

BAYESIAN GENERALIZED DISSIMILARITY MODEL: THE POSTERIOR MEAN EXISTENCE AND ITS IMPLEMENTATION ON THE COMMUNITY ECOLOGY ANALYSIS

 Suci Astutik¹,  Evellin Dewi Lusiana^{1*},  Nurjannah¹,  Abu Bakar Sambah²

¹Department of Statistics, Faculty of Mathematics and Natural Science, Brawijaya University, Malang, Indonesia

²Faculty of Fisheries and Marine Science, Brawijaya University, Malang, Indonesia

Abstract. The Bayesian Generalized Dissimilarity Model (BGDM) is proposed in this paper for estimating the parameters of the Generalized Dissimilarity Model (GDM) using Bayesian approaches. The logit link function is used to estimate the GDM's response variable, which is characterized as a binomial process. However, the logit function in Generalized Linear Model (GLM) is sensitive to the data separation issue, resulting in the estimator's non-existence. This research was aimed to assess and evaluate the existence of BGDM estimator (posterior mean) under different prior theoretically and empirically. Parameter estimation of BGDM uses the Hamiltonian Monte Carlo (HMC) technique for carrying out the Markov Chain Monte Carlo (MCMC) approach that required for solving the complex integration in Bayesian procedure. The paper presents theoretical and simulation study results that suggest that the posterior mean can be used as the estimator of the BGDM, regardless of its prior distribution, as long as there is no solitary separator in the data. However, if there is any solitary separator in the data, the posterior mean does not exist under a prior distribution that has an undefined mean. Moreover, BGDM also has better estimation results in modelling community ecology data.

Keywords: binomial regression, data separation, logit link function, pair-site dissimilarity, parameter estimation.

AMS Subject Classification: 62-07, 62P12, 62F15, 62C10.

Corresponding author: Evellin Dewi Lusiana, Department of Statistics, Faculty of Mathematics and Natural Science, Brawijaya University, Malang, Indonesia, Tel.: +6285253417952, e-mail: evellinlusiana@ub.ac.id

Received: 13 March 2024; Revised: 28 October 2024; Accepted: 5 December 2024; Published: 10 April 2025.

1 Introduction

Generalized Dissimilarity Model (GDM) is a statistical method for evaluating pairwise biodiversity dissimilarities in ecological study (Ferrier et al., 2007). GDM is expanded from Generalized Linear Model (GLM) by setting the fitted dissimilarity function be increasingly monotonic using I-spline function (Mokany et al., 2022). This is a manifestation of the ecological assumption that widening the distance among locations over an environmental gradient can further increase compositional dissimilarity (Ferrier et al., 2007). Like other GLM methods, the GDM makes use of a link function to analyze the relationship between the expected pair-site dissimilarity and a scaled distance between sites according on various environmental variables. The classic GDM (Ferrier et al., 2007) and Sparse GDM that developed to model sparse dataset (Leitao et al., 2015) used negative binomial link function. Meanwhile, Woolley (2017) proposed Bayesian

How to cite (APA): Astutik, S., Lusiana, E.D., Nurjannah, & Sambah, A.B. (2025). Bayesian generalized dissimilarity model: The posterior mean existence and its implementation on the community ecology analysis. *Advanced Mathematical Models & Applications*, 10(1), 160-173 <https://doi.org/10.62476/amma101160>

Bootstrap GDM (BBGDM) by recognizing that the pair-site dissimilarity within GDM framework follows a binomial process, the logit link function was chosen since it is widely utilized in binomial GLM, particularly in the logistic regression model.

Logistic regression has been widely used in both classical and Bayesian approaches. It has been shown that under a data separation condition, the estimates used in Maximum Likelihood Estimation (MLE) for logistic regression are infinite or do not exist (Albert & Anderson, 1984). Data separation can be described as a condition, for example, if there exists a linear function of the covariates for which the positive values of the function correspond to those units with higher response values, while negative values of the function correspond to units with lower response values (Copson, 2015). In addition, it is worth noting that the existence of a posterior mean in Bayesian analysis is equivalent to the existence of MLE estimates in the classical approach. However, specific prior distribution, such as Cauchy prior, which does not have finite mean, when it is used under the separation condition, then the posterior mean for the logistic regression model will not exist (Copson, 2015).

Previous work of Woolley (2017) focused on the use of resampling technique such as Bayesian Bootstrap to enable the characterization of uncertainty in GDM parameter estimation, because the classic GDM only provide the inferential analysis using permutation test. Thus, there is no concerned yet of using different parameter estimation methods to obtain better estimates results. White et al. (2024) utilized Bayesian framework in GDM's parameter estimation by using weakly informative prior. Meanwhile, in this work, we used Bayesian technique with informative prior for estimating parameter of GDM. Bayesian technique provides significant benefits over the frequentist technique, particularly when the model coefficients are known in advance (Gray et al., 2015). Due to the fact that the GDM framework necessitates an underlying ecological premise (Ferrier et al., 2007), thus the Bayesian technique is appropriate for this kind of model. By utilizing a non-negative informative prior distribution, the monotonic fitted function condition for the GDM could be met. Furthermore, the Bayesian GDM (BGDM) estimators had shown to have better performance compared to those of GDM's (Lusiana et al., 2023). Nonetheless, the issue of data separation should be investigated for the GDM, just as it is in logistic regression. Therefore, we aim to examine the influence of data separation on the BGDM's posterior mean theoretically and empirically as well as its implementation in community ecology problems.

2 Materials and Methods