

Article

Categorical Data Analysis Using a Skewed Weibull Regression Model

Renault Caron ¹, Debajyoti Sinha ², Dipak K. Dey ³ and Adriano Polpo ^{1,*}

¹ Department of Statistics, Federal University of São Carlos, São Carlos 13565-905, Brazil; renaultcaron@hotmail.com

² Department of Statistics, Florida State University, Tallahassee, FL 32306, USA; sinhad@stat.fsu.edu

³ Department of Statistics, University of Connecticut, Storrs, CT 06269, USA; dipak.dey@uconn.edu

* Correspondence: polpo@ufscar.br

Received: 24 November 2017; Accepted: 27 February 2018; Published: 7 March 2018

Abstract: In this paper, we present a Weibull link (skewed) model for categorical response data arising from binomial as well as multinomial model. We show that, for such types of categorical data, the most commonly used models (logit, probit and complementary log–log) can be obtained as limiting cases. We further compare the proposed model with some other asymmetrical models. The Bayesian as well as frequentist estimation procedures for binomial and multinomial data responses are presented in detail. The analysis of two datasets to show the efficiency of the proposed model is performed.

Keywords: asymmetric model; binomial response; multinomial response; skewed link; Weibull distribution

1. Introduction

The statistical problem of estimating binary response variables is very important in many areas including social science, biology and economics [1]. The vast bibliography of categorical data presents the big evolution of the methods that handle appropriately binary and polychotomous data. More details can be found in Agresti [2]. Generalized linear model (GLM) has a wide range of tools in regression for count data [3]. Two important and commonly used symmetric link functions in GLM are the logit and probit links [4]. Many studies have investigated the limitations of these symmetric link functions. It is well accepted that when the probability of the binary response approaches 0 at a different rate from the rate (as a function of covariate) it approaches 1, symmetric link functions cannot be appropriate [5]. Many parameteric classes of link functions are in the literature, including the power transform of logit link by Aranda-Ordaz [6] and the a general link class of Chen et al. [5]. Other works with one-parameter class include Guerrero and Johnson [7], Morgan [8], Whittmore [9] and a host of others. Existing models for two-parameter families include Stukel [10], Prentice [11], Pregibon [12], Czado [13] and Czado [14].

Stukel's model with transformation of both tails of logit link is very general and can approximate many important links including probit, logit and complementary log–log. However, the Bayesian analysis of Stukel's model is not straightforward to implement, particularly in presence of multiple covariates and noninformative improper priors. The model proposed by Chen et al. [5], which includes the skew-probit model, uses a latent variable approach [15] that is convenient for sampling from the posterior distribution. Using the Albert and Chib [15] technique, Kim et al. [16] proposed the generalized t-link models, Naranjo et al. [17] proposed the asymmetric exponential power (AEP) model, and Rubio and Liseo [18] discuss the Jeffreys prior for skew-symmetric models. However the frequentist analysis for these models are not trivial. For the skew-probit model, The existence of the maximum likelihood estimator (MLE) of the linear regression parameters (β) can be proved only under the restrictive condition that the skewness parameter of the link function is known [19].

The majority of the works in literature are devoted to the models for binary response data. For the case of multinomial data, the multinomial extension of the logit link [20] (Chapter 8) and associated inference tools are simple to perform, and the marginal distribution of each component preserves the logit link. As mentioned before, the symmetric link may not be always appropriate. This is even clearer in multinomial data, where the sense of symmetric link is not simple to state. Generally, some categories has few observations when compared to the other ones, suggesting the idea of asymmetric distribution. We are also not aware of any model with asymmetric link function for multinomial data.

Caron and Polpo [21] briefly suggested an asymmetrical link function, called Weibull link, exclusively for binary response data. The use of Weibull distribution in survival/reliability analysis is well known. One important fact is the simplicity of the distribution, which has an analytic expression for the distribution function. Our proposed link model, based in the Weibull distribution, preserves this simplicity and it is a good option for the analysis of binary data.

In this paper, we take the Bayesian route and extend their work to multinomial data. Further we present for the first time the associated Bayesian inference tools and explore the properties of the proposed link function. We show that the benefits of this model are as follows: (1) flexibility of the Weibull distribution; (2) logit, probit and complementary log–log links as limiting cases; (3) case of implementation of both frequentist and Bayesian inferences; and (4) a general extension to handle multinomial response. The implementation of the associated Markov chain Monte Carlo (MCMC) algorithm to sample from posterior distribution is not complicated. In addition, we develop an Empirical Bayes tool [22,23] to obtain the prior when there is no relevant prior information available to the statistician.

We illustrate the use of Weibull link via analysis of two following data examples. (1) For the experiment to study the potencies of three poisons [24], the main binary response is whether the insect is alive after being treated with assigned dose level. For this example, we compare our Weibull link model with other asymmetric and symmetric link models. (2) The main response of the study by Grazeffe et al. [25] is the multiple levels of DNA damage in circulating hemocytes of each adult snail irradiated with an assigned dose. This study is used to illustrate the analysis of multinomial response data under Weibull link model, and comparing the results with those obtained by Grazeffe et al. [25] using logistic regression.

The article is organized as follows. In Section 2, we present the Weibull model, its novel properties and some approximations of the link function. In Section 3, we present the estimation procedures using MLE as well as the Bayesian estimation. In Section 3, we also present the estimation procedure for multinomial response. Section 4 is devoted to illustrating the Weibull link for analyzing two real datasets, and comparison with other existing models. Finally, Section 5 presents some future considerations and final comments.

2. Weibull Regression Model

2.1. Link function

Let $\mathbf{X} = (\mathbf{1}, \mathbf{X}_1, \dots, \mathbf{X}_r)'$ be the design matrix, where $\mathbf{1}$ is a vector with all values equal to 1, $j = 1, \dots, r$. We denote the vector of binary response variable as \mathbf{Y} . Similar to GLM, our interest lies in modeling the probability $\Pr[Y_i = 1 \mid \eta_i] = \mu(\eta_i) = E(Y_i)$ as $\Pr[Y_i = 1 \mid \eta_i] = g^{-1}(\eta_i)$, $i = 1, \dots, n$, where $\boldsymbol{\eta} = \boldsymbol{\beta}\mathbf{X}$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_r)$ are the linear coefficients, and $g(\cdot)$ is the link function. The link function relates the covariates \mathbf{X} with the mean response $\mu = E(Y \mid X)$. In this case, the g^{-1} is a cumulative distribution function (cdf) on the real line. Our interest is a link function that can accommodate symmetric and asymmetric tails which has a simple parametric functional form, and can be easily tractable. To obtain these goals, we use the cdf of Weibull distribution

$$F(\eta) = 1 - \exp\{-(\eta - \alpha)^\gamma\} \mathbb{I}_{(\eta > \alpha)}, \quad (1)$$

for g^{-1} , where $\alpha \in \mathbb{R}$ is the location/threshold parameter, $\gamma > 0$ is the shape parameter, and $\mathbb{I}_{(\eta > \alpha)}$ is the indicator of $\eta > \alpha$. \mathbb{I}_A is the indicator function of event A , that is $\mathbb{I}_A = 1$ and $\mathbb{I}_{A^c} = 0$.

Alternatively, the Weibull link function is defined as

$$\begin{aligned} \eta &= g(\mu) = [-\log(1 - \mu)]^{\frac{1}{\gamma}}, \\ \mu &= g^{-1}(\eta) = 1 - \exp\{-\eta^\gamma\}, \end{aligned} \tag{2}$$

where $\mu(\eta) = E(Y | \eta) \geq 0$, $\gamma > 0$, and $\eta > 0$.

Note that, in the above parameterization, the restriction of $\eta > 0$ is not a problem because the parameter β_0 plays the role of both the location/threshold parameter α and the intercept of linear predictor $\eta = \beta X$. By doing this, we avoid the identifiability problem in estimation of β_0 , also we have a more parsimonious model. The skewness of the Weibull link depends only on the parameter γ , and can be evaluated by $(\Gamma_3 - 3\Gamma_2\Gamma_1 + 2\Gamma_1^3)/(\Gamma_2 - \Gamma_1^2)^{3/2}$, where $\Gamma_j = \Gamma(1 + j/\gamma)$ and $\Gamma(\cdot)$ is the Gamma function. The skewness lies in the interval $(-1.1395, \infty)$. We also evaluated the Arnold–Groeneveld (AG) skewness measure [26], which is a skewness measure related to the mode of a distribution. Again, the AG skewness depends only on the parameter γ , and can be evaluated as $2 \exp\{(1 - \gamma)/\gamma\} - 1$, and lies in the interval $(-0.26424, \infty)$. However, sometimes, a model with skewness lower than -1.1395 is desired; in this case, we can use the reflected Weibull distribution to define the link as $\mu = g^{-1}(\omega) = \exp\{-\eta^\gamma\}$, and the skewness lies in the interval $(-\infty, 1.1395)$. The different forms of Weibull link are shown in Figure 1 with solid line for the Weibull link and dashed line for the reflected Weibull link.

