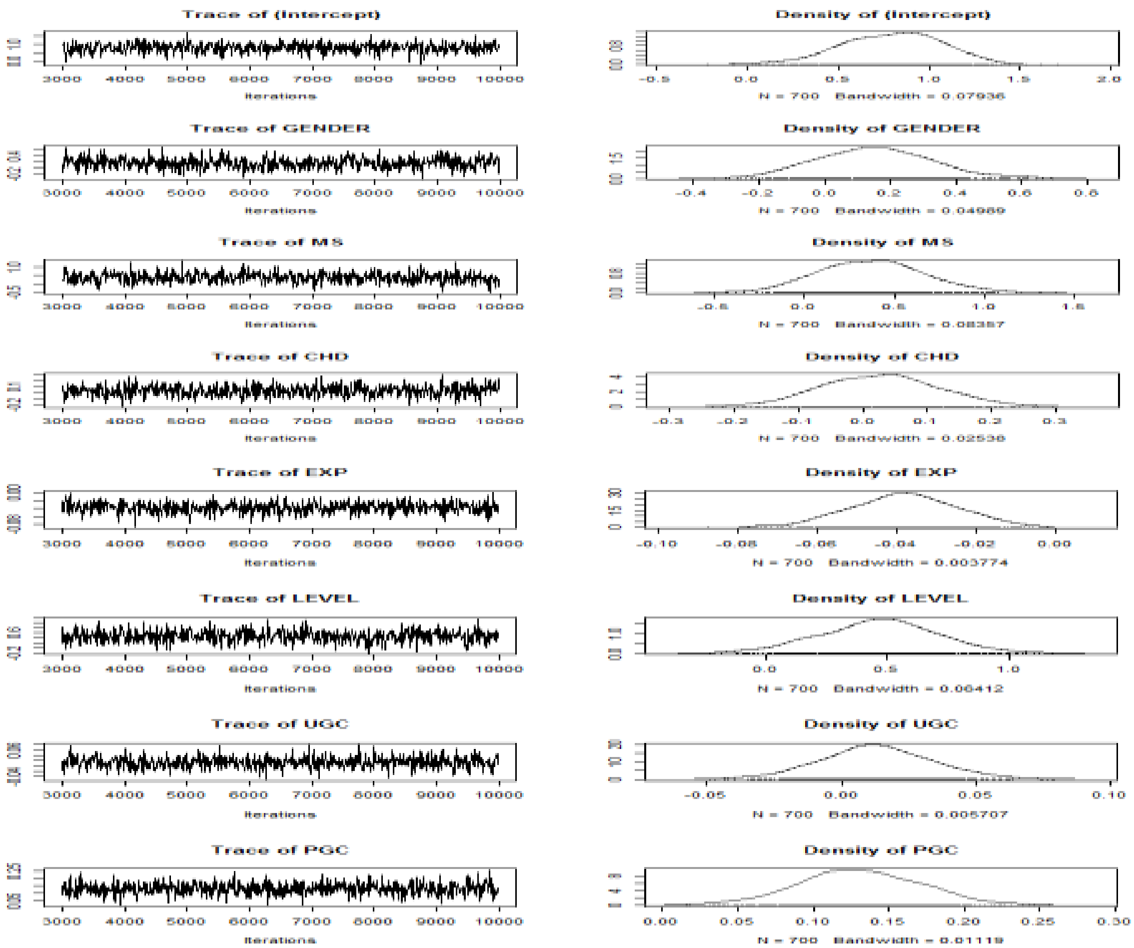


Figure 3: Trace of Sampled Output and Density Estimates of the Covariates for over-dispersed with Bayesian Glmm

ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.7, No.9, 2017

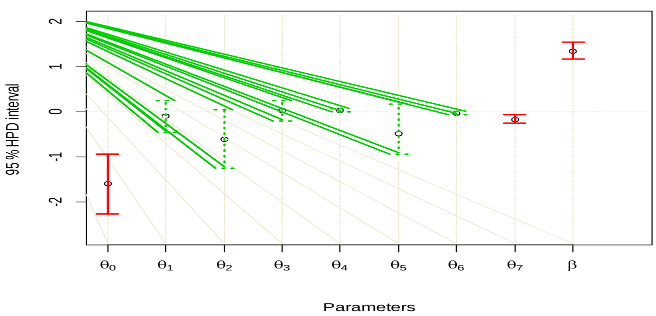
Bayesian techniques have proven to perform better than the frequentist, particularly in the case the Bayesian MCMCglmm for over-dispersed and over-dispersed excess zeros for the real life count data considered in this study.