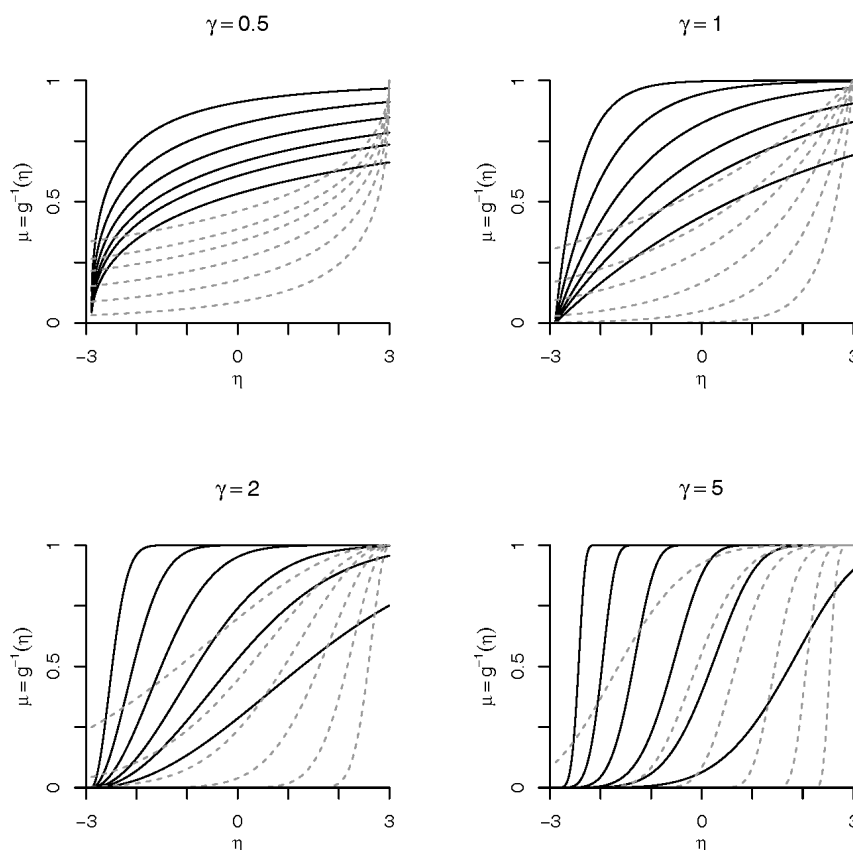


Figure 1. Forms of Weibull link. Solid lines are for the Weibull link and dashed lines are for the reflected Weibull link. We have used $\eta = \beta_0 + \beta_1 x$, where x is a grid in $(0.0001, 5.9)$, $\beta_0 = -2.9$, and for β_1 we have considered the values 0.2, 0.3, 0.4, 0.6, and 2.

2.2. Special Cases

The choice of the Weibull distribution as link function is due to its flexible properties. Rinne [27] (Chapter 3) discusses the various properties of Weibull along with Weibull distribution as approximation to some symmetrical distributions. We highlight the relations of Weibull with the normal and logistic distributions, because they explain the relations of Weibull link with probit and logit link functions. Based on results of Rinne [27], we have

$$g_1^{-1}(\eta) = 1 - \exp \left\{ -(0.90114 + 0.27787\eta)^{3.60235} \right\} \approx \Phi(\eta),$$

$$g_2^{-1}(\eta) = 1 - \exp \left\{ -(0.89864 + 0.16957\eta)^{3.50215} \right\} \approx \frac{\exp(\eta)}{1 + \exp(\eta)},$$

where Φ is the distribution function of the standard normal distribution. These results show that Weibull link can approximate the probit link and the logit link. The degrees of these approximations are illustrated in Figure 2.

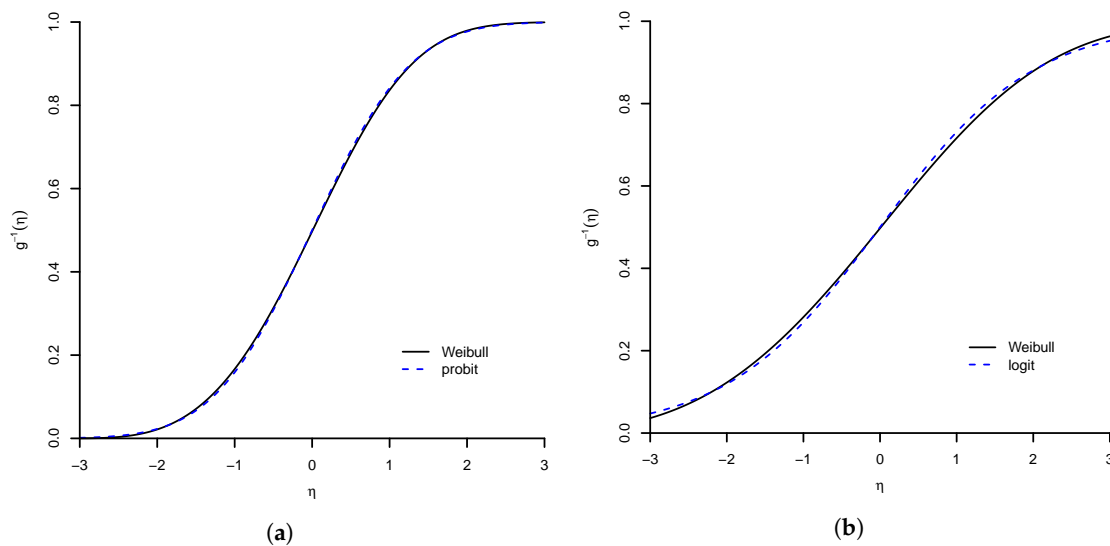


Figure 2. Similarity of Weibull link with: (a) probit link; and (b) logit link. The maximum absolute distance between Weibull link and probit link is 0.0078, and with logit link is 0.0148.

We have the following proposition for another important case of link, the complementary log–log link [4].

Proposition 1. *The complementary log–log link defined by $g^{-1}(\eta) = 1 - \exp\{-\exp(\eta)\}$ is a limiting case of the Weibull link because*

$$\lim_{\gamma \rightarrow \infty} \left\{ 1 - \exp \left[- \left(1 + \frac{\eta}{\gamma} \right)^\gamma \right] \right\} = 1 - \exp\{-\exp(\eta)\}. \tag{3}$$

Proof. Taking $\alpha = -1$ in Equation (1) and dividing η by γ , without loss of generality, we can rewrite the Weibull link given in Equation (2) as:

$$g^{-1}(\eta) = 1 - \exp \left\{ - \left(1 + \frac{\eta}{\gamma} \right)^\gamma \right\}.$$

Now, taking the limit $\gamma \rightarrow \infty$ of $g^{-1}(\eta)$ completes the proof. \square

Given this result, we can say that for a dataset when the estimated value of γ is large then the complementary log–log link should be appropriate. Using the reflected Weibull link, we have a similar

result with the log–log link, defined as $g^{-1}(\eta) = \exp\{-\exp(-\eta)\}$ [4]. The complementary log–log and log–log link as limiting cases are illustrated in Figure 3.

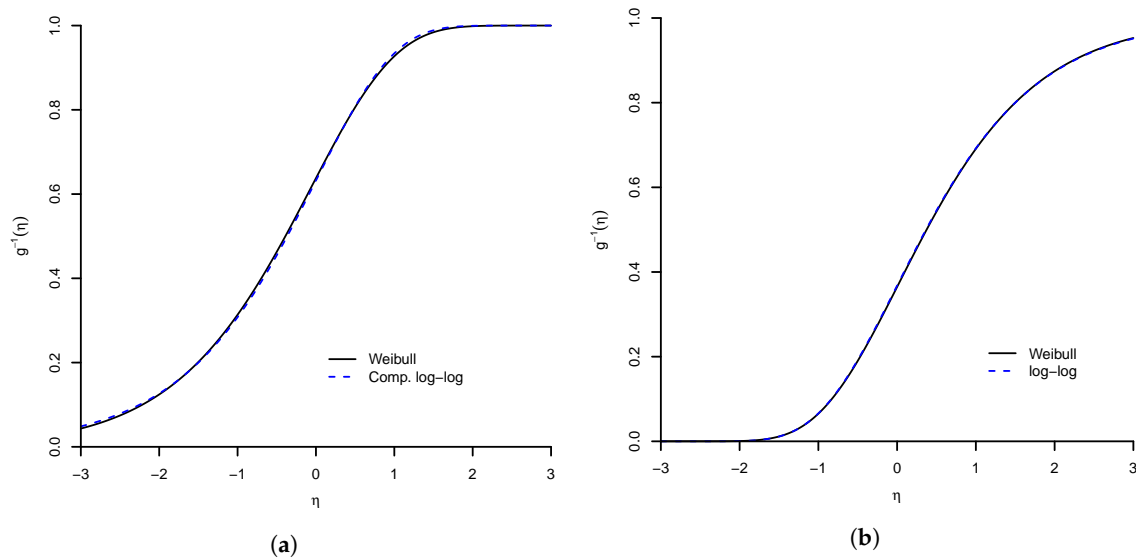


Figure 3. Similarity of Weibull link with: (a) complementary log–log link; and (b) log–log link. The maximum absolute distance between Weibull link and complementary log–log link is 0.0082 (the value of γ parameter was 21.03), and with log–log link is 0.0031 (the value of γ parameter was 114.45).

3. Estimation

3.1. Binomial Data

Consider a sample of size n from the binary variable/response Y , with $\Pr[Y_i = 1] = p_i$ for $i = 1, \dots, n$. We denote the observed data as $\mathcal{D} = \{n, Y = \mathbf{y}, X = \mathbf{x}\}$, where $\mathbf{y} = (y_1, \dots, y_n)$ is the observed vector of $Y = (Y_1, \dots, Y_n)$, and $\mathbf{x} = (\mathbf{1}, x_1, \dots, x_r)'$, is the observed covariate matrix of $X = (\mathbf{1}, X_1, \dots, X_r)'$. The likelihood function for the Weibull link can be written as

$$\begin{aligned}
 L(\boldsymbol{\beta}, \gamma \mid \mathcal{D}) &\propto \prod_{i=1}^n p_i^{y_i} (1 - p_i)^{1-y_i} \\
 &\propto \prod_{i=1}^n [1 - \exp\{-\eta_i^\gamma\}]^{y_i} [\exp\{-\eta_i^\gamma\}]^{1-y_i},
 \end{aligned} \tag{4}$$

and the log-likelihood as

$$l(\boldsymbol{\beta}, \gamma \mid \mathcal{D}) \propto \sum_{i=1}^n [y_i \log\{1 - \exp(-\eta_i^\gamma)\} - (1 - y_i)\eta_i^\gamma], \tag{5}$$

where η_i is the i -th element of the vector $\boldsymbol{\eta} = \boldsymbol{\beta}X$, and $\boldsymbol{\beta}, \gamma$ are the parameters to be estimated.

A numerical method such as Nelder and Mead [28] can be used to obtain the MLE for $(\boldsymbol{\beta}, \gamma)$. The expression of the gradient vector and Hessian matrix are given in Appendix A. Using the gradient vector and Hessian matrix, it is simple to implement a Newton–Raphson algorithm to obtain the MLE. As initial guesses for the numerical algorithm, we suggest to use the estimator $\tilde{\boldsymbol{\beta}}_{i,probit}$ under probit model for β_i ($i \neq 0$), $\tilde{\beta}_{0,guess} = -\min(\tilde{\boldsymbol{\beta}}_{probit} \mathbf{x}) + 0.001$ for β_0 , and 3.60235 for γ . The initial guesses $(\boldsymbol{\beta}, \gamma)$ can be interpreted as the Weibull link being an approximate probit link.

For the Bayesian analysis, the posterior density is

$$p(\boldsymbol{\beta}, \gamma \mid \mathcal{D}) \propto L(\boldsymbol{\beta}, \gamma \mid \mathcal{D})p(\boldsymbol{\beta}, \gamma), \tag{6}$$

where $p(\beta, \gamma)$ is the joint prior. We suggest using the hierarchical Bayes model. Assuming the parameters are a priori independent, the first level of hierarchy has γ following a gamma distribution with mean m_γ and variance v_γ , and β with multivariate normal distribution with mean vector m_β and covariance matrix $v_\beta I$, where I is the identity matrix. The values of v_γ and v_β are fixed, and for m_γ and m_β we consider a prior, that is $p(m_\gamma) = p(m_\beta) \propto 1$. Arguably, we can use the mode of the integrated likelihood of (m_γ, m_β) to determine a prior distribution [23]. The hyper-parameters v_γ and v_β are viewed as prior precision parameters. The EM (Expectation–Maximization) algorithm [29] can be used to obtain the estimates of m_γ and m_β . The MCMC procedure is used to generate a sample from the posterior distribution. For the MCMC procedure, we used a Gibbs sampler with Metropolis–Hasting. The convergence of the chain was monitored using ergodic means. We omit the details about these computational tools because they are already well known tools and are not the main subject of the present paper. In addition, it was not necessary to develop any special scheme to sample from the posterior chain.

Another advantage of the Weibull link is that the posterior distributions are proper even when we use a wide range of non-informative priors. The Jeffreys’ prior for the parameter β has the form $p(\beta | \gamma) \propto |I(\beta | \gamma)|^{1/2}$, where the Fisher information matrix $I(\beta | \gamma)$ can be obtained by taking the expectation of the Hessian matrix given in Appendix A.

Considering the improper prior $p(\beta) \propto 1$, and the non-informative prior $p(\gamma) \propto 1/\gamma^c$, for $\gamma > 1$ and $c > 1$ a known constant [30], we have the non-informative prior distribution

$$p(\beta, \gamma) \propto p(\beta)p(\gamma) \propto \frac{1}{\gamma^c}. \tag{7}$$

With this constraint (in the parameter γ of the Weibull link), the skewness lies in the interval $(-1.1395, 2]$, which is still a flexible link. For the improper prior of Equation (7), the propriety of the resulting posterior distribution in Equation (6) is stated in Theorem 1.

Theorem 1. Let $z_i = -1$ when $y_i = 0$ and $z_i = 1$ when $y_i = 1$, and X^* be the matrix with rows $z_i x'_i$. Suppose that the design matrix X is of full rank, and there exists a positive vector $\mathbf{a} = (a_1, \dots, a_n)' \in \mathbb{R}^n$, with $a_i > 0$, for $i = 1, \dots, n$, such that $X^{*t} \mathbf{a} = 0$, under the non-informative prior of Equation (7), then the posterior density Equation (6) is proper.

Proof. Let u, u_1, \dots, u_n be independent random variables with common Weibull distribution with shape parameter γ . For $0 < k < \infty$, we have that $E(|u|^k) = \Gamma(1 + k/\gamma) < \infty$. Observing that $1 - F(x) = E[\mathbb{I}(u > x)]$ and $F(x) = E[\mathbb{I}(u \leq x)]$, where \mathbb{I} is an indicator function. Then, we have $[F(x'_i \beta)]^{y_i} [1 - F(x'_i \beta)]^{1-y_i} \leq E(z_i u_i \geq z_i x'_i \beta)$ and $[F(x'_i \beta)]^{y_i} [1 - F(x'_i \beta)]^{1-y_i} \geq E(z_i u_i > z_i x'_i \beta)$. Let $\mathbf{u}^* = (z_1 u_1, \dots, z_n u_n)$. By the Fubini’s theorem, we get

$$\begin{aligned} & \int_1^\infty \int_{\mathbb{R}^k} L(\beta, \gamma | \mathbf{yX}) \frac{1}{\gamma^c} d\beta d\gamma \\ &= \int_1^\infty \frac{1}{\gamma^c} \int_{\mathbb{R}^n} \int_{\mathbb{R}^k} \mathbb{I}(z_i u_i > z_i x'_i \beta, 1 \leq i \leq n) d\beta dF(\mathbf{u}) d\gamma \\ &= \int_1^\infty \frac{1}{\gamma^c} \int_{\mathbb{R}^n} \int_{\mathbb{R}^k} \mathbb{I}(X^* \beta \leq \mathbf{u}^*) d\beta dF(\mathbf{u}) d\gamma. \end{aligned}$$

From Lemma 4.1 of Chen and Shao [31] there exists a constant K depending only on X^* such that

$$\int_{\mathbb{R}^k} \mathbb{I}(X^* \beta \leq \mathbf{u}^*) d\beta \leq K \|\mathbf{u}^*\|^k,$$

which yields

$$\int_1^\infty \int_{\mathbb{R}^k} L(\beta, \gamma | \mathbf{yX}) \frac{1}{\gamma^c} d\beta d\gamma < \infty,$$

by $E(|u|^k) < \infty$, and $\int_1^\infty 1/\gamma^c d\gamma < \infty$ for $c > 1$. \square

This prior give a constraint in the parameter γ . However, any proper prior can be used with the proposed model, avoiding any constraint problem in the parameter γ .