Table 4: Model Selection for Over-dispersed and Over-dispersed Excess Zero for Frequentist Models

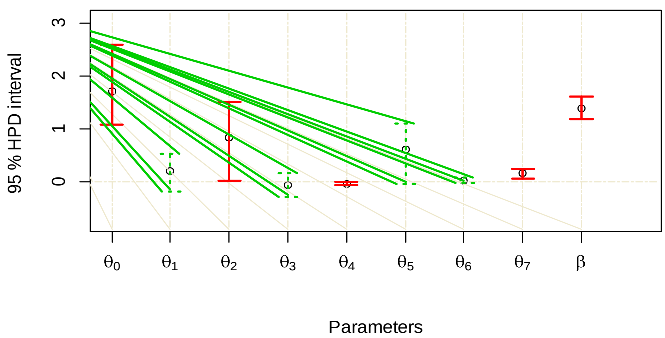
Model	Inf. Criteria	Over-dispersed (loglik)		Over-dispersion Excess zero(loglik)	
Poison	AIC	765.81		551.45	
	BIC	761.33	(-373.90)	576.97	(-266.73)
Negbin	AIC	720.927		492.28	
	BIC	746.455	(-351.46)	517.81	(-237.14)
ZIP	AIC	795.70		486.16*	
	BIC	768.00	(-381.90)	458.40	(-227.1)
ZINB	AIC	720.33		486.89	
	BIC	690.70	(-343.19)	537.09	(-226.45)
Hurdle Pois	AIC	795.56		487.81	
	BIC	767.76	(-381.78)	460.01	(-227.91)
Hurdle NB	AIC	720.31*		488.68	
	BIC	690.52*	(-343.16)	454.68*	(-227.34)
CMP	AIC	866.00		515.00	
	BIC	866.69	(866.00)	537.91	(-249.39)
D-WB	AIC	723.45		494.86	
	BIC	708.02	(-871.49)	497.92	(-237.762)

Table 5: Model Selection for Over-dispersed and Over-dispersed Excess Zero for Bayesian Models

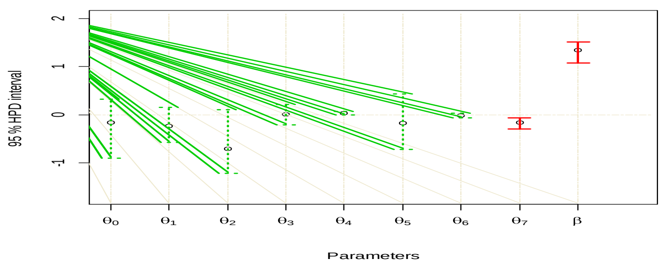
Model	AIC	BIC	CAIC	QIC	DIC	PBIC
$\log it_U : BDW(q, \beta)$	722.877*	748.404*	757.404*	5.90815*	721.832	730.688
$\log_U : BDW(q, \beta)$	723.246	748.772	757.773	5.91108	721.3956	730.077 *
$\log it_O : BDW(q, \beta)$	548.40	568.251	550.10	-	638.673*	-
$\log_O : BDW(q, \beta)$	494.557	520.0839	528.0839*	4.09609	493.8645	502.904
$BPglmm_U$	493.965*	519.4918*	528.4918	4.09139*	493.4428	502.487*
$BPglmm_O$	403.36	423.22	405.06	-	463.052*	-



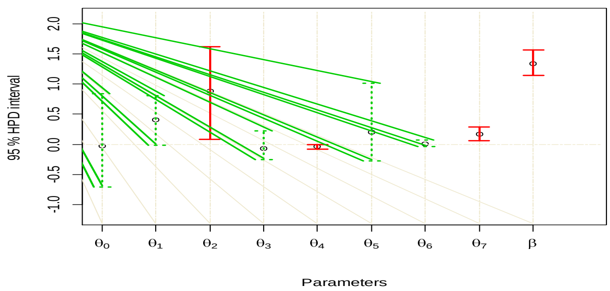
Over-dispersed log



Over-dispersed logit



Over-dispersed logit



Over-dispersed logit

Figure 4: Marginal Densities of Parameters and the 95% HPD

Table 6: Posterior Mean and Confidence Interval for Bayesian GLMM for the period of one year