3.2. Multinomial Data

For multinomial responses, we have that $Y_i \in \{1, \dots, \mathcal{K}\}$, and $p_k = \Pr(Y_i = k)$, for $k = 1, \dots, \mathcal{K}$ and $\sum_{j=1}^{\mathcal{K}} p_j = 1$. The logistic multinomial regression model consider a reference category, generally the category \mathcal{K} , and have a link function

$$p_k = g_k^{-1}(\eta_k) = \frac{\exp(\eta_k)}{1 + \sum_{k=1}^{\mathcal{K}-1} \exp(\eta_k)}, \quad k = 1, \dots, \mathcal{K} - 1, \quad \text{and} \quad p_{\mathcal{K}} = 1 - \sum_{k=1}^{\mathcal{K}-1} p_k = \frac{1}{1 + \sum_{k=1}^{\mathcal{K}-1} \exp(\eta_k)},$$

where $\eta_k = \beta_k X$, $\beta_k = \{\beta_{k0}, \beta_{k1}, \dots, \beta_{kr}\}$. The likelihood function for multinomial response data \mathcal{D} is

$$L(\mathbf{p}|\mathcal{D}) \propto \prod_{i=1}^n \prod_{k=1}^{\mathcal{K}} p_k^{\mathbb{I}(y_i=k)} = \prod_{k=1}^{\mathcal{K}} p_k^{s_k}, \tag{8}$$

where $s_k = \sum_{i=1}^n \mathbb{I}(y_i = k)$, and \mathbb{I}_A is the indicator function of event A , that is $\mathbb{I}_A = 1$ and $\mathbb{I}_{A^c} = 0$.

Note that $\sum_{k=1}^{\mathcal{K}} s_k = n$.

Using a reparameterization [32] of \mathbf{p} as $p_1 = \theta_1$, $p_k = \theta_k \prod_{\ell=1}^{k-1} (1 - \theta_\ell)$, for $k = 1, \dots, \mathcal{K} - 1$, and $p_{\mathcal{K}} = \prod_{\ell=1}^{\mathcal{K}-1} (1 - \theta_\ell)$ the likelihood function in Equation (8) can be rewritten as

$$L(\boldsymbol{\theta}|\mathcal{D}) \propto \prod_{k=1}^{\mathcal{K}-1} \theta_k^{s_k} (1 - \theta_k)^{n - \sum_{\ell=1}^k s_\ell} = \prod_{k=1}^{\mathcal{K}-1} L(\theta_k|\mathcal{D}). \tag{9}$$

This shows that the estimation for multinomial data is equivalent to estimating $\mathcal{K} - 1$ binomial response models. We can consider any link function for binary data, taking $\theta_k = g^{-1}(\eta_k)$. For the MLE, we have $\hat{p}_1 = \hat{\theta}_1$, $\hat{p}_k = \hat{\theta}_k \prod_{\ell=1}^{k-1} (1 - \hat{\theta}_\ell)$, and $\hat{p}_{\mathcal{K}} = \prod_{\ell=1}^{\mathcal{K}-1} (1 - \hat{\theta}_\ell)$. For Bayesian estimation, we generate a sample from the posterior distribution of each θ_k , then we can do the transformation to obtain the estimators of \mathbf{p} . Considering the Weibull link function, we need to generate a sample from the posterior of γ_k and β_k , for each $k = 1, \dots, \mathcal{K} - 1$, and then perform the proper transformation to obtain the sample from the posterior of θ_k . In this case, the prior of θ_k can be viewed as a transformation of the priors of γ_k and β_k . Thus, for both MLE and Bayesian estimator, we can use the procedures described in Section 3.1. The partition scheme presented to solve the multinomial model estimation is intuitive. For more details about the reparameterization used here, see Pereira and Stern [32].

3.3. Model Selection and Diagnostics

In the case of binomial data, to compare models within frequentist set up, we use the Akaike Information Criterion (AIC) [33] and the Bayesian Information Criterion (BIC) [34]. For Bayesian analysis, we use long established tool of Deviance Information Criterion (DIC) [35]. We omit the details of these popular tools for the sake of brevity. In addition, for Bayesian analysis we use the $\Pr(\mathcal{D}|M)$ [36], where \mathcal{D} is the observed data and M is the used model. $\Pr(\mathcal{D}|M)$ is approximated by $\{1/m \sum_{i=1}^m \Pr(\mathcal{D}|M, \theta_i)^{-1}\}^{-1}$, where θ_i is the i -th sample from the posterior distribution of θ under model M , given the data \mathcal{D} . This measure is directly related to the Bayes Factor (BF). If the interest is to evaluate the BF_{10} of the model M_1 against M_0 , considering $\Pr(M_0) = \Pr(M_1) = 0.5$, we have

that $BF_{10} = \Pr(\mathcal{D}|M_1)/\Pr(\mathcal{D}|M_0)$. For both Bayesian and frequentist paradigms, we use: a version of Kolmogorov–Smirnov statistics (KS) as measure of goodness of fit (KS is defined as $KS = \max_i |y_i - \hat{y}_i|$, the maximum absolute error of the predicted and the observed frequencies, where \hat{y}_i is the predicted value of y_i); the Mean Absolute Error (MAE) defined as $MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$; and the Brier Score (B-S) defined by Brier [37].

4. Data Example

In the data examples, we compare the proposed link function with some others links. Table 1 presents the link functions considered.

Table 1. Link functions.

| | Link Function | Parametric Space |
|-----------------------|--|-------------------------------------|
| Weibull | $g^{-1}(\eta) = 1 - \exp(-(\eta^\gamma))$ | $\gamma > 0, \eta > 0$ |
| reflected Weibull | $g^{-1}(\eta) = \exp(-(\eta^\gamma))$ | $\gamma > 0, \eta > 0$ |
| AEP | if $\eta \leq 0$, take $\omega = (-2\eta\Gamma(1 + 1/\theta_1))^{\theta_1}$ $g^{-1}(\eta) = \eta \exp(-\omega)/4 + F_G(\omega, 1 + 1/\theta_1, 1)/2$ if $\eta > 0$, take $\omega = (2\eta\Gamma(1 + 1/\theta_2))^{\theta_2}$ $g^{-1}(\eta) = 0.5 + \eta \exp(-\omega)/4 + F_G(\omega, 1 + 1/\theta_2, 1)/2$ | $\theta_1 > 0$ and $\theta_2 > 0$ |
| Aranda–Ordaz | $g^{-1}(\eta) = 1 - (\alpha \exp(\eta) + 1)^{-1/\alpha}$ | $\alpha > 0$ |
| complementary log–log | $g^{-1}(\eta) = 1 - \exp(-\exp(\eta))$ | |
| log–log | $g^{-1}(\eta) = \exp(-\exp(\eta))$ | |
| logit | $g^{-1}(\eta) = \frac{\exp(\eta)}{1 + \exp(\eta)}$ | |
| Prentice | $g^{-1}(\eta) = F_B\left(\frac{1}{1 + \exp\{-\eta\}}, \lambda_1, \lambda_2\right)$ | $\lambda_1 > 0$ and $\lambda_2 > 0$ |
| probit | $g^{-1}(\eta) = F_N(\eta)$ | |
| skew-probit | $g^{-1}(\eta) = F_{SN}(\eta, \delta)$ | $\delta \in (-1, 1)$ |
| Stukel | if $\eta > 0$ and $\alpha_1 > 0$, $h(\eta) = \frac{\exp(\alpha_1\eta) - 1}{\alpha_1}$ if $\eta > 0$ and $\alpha_1 = 0$, $h(\eta) = \eta$ if $\eta > 0$ and $\alpha_1 < 0$, $h(\eta) = \frac{-\log(1 - \alpha_1\eta)}{\alpha_1}$ if $\eta < 0$ and $\alpha_2 > 0$, $h(\eta) = \frac{-\exp(-\alpha_2\eta) - 1}{\alpha_2}$ if $\eta < 0$ and $\alpha_2 = 0$, $h(\eta) = \eta$ if $\eta < 0$ and $\alpha_2 < 0$, $h(\eta) = \frac{\log(1 + \alpha_2\eta)}{\alpha_2}$ $g^{-1}(\eta) = \frac{1}{1 + \exp\{-h(\eta)\}}$ | |

AEP is the asymmetric exponential power link from Naranjo et al. [17]; $\Gamma(\cdot)$ is the mathematical gamma function; $F_G(\cdot, a, b)$ is the distribution function of a random variable with distribution Gamma with shape a and scale b ; $F_B(\cdot, \lambda_1, \lambda_1)$ is the distribution function of a random variable with distribution Beta(λ_1, λ_2); $F_N(\cdot)$ is the distribution function of a random variable with distribution normal, with mean zero and variance 1; and $F_{SN}(\cdot, \delta)$ is the distribution function of a random variable with distribution skew-normal, with mean zero, variance 1 and asymmetric parameter δ [38].

4.1. Binary Data Example

We analyze the study of relative potency of three different poisons: Rotenone, Deguelin and Mixture [24]. The experiment was to test the different poisons with different doses, with objective to understand the potency of the poisons. Five doses for rotenone, six doses for deguelin and six doses for the mixture were considered. For each dose and poison, around 50 insects were considered by observing how many insects were killed. The data are presented in Table 2. We consider that the response variable is binary with $Y = 1$ representing the insect killed, and as covariates: X_1 as the log (Dose), X_2 as an indicator of Rotonone, and X_3 as an indicator of Deguelin. The mixture of poisons

is considered as the reference poison (that is, $X_2 = 0$ and $X_3 = 0$). Our main objective is to find the model that better represents (fits) these Data. We are not looking for the “best” poison or dose.

Table 2. Relative potency of Rotenone, a Deguelin concentrate, and a Mixture of two.

| Rotenone | | | Deguelin | | | Mixture | | |
|------------|------|----------|------------|------|----------|------------|------|----------|
| log (Dose) | Dead | <i>n</i> | log (Dose) | Dead | <i>n</i> | log (Dose) | Dead | <i>n</i> |
| 1.01 | 44 | 50 | 1.70 | 48 | 48 | 1.40 | 48 | 50 |
| 0.89 | 42 | 49 | 1.61 | 47 | 50 | 1.31 | 43 | 46 |
| 0.71 | 24 | 46 | 1.48 | 47 | 49 | 1.18 | 38 | 48 |
| 0.58 | 16 | 48 | 1.31 | 34 | 48 | 1.00 | 27 | 46 |
| 0.41 | 6 | 50 | 1.00 | 18 | 48 | 0.71 | 22 | 46 |
| - | - | - | 0.71 | 16 | 49 | 0.40 | 7 | 47 |

We obtain the MLE for Weibull parameters and for comparison we also estimated the parameters of complementary log–log, Stukel, probit, logit, Aranda–Ordaz, and Prentice models. Table 3, presents some statistics of each model to compare them. The best models, based on AIC, are complementary log–log and Weibull. The advantage of the complementary log–log is that this model has one fewer parameter. However, $\hat{\gamma} = 114.5084$ (Table 4), indicating that the Weibull model is going to the complementary log–log model. As expected, the Weibull model performs similar to the complementary log–log (see Proposition 1). The model with lowest KS is the logit model. The estimated logit and Weibull links are illustrated in Figure 4. The model with lowest MAE is the Stukel model. The estimated parameter values of Weibull logit, and Stukel models are given in Table 4.

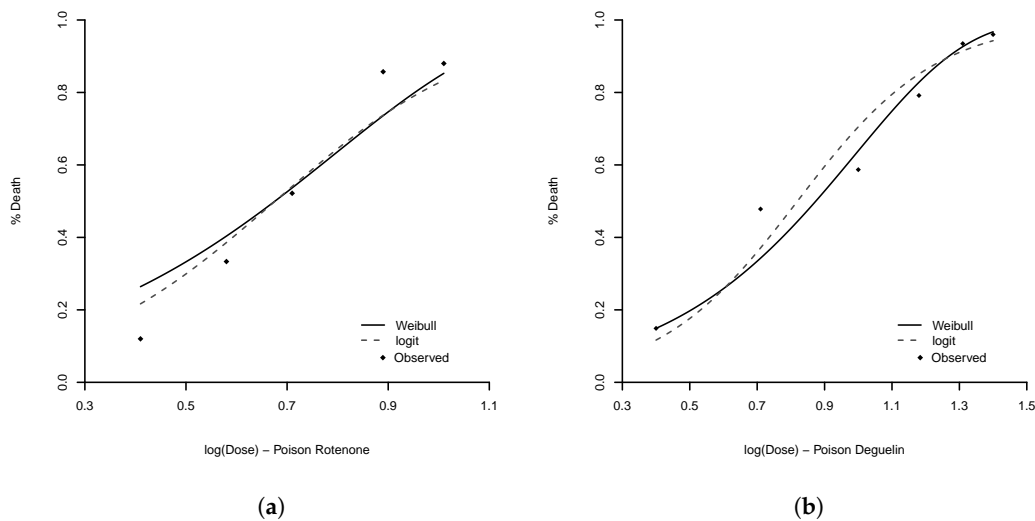
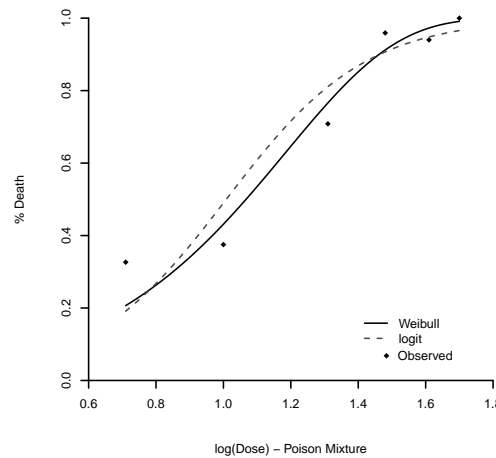


Figure 4. Cont.



(c)

Figure 4. Comparison of MLE for Weibull (solid line) and logit link (dashed line) for three types of poisons: (a) Rotonone; (b) Deguelin; and (c) Mixture. The dots are the observed proportions.

Table 3. Comparison of the link functions under MLE.

| Model | log(L) | AIC | BIC | KS | MAE | B-S |
|---------------|---------|---------------|---------------|---------------|---------------|---------------|
| Stukel | −369.51 | 751.01 | 779.26 | 0.1704 | 0.0482 | 0.1474 |
| probit | −369.66 | 751.32 | 779.56 | 0.1583 | 0.0492 | 0.1475 |
| comp. log–log | −370.32 | 748.66 | 767.48 | 0.1451 | 0.0551 | 0.1476 |
| Weibull | −370.34 | 750.69 | 774.22 | 0.1440 | 0.0553 | 0.1477 |
| logit | −372.57 | 753.14 | 771.97 | 0.1292 | 0.0656 | 0.1486 |
| Aranda–Ordaz | −373.41 | 754.82 | 773.65 | 0.1351 | 0.0668 | 0.1487 |
| Prentice | −374.90 | 759.80 | 783.33 | 0.3674 | 0.1396 | 0.1750 |

Table 4. MLE for the binomial example.

| Parameter | Model [Estimate (SE)] | | |
|-----------|-------------------------------|------------------|-------------------------------|
| | Weibull | Logit | Stukel |
| β_0 | 0.9735 (0.0110) | −3.9559 (0.3546) | −5.1973 (2.0524) |
| β_1 | 0.0266 (0.0111) | 4.8273 (0.3394) | 5.5892 (1.9677) |
| β_2 | 0.0053 (0.0024) | 0.6910 (0.2308) | 1.3233 (0.6397) |
| β_3 | −0.0051 (0.0024) | −0.9125 (0.2449) | −1.0658 (0.4397) |
| - | $\gamma = 114.5084 (47.9818)$ | - | $\alpha_1 = 0.1732 (0.2871)$ |
| - | - | - | $\alpha_2 = -0.9663 (0.8492)$ |

Another important models are the skew-probit proposed by Chen et al. [5] and AEP proposed by Naranjo et al. [17]. To compare with the Weibull model we perform a Bayesian analysis for these models. The priors for the parameters of the Weibull model are the same as that described in Section 3.1. We use the values $v_\gamma = 100$ and $v_\beta = 25$. The estimated values for the hyper-parameters of first hierarchical level are $\widehat{m}_\gamma = 9.1089$ and $\widehat{m}_\beta = (0.1588, 0.8879, 0.1261, -0.1717)$. For the priors of skew-probit model, we used a uniform distribution over the interval $(-1, 1)$ for the asymmetry parameter and independent normal distribution with mean 0 and variance 25 for each $\beta_j, j = 0, \dots, 3$. For the priors of AEP model, we used independent priors: gamma distribution with mean 1 and variance 100 for the parameters θ_1 and θ_2 , and normal distribution with mean 0 and variance 25 for each $\beta_j, j = 0, \dots, 3$. Table 5 presents the model selection criteria (DIC, KS, MAE, B-S and $\Pr(\mathcal{D}|M)$)

for the three competing models. For criteria DIC, KS, MAE and B-S, a smaller value indicates a better agreement between the model and observed data. For criterion $\Pr(\mathcal{D}|M)$, a higher value indicates a better agreement between the model and observed data. We note that the Weibull model has better KS; AEP has the better DIC and MAE; and skew-probit has better $\Pr(\mathcal{D}|M)$. The B-S was very similar for all models. For the three models, the posterior mean of relevant parameters are given in Table 6.