	Post.mean	l-95%CI	u-95%
Intercept	1.466	-2.840	6.014
GENDER	4.747	0.085	8.874
MS	5.048	0.093	10.84
CHD	0.5033	0.0657	1.423
EXP	0.233	0.0679	0.488
LEVEL	7.387	0.080	9.604
UGC	0.242	0.066	0.515
PGC	0.404	0.074	0.978

Table 7: Posterior Mean and Confidence Interval for Bayesian GLMM for the period of three years

	Post.mean	l-95%CI	u-95%
Intercept	0.147	-4.748	4.864
GENDER	5.854	0.069	14.26
MS	5.986	0.067	16.5
CHD	0.483	0.056	1.330
EXP	0.263	0.072	0.555
LEVEL	3.640	0.081	9.637
UGC	0.355	0.078	0.814
PGC	0.485	00728	1.193

From the posterior means and confidence intervals in table 6, and 7, it is observed that Gender, Level, and Marital Status really contribute to the variability of number of article production by lecturers. Number of children, year of experience, undergraduate courses taught, and postgraduate courses taught do not have much impact on number of journal article produced by lecturers. Result from frequentist approach further explained with Zero-inflated Negative Binomial as follows:

From table (8) below, for every increase in male gender in the system, the number of publications produced will increase by a factor of 1.29. Addition of married individual into the system increases number of publication by 2.76. Also, addition of a child into the family of a lecturer reduces the number of article publication by a factor of 0.894. Increase in year of experience does not necessarily increase publication output; it reduces it by a factor of 0.99. Every increase in number of lecturer in senior cadre increases publication output by 1.47. Every increase in undergraduate courses taught increases publication output by a factor of 1.02.

Table 8: Coefficient and Confidence Interval for number of publication in one year

	CO	2.5 pct	97.5 pct
count_(Intercept)	7.588246e-01	4.419457e-01	1.302908e+00
count_GENDER	1.290373e+00	9.286388e-01	1.793015e+00
count_MS	2.764620e+00	1.564341e+00	4.885844e+00
count_CHD	8.944647e-01	7.618634e-01	1.050145e+00
count_EXP	9.906919e-01	9.637305e-01	1.018408e+00
count_LEVEL	1.474057e+00	9.348171e-01	2.324353e+00
count_UGC	1.019014e+00	9.873434e-01	1.051700e+00
count_PGC	1.066096e+00	9.921887e-01	1.145508e+00

From table (9) below, for every increase in male gender in the system, the number of publications produced will increase by a factor of 1.18. Addition of married individual into the system increases number of publication by 1.51. Also, addition of a child into the family of a lecturer reduces the number of article publication by a factor of 0.99. Increase in year of experience does not necessarily increase publication output; it reduces it by a factor of 0.98. Every increase in number of lecturer in senior cadre increases publication output by 1.16. Every increase in undergraduate courses taught reduces publication output by a factor of 0.99.

Table 9: Coefficient and Confidence Interval for number of publication in three years

	CO	2.5 pct	97.5 pct
count_(Intercept)	3.9252303	2.97641196	5.1765121
count_GENDER	1.1811979	1.00906103	1.3826999
count_MS	1.5098190	1.13049070	2.0164283
count_CHD	0.9940997	0.92085986	1.0731646
count_EXP	0.9804618	0.96820485	0.9928739
count_LEVEL	1.1649363	0.92776472	1.4627378
count_UGC	0.9962218	0.98053341	1.0121612
count_PGC	1.1123543	1.07318927	1.0121612

6. Summary and Conclusion

In this study count data has been fitted using frequenist models along with well-known Bayesian regression models of Bayesian Discrete Weibull and MCMCglmm. The performance of these models that are being used for fitting count data is closely observed using simulated data and real life data. Also, application of the proposed models have been considered in academic domain, particularly analysing counts of number of article publication(s) by lecturers in a private University in Nigeria; for both one year period and three year period respectively. Simulation was carried out for both over- and under-dispersed data where response variable is taken from from Discrete Weibul distribution and predictors from uniform distribution respectively, so as to test the performance of these models and draw comparison among them, both from Bayesian and frequentist techniques.