Table 5. Comparison of the link functions under Bayesian estimation.

| | DIC | KS | MAE | B-S | $\Pr(\mathcal{D} M)$ |
|-------------|---------------|---------------|---------------|---------------|----------------------|
| AEP | 749.01 | 0.1826 | 0.0564 | 0.1472 | 744.66 |
| Weibull | 751.38 | 0.1278 | 0.0644 | 0.1484 | 748.42 |
| skew-probit | 751.92 | 0.1432 | 0.0662 | 0.1486 | 749.13 |

Table 6. Bayesian estimates for binomial example.

| Parameter | Model [Posterior Mean (Standard Deviation)] | | |
|-----------|---|-----------------------------|------------------------------|
| | Weibull | Skew-Probit | AEP |
| β_0 | 0.2661 (0.2022) | −2.1900 (0.4006) | −4.9790 (1.5058) |
| β_1 | 0.7799 (0.2332) | 2.6464 (0.2226) | 5.2736 (1.4799) |
| β_2 | 0.1152 (0.0402) | 0.3756 (0.1258) | 1.2292 (0.4575) |
| β_3 | −0.1474 (0.0576) | −0.4982 (0.1309) | −1.0145 (0.3830) |
| - | $\gamma = 4.0285 (1.1992)$ | $\delta = -0.0434 (0.5506)$ | $\theta_1 = 0.4491 (0.1152)$ |
| - | - | - | $\theta_2 = 0.9057 (0.1904)$ |

4.2. Multinomial Data Example

Grazeffe et al. [25] reported a study of DNA mutation of the cells of adult snails, each irradiated with a single dose of gamma radiation. They recorded four categories of DNA mutation with $Y = 1, 2, 3$ and 4 representing no mutation, low, intermediate, and high DNA mutation respectively. The snails are randomized into five different dose levels with $X \in \{0, 2.5, 5, 10, 20\}$. The data are presented in Table 1 of Grazeffe et al. [25]. The objective is to compare effects of different dose levels on DNA mutation Y ($Y = 1$ for C0, $Y = 2$ for C1, $Y = 3$ for C2 and $Y = 4$ for C3). We illustrate the use of Weibull link model, under frequentist approach, for analysis of this study with multinomial responses. Further, in Table 10, we compare our estimates of $\Pr[Y = k | x]$ with those obtained by Grazeffe et al. [25] based on the logit link model, and the other models discussed here.

For a proper comparison with previous method of Grazeffe et al. [25], we obtain the MLE with only X and X^2 as covariates. We use the reflected Weibull link, because this model has lowest values of KS and MAE than those for Weibull link. To obtain the estimation of the reflected Weibull model we first estimate the values of $\theta_1, \theta_2, \theta_3$. To simplify, consider the three binary variables Z_1, Z_2 and Z_3 , where $\theta_k = \Pr(Z_k = k), k = 1, 2, 3$. Then, using the results in Section 3.2, we construct Table 7 with the observed values of Z_s , and estimate the models for Z_s .

Table 7. Observed values of constructed variables Z_1, Z_2 and Z_3 .

| Dose (X) | Z_1 | | Z_2 | | Z_3 | |
|----------|-------|-----|-------|-----|-------|-----|
| | 0 | 1 | 0 | 1 | 0 | 1 |
| 0 | 446 | 654 | 321 | 125 | 249 | 72 |
| 2.5 | 458 | 442 | 280 | 178 | 175 | 105 |
| 5 | 703 | 197 | 450 | 253 | 277 | 173 |
| 10 | 841 | 159 | 545 | 296 | 281 | 264 |
| 20 | 842 | 58 | 793 | 49 | 660 | 133 |

The parameter estimates for the three binary models are presented in Table 8, and we have $\hat{\theta}_1(x) = e^{-(0.0234 - 1.6395x + 0.6748x^2)^{0.1742}}$, $\hat{\theta}_2(x) = e^{-(1.0930 - 0.0368x + 0.0030x^2)^{2.3604}}$ and $\hat{\theta}_3(x) = e^{-(1.2429 - 0.0866x + 0.0047x^2)^{1.7562}}$.

Table 8. MLE of Weibull model for the multinomial example.

| Parameter | Model [Estimate (SE)] | | |
|-----------|-------------------------|-------------------------|-------------------------|
| | Y = 1 (Z ₁) | Y = 2 (Z ₂) | Y = 3 (Z ₃) |
| γ | 0.1742 (0.0209) | 2.3604 (0.1905) | 1.7562 (0.1415) |
| β_0 | 0.0234 (0.0123) | 1.0930 (0.0258) | 1.2429 (0.0369) |
| β_1 | -1.6395 (0.8355) | -0.0368 (0.0067) | -0.0866 (0.0078) |
| β_2 | 0.6748 (0.3295) | 0.0030 (0.0002) | 0.0047 (0.0003) |

As described in Section 3.2, we have $\hat{p}_1(x) = \hat{\theta}_1(x)$, $\hat{p}_2(x) = [1 - \hat{\theta}_1(x)]\hat{\theta}_2(x)$, $\hat{p}_3(x) = [1 - \hat{\theta}_1(x)][1 - \hat{\theta}_2(x)]\hat{\theta}_3(x)$, and $\hat{p}_4(x) = [1 - \hat{\theta}_1(x)][1 - \hat{\theta}_2(x)][1 - \hat{\theta}_3(x)]$, where $\hat{p}_k(x)$ is the estimated value of $\Pr[Y = k|x]$, $k = 1, \dots, 4$.

Table 9 presents the inferential statistics for model comparisons. All statistics indicate a preference for Weibull link model. The main difference for the Stukel model was because Weibull model has three parameters fewer than the Stukel model in this multinomial example.

Table 9. Comparison of the link functions for multinomial example.

| | log(L) | AIC | BIC | KS | MAE | B-S |
|---------------|-----------|----------|----------|-------|--------|--------|
| Weibull | -5654.224 | 11332.45 | 11410.16 | 0.030 | 0.0095 | 0.6324 |
| Stukel | -5654.961 | 11339.92 | 11437.07 | 0.031 | 0.0100 | 0.6325 |
| Prentice | -5667.156 | 11364.31 | 11461.46 | 0.066 | 0.0158 | 0.6341 |
| Aranda-Ordaz | -5671.681 | 11361.36 | 11419.65 | 0.733 | 0.3116 | 1.1898 |
| logit | -5672.196 | 11362.39 | 11420.68 | 0.075 | 0.0171 | 0.6348 |
| comp. log-log | -5672.799 | 11363.60 | 11421.89 | 0.936 | 0.3129 | 1.2876 |
| probit | -5673.003 | 11364.01 | 11422.29 | 0.079 | 0.0175 | 0.6348 |
| log-log | -5676.075 | 11370.15 | 11428.44 | 0.693 | 0.2099 | 0.9490 |

The estimated frequencies, under Weibull, Stukel and logit models, of DNA mutation for each class is presented in Table 10, and illustrated in Figure 5. This figure shows that the Weibull link model has a better fit for categories $Y = 1$ and 4, when compared with the logit link model. For categories $Y = 2$ and 3, both models have comparable performances. Weibull and Stukel models have similar values of estimated frequencies.

Table 10. Relative frequencies of mutation (observed and model’s estimates).

| Dose (Gy) | Model | Mutation Classes (Y) | | | |
|-----------|----------|----------------------|-------|-------|-------|
| | | 1 | 2 | 3 | 4 |
| 0 | Observed | 0.595 | 0.114 | 0.065 | 0.226 |
| | Weibull | 0.595 | 0.120 | 0.065 | 0.220 |
| | Logit | 0.606 | 0.114 | 0.064 | 0.215 |
| | Stukel | 0.594 | 0.120 | 0.065 | 0.221 |
| 2.5 | Observed | 0.491 | 0.198 | 0.117 | 0.194 |
| | Weibull | 0.491 | 0.178 | 0.112 | 0.219 |
| | Logit | 0.430 | 0.201 | 0.123 | 0.246 |
| | Stukel | 0.490 | 0.178 | 0.113 | 0.218 |
| 5 | Observed | 0.214 | 0.281 | 0.192 | 0.308 |
| | Weibull | 0.233 | 0.289 | 0.200 | 0.278 |
| | Logit | 0.289 | 0.273 | 0.183 | 0.255 |
| | Stukel | 0.233 | 0.289 | 0.201 | 0.277 |
| 10 | Observed | 0.159 | 0.296 | 0.264 | 0.281 |
| | Weibull | 0.137 | 0.302 | 0.264 | 0.296 |
| | Logit | 0.136 | 0.298 | 0.268 | 0.298 |
| | Stukel | 0.135 | 0.304 | 0.263 | 0.298 |
| 20 | Observed | 0.064 | 0.054 | 0.148 | 0.733 |
| | Weibull | 0.075 | 0.054 | 0.147 | 0.725 |
| | Logit | 0.067 | 0.055 | 0.148 | 0.730 |
| | Stukel | 0.077 | 0.054 | 0.146 | 0.723 |

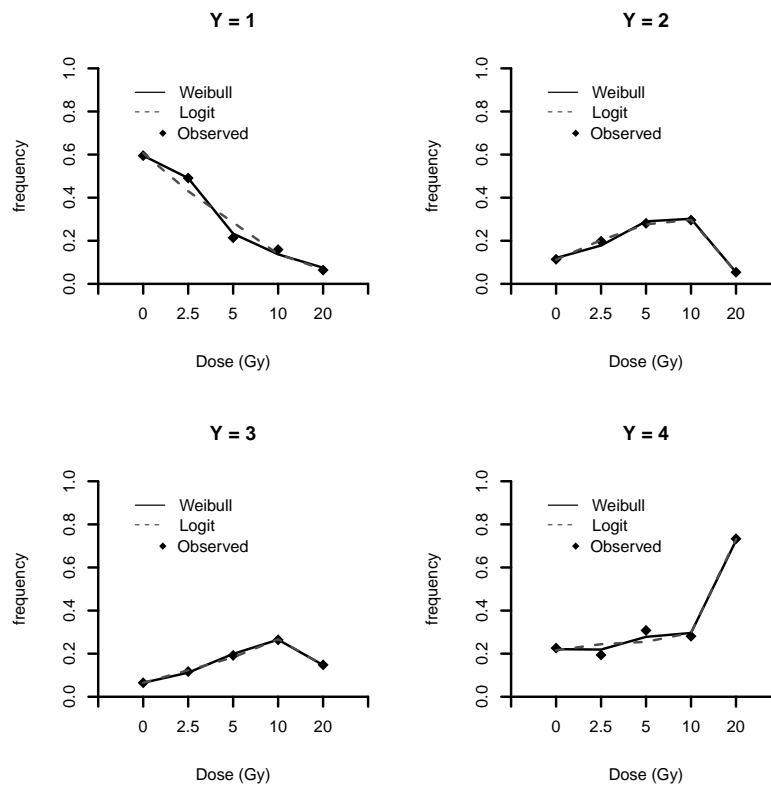


Figure 5. Estimated population frequencies.

5. Final Comments

In this paper, we have presented a Weibull model to estimate the problem of binary and multinomial regression analysis. The model is very flexible and capable to handle with many different types of data. The comparison with other skew-link model, in binomial data example (Section 4.1), shows that the performance of the Weibull link was good when compared to the others models. The model with worst measures was the Prentice model. All others had an equivalent result. We are convinced that our proposed model is a good option. A good feature of the model is that the logit, probit, complementary log–log, and log–log link functions are approximations of Weibull link. Then, the proposed model can accommodate even symmetric link function. For the flexibility of the Weibull link model, we are comfortable to suggest its use in practice.

Other aspect of the proposed Weibull model is that the associated numerical procedure of MLE is very simple to implement, particularly in comparison to other competing. For Bayesian estimates, we also suggest an Empirical Bayes approach to determine the prior. Under full Bayesian estimation, we compare the model with the skew-probit model [5] and AEP model [17], in Section 4.1. Again, all models had similar results, however the KS of Weibull model were the measures with the greatest differences among all models. The performance of our model was good, even under full Bayesian framework, in binomial data example (Section 4.1).

We also develop a partition scheme for the multinomial regression model simplifying the problem to $\mathcal{K} - 1$ binomial regression analysis. This is a general scheme that can be used for other link functions, which opens a vast options to estimate multinomial data. In Section 4.2, we analyze a multinomial data problem, where the Weibull model had the best measure values when compared with all other models. Our perspective is that the Weibull model is a good option for binary/multinomial regression, mainly due to its simplicity. We have analytic form for the link function, as well as for the gradient and Hessian matrix.

Acknowledgments: Research work of this article was partially supported by grants from the Pfeiffer Cancer Research Foundation and National Cancer Institute (R03CA205018-01) of USA, and from National Council for Scientific and Technological Development (308776/2014-3) of Brazil. The agencies had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author Contributions: The authors contributed equally to this work. All authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Gradient and Hessian for Log-Likelihood of Weibull Model

Let $\xi_{i1} = \exp\{-\eta_i^\gamma\}$, $\xi_{i2} = \exp\{-2\eta_i^\gamma\}$, $i = 1, \dots, n$.

Appendix A.1. Gradient

The gradient vector is $g = (g_1, g_2, \dots, g_r)$, where

$$g_1 = \frac{\partial l(\gamma, \beta | \mathcal{D})}{\partial \gamma} = \sum_{i=1}^n -(1 - y_i)\eta_i^\gamma \log(\eta_i) + \frac{y_i \xi_{i1} \eta_i^\gamma \log(\eta_i)}{1 - \xi_{i1}};$$

$$g_2 = \frac{\partial l(\gamma, \beta | \mathcal{D})}{\partial \beta_0} = \sum_{i=1}^n -\gamma(1 - y_i)\eta_i^{\gamma-1} + \frac{\gamma y_i \xi_{i1} \eta_i^{\gamma-1}}{1 - \xi_{i1}};$$

for $j = 1, \dots, r - 2$,

$$g_{(j+2)} = \frac{\partial l(\gamma, \beta | \mathcal{D})}{\partial \beta_j} = \sum_{i=1}^n -\gamma x_{ij}(1 - y_i)\eta_i^{\gamma-1} + \frac{\gamma x_{ij} y_i \xi_{i1} \eta_i^{\gamma-1}}{1 - \xi_{i1}}.$$

Appendix A.2. Hessian Matrix

The Hessian matrix H is

$$H = \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1r} \\ h_{21} & h_{22} & \dots & h_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ h_{r1} & h_{r2} & \dots & h_{rr} \end{bmatrix},$$

where, for $j = 1, \dots, r - 2$,

$$h_{11} = \frac{\partial^2 l(\gamma, \beta | \mathcal{D})}{\partial \gamma \partial \gamma} = \sum_{i=1}^n -(1 - y_i)[\log(\eta_i)]^2 \eta_i^\gamma + \frac{\{\xi_{i1} y_i [\log(\eta_i)]^2\} [\eta_i^\gamma - \eta_i^{2\gamma}]}{1 - \xi_{i1}} - \frac{\xi_{i2} y_i [\log(\eta_i)]^2 \eta_i^{2\gamma}}{(1 - \xi_{i1})^2};$$

$$h_{21} = \frac{\partial^2 l(\gamma, \beta | \mathcal{D})}{\partial \gamma \partial \beta_0} = \frac{\partial^2 l(\gamma, \beta | \mathcal{D})}{\partial \beta_0 \partial \gamma} = h_{12};$$

$$h_{(j+2)1} = \frac{\partial^2 l(\gamma, \beta | \mathcal{D})}{\partial \gamma \partial \beta_j} = \frac{\partial^2 l(\gamma, \beta | \mathcal{D})}{\partial \beta_j \partial \gamma} = h_{1(j+2)};$$

$$h_{12} = \frac{\partial^2 l(\gamma, \beta | \mathcal{D})}{\partial \beta_0 \partial \gamma} = \sum_{i=1}^n -[1 + \gamma \log(\eta_i)](1 - y_i)\eta_i^{\gamma-1} + \frac{[1 + (1 - \eta_i^\gamma)\gamma \log(\eta_i)]\xi_{i1}y_i\eta_i^{\gamma-1}}{1 - \xi_{i1}} - \frac{\xi_{i2}\gamma y_i \log(\eta_i)\eta_i^{2\gamma-1}}{(1 - \xi_{i1})^2};$$

$$h_{22} = \frac{\partial^2 l(\gamma, \beta | \mathcal{D})}{\partial \beta_0 \partial \beta_0} = \sum_{i=1}^n -(\gamma - 1)\gamma(1 - y_i)\eta_i^{\gamma-2} + \frac{[(\gamma - 1) - \gamma\eta_i^\gamma]\xi_{i1}\gamma y_i\eta_i^{\gamma-2}}{1 - \xi_{i1}} - \frac{\xi_{i2}\gamma^2 y_i\eta_i^{2\gamma-2}}{(1 - \xi_{i1})^2};$$