Simulation study shows that Bayesian estimation method performs well compared to frequentist. In either case, Bayesian Discrete Weibull shows superior technique than frequentist DW; also, Poisson model based on MCMCglmm shows a superior technique to frequentist Poisson particularly for over-dispersed and over-dispersed excess zeros, but not under-dispersed data.

The study further shows that BDW with logit link seems to perform credibly well when covariates are few, that is, in the case of simulation study carried out with two covariates. The real life data set consist of seven covariates, and the log link outperforms the logit link. Also MCMCglmm based on Poisson distribution is observed to perform well with over-dispersed data as noted by Hadfield (2010), but not in the case of under-dispersion as shown in the simulation study. The applicability Bayesian models to real life datasets give a meaningful result and reasonable inference is drawn from it.

Based on the results obtained, the following are hereby recommended:

- (i) In the class of frequentist models, discrete Weibul fits count data well, it is more suitable for under-dispersed among other models under examination. Therefore, it is recommended for use.
- (ii) Negative Binomial and Hurdle negative Binomial performs well both for under-dispersed and over-dispersed data. This is demonstrated in the simulation and real life data.
- (iii) In the class of Bayesian models, Bayesian discrete Weibull with logit link should be considered when predictors are few, but when number predictors is large, but Bayesian discrete Weibull with log link should be considered when predictors are few.
- (iv) Bayesian Generalized mixture model have proven to outperform Bayesian discrete Weibull for over-dispersed data and not under-dispersed. It is therefore recommended for use for over-dispersed and over-dispersed excess zeros.

Reference

Bolstad W. M (2007). Introduction to Bayesian Statistics, 2nd ed., *John Willey & sons, Inc., publication*

Cameron, A. C., Trivedi, P. K. (2005), Microeconometrics Methods and Application; Cambridge University Press

Chandra, N and V.K. Rathaur (2017) Bayes Estimation of Augmenting Gamma Strength Reliability of a System under Non-informative Prior Distributions . *Calcutta Statistical Association Bulletin. Volume: 69 issue: 1, page(s): 87-102. https://doi.org/10.1177/0008068317696574*

Chanialidis, . C (2015) *Bayesian mixture models for count data.* PhD thesis. http://theses.gla.ac.uk/6371/

Gelman, .A; Carlin, J. B.; Stern, H. S., Duson D.B, Vehtari A and Rubin, D. B (2013). Bayesian Data Analysis. Chapman and Hall/CRC, 2nd ed. Edition.

Hadfield, J. D (2010). MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. Journal of Statistical Software, 33(2), 1-22. URL http://www.jstatsoft.org/v33/i02/.

Haselimashhadi. H (2016). BDWreg: Bayesian Inference for Discrete Weibull Regression. R package version 1.0.0. https://CRAN.R-project.org/package=BDWreg

Haselimashhadi,. H, Vinciotti,. V and Yu. K (2016). A new Bayesian regression model for counts in medicine

Hastings, W. K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57 (1), 97–109.

Hilbe. J.M (2016). COUNT: Functions, Data and Code for Count Data. R package version 1.3.4. https://CRAN.R-project.org/package=COUNT

Kalktawi, H.S, Yu. K And Vinciotti,. V (2016). A Simple and Adaptive Dispersion Regression Model for Count Data

Klein. N, Kneib. T, and Lang, S. (2015). Bayesian Generalized Additive Models for Location, Scale, and Shape for Zero-Inflated and Over-dispersed Count Data. Journal of the American Statistical Association, Vol 110, issue 509, page 405-419.

Murat,. N, Ali Mehmet, and Cengiz, B. K (2015). A Bayesian Appproach for Zero Inflate d Count Data in Psychology research. Adv Res, 6(1): 20-26.

Nagakawa, T., Osaki, S., 1975. The discrete Weibull distribution. IEEE transactions on reliability R-24 (5).

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.Rproject.org/.

Saengthong. P , Bodhisuwan. W, and Thongteeraparp. A. (2015): The zero inflated negative binomial – Crack distribution: some properties and parameter estimation.