$$h_{(j+2)2} = \frac{\partial^2 l(\gamma, \beta | \mathcal{D})}{\partial \beta_0 \partial \beta_j} = \sum_{i=1}^n -(\gamma - 1)\gamma x_{ij}(1 - y_i)\eta_i^{\gamma-2} + \frac{[(\gamma - 1) - \gamma\eta_i^\gamma]\xi_{i1}\gamma x_{ij}y_i\eta_i^{\gamma-2}}{1 - \xi_{i1}} - \frac{\xi_{i2}\gamma^2 x_{ij}y_i\eta_i^{2\gamma-2}}{(1 - \xi_{i1})^2};$$

$$h_{1(j+2)} = \frac{\partial^2 l(\gamma, \beta | \mathcal{D})}{\partial \beta_j \partial \gamma} = \sum_{i=1}^n -[x_{ij}(1 - y_i)\eta_i^{\gamma-1}][1 + \gamma \log(\eta_i)] + \frac{\xi_{i1}x_{ij}y_i\eta_i^{\gamma-1}[1 + \gamma \log(\eta_i)(1 - \eta_i^\gamma)]}{1 - \xi_{i1}} - \frac{\xi_{i2}\gamma x_{ij}y_i \log(\eta_i)\eta_i^{2\gamma-1}}{(1 - \xi_{i1})^2};$$

$$h_{2(j+2)} = \frac{\partial^2 l(\gamma, \beta | \mathcal{D})}{\partial \beta_j \partial \beta_0} = \frac{\partial^2 l(\gamma, \beta | \mathcal{D})}{\partial \beta_0 \partial \beta_j} = h_{(j+2)2};$$

for $k = 1, \dots, r - 2$ and $k \neq j$,

$$h_{(k+2)(j+2)} = \frac{\partial^2 l(\gamma, \beta | \mathcal{D})}{\partial \beta_j \partial \beta_k} = \sum_{i=1}^n -(\gamma - 1)\gamma x_{ij}x_{ik}(1 - y_i)\eta_i^{\gamma-2} + \frac{[(\gamma - 1) - \gamma\eta_i^\gamma]\xi_{i1}\gamma x_{ij}x_{ik}y_i\eta_i^{\gamma-2}}{1 - \xi_{i1}} - \frac{\xi_{i2}\gamma^2 x_{ij}x_{ik}y_i\eta_i^{2\gamma-2}}{(1 - \xi_{i1})^2} = h_{(j+2)(k+2)}.$$

References

1. Agresti, A.; Finlay, B. *Statistical Methods for the Social Sciences*, 4th ed.; Pearson Prentice Hall: Upper Saddle River, NJ, USA, 2009.
2. Agresti, A. *Categorical Data Analysis*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2002.
3. Nelder, J.A.; Wedderburn, R. Generalized Linear Models. *J. R. Stat. Soc. Ser. A* **1972**, *135*, 370–384.
4. McCullagh, P.; Nelder, J.A. *Generalized Linear Models*, 2nd ed.; Chapman and Hall: London, UK, 1989.
5. Chen, M.H.; Dey, D.K.; Shao, Q.M. A New Skewed Link Model for Dichotomous Quantal Response Data. *J. Am. Stat. Assoc.* **1999**, *94*, 1172–1186.
6. Aranda-Ordaz, F.J. On Two Families of Transformations to Additivity for Binary Response Data. *Biometrika* **1981**, *68*, 357–363.
7. Guerrero, V.M.; Johnson, R.A. Use of the Box-Cox Transformation With binary Response Models. *Biometrika* **1982**, *69*, 309–314.
8. Morgan, B.J.T. Observations on Quantitative Analysis. *Biometrics* **1983**, *39*, 879–886.
9. Whittmore, A.S. Transformation to Linearity in Binary Regression. *SIAM J. Appl. Math.* **1983**, *43*, 703–710.
10. Stukel, T. Generalized Logistic Models. *J. Am. Stat. Assoc.* **1988**, *83*, 426–431.
11. Prentice, R. Generalization of the Probit and Logit Models. *Biometrics* **1976**, *32*, 761–768.
12. Pregibon, D. Goodness of Link Test for Generalized Linear Models. *Appl. Stat.* **1980**, *29*, 338–345.
13. Czado, C. On Link Selection in Generalized Linear Models. In *Advances in GLIM and Statistical Modelling*; Lecture Notes in Statistics; Fahrmeir, L., Francis, B., Gilchrist, R., Tutz, G., Eds.; Springer: Berlin/Heidelberg, Germany, 1992; Volume 78, pp. 60–65.
14. Czado, C. Parametric Link Modification of Both Tails in Binary Regression. *Stat. Papers* **1994**, *35*, 189–201.
15. Albert, J.H.; Chib, S. Bayesian Analysis of Binary and Polychotomous Response Data. *J. Am. Stat. Assoc.* **1993**, *88*, 669–679.
16. Kim, S.; Chen, M.H.; Dey, D.K. Flexible generalized t-link models for binary response data. *Biometrika* **2008**, *95*, 93–106.
17. Naranjo, L.; Peres, C.J.; Martin, J. Bayesian analysis of some models that use the asymmetric exponential power distribution. *Stat. Comput.* **2015**, *25*, 497–514.
18. Rubio, F.J.; Liseo, B. On the independence Jeffreys prior for skew-symmetric models. *Stat. Probab. Lett.* **2014**, *85*, 91–97.
19. Bazán, J.L.; Bolfarine, H.; Branco, M.D. A Framework for Skew-Probit Links in Binary Regression. *Commun. Stat. Theory Methods* **2010**, *39*, 678–697.
20. Hosmer, D.W.; Lemeshow, S. *Applied Logistic Regression*, 2nd ed.; John Wiley & Sons: New York, NY, USA, 2000.
21. Caron, R.; Polpo, A. Binary data regression: Weibull distribution. *AIP Conf. Proc.* **2009**, *1193*, 187–193.
22. Robbins, H. An Empirical Bayes Approach to Statistics. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA, USA, December 1954 and July–August 1955; University of California Press: Berkeley, CA, USA, 1956; Volume 1, pp. 157–163.
23. Carlin, B.P.; Louis, T.A. *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd ed.; Chapman & Hall/CRC: London, UK, 2000.
24. Finney, D.J. *Probit Analysis*; University Press: Cambridge, UK, 1947.
25. Grazeffe, V.S.; Tallarico, L.F.; Pinheiro, A.S.; Kawano, T.; Suzuki, M.F.; Okazaki, K.; Pereira, C.A.B.; Nakano, E. Establishment of the Comet Assay in the Freshwater Snail *Biomphalaria Glabrata* (Say, 1818). *Mutat. Res.* **2008**, *654*, 58–63.
26. Arnold, B.C.; Groeneveld, R.A. Measuring skewness with respect to the mode. *Am. Stat.* **1995**, *49*, 34–38.
27. Rinne, H. *The Weibull Distribution: A Handbook*; CRC Press: Boca Raton, FL, USA, 2009.
28. Nelder, J.A.; Mead, R. A Simplex Algorithm for Function Minimization. *Comput. J.* **1965**, *7*, 308–313.
29. McLachlan, G.J.; Krishnan, T. *The EM Algorithm and Extensions*; John Wiley & Sons: Hoboken, NJ, USA, 2008.
30. Sun, D. A note on noninformative priors for Weibull distributions. *J. Stat. Plan. Inference* **1997**, *61*, 319–338.
31. Chen, M.H.; Shao, Q.M. Property of Posterior Distribution for Dichotomous Quantal Response Models with General Link Functions. *Proc. Am. Math. Soc.* **2000**, *129*, 293–302.
32. Pereira, C.; Stern, J. Special Characterizations of Standard Discrete Models. *REVSTAT Stat. J.* **2008**, *6*, 199–230.
33. Akaike, H. A New Look at the Statistical Model Identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723.
34. Schwarz, G.E. Estimating the Dimension of a Model. *Ann. Stat.* **1978**, *6*, 461–464.

35. Spiegelhalter, D.J.; Best, N.G.; Carlin, B.P.; van der Linde, A. Bayesian Measures of Model Complexity and Fit. *J. R. Stat. Soc. Ser. B* **2002**, *64*, 583–639.
36. Kass, R.E.; Raftery, A.E. Bayes Factors. *J. Am. Stat. Assoc.* **1995**, *90*, 773–795.
37. Brier, G.W. Verification of forecasts expressed in terms of probability. *Mon. Weather Rev.* **1950**, *78*, 1–3.
38. Azzalini, A.; Capitanio, A. *The Skew-Normal and Related Families*; Cambridge University Press: Cambridge, UK, 2014.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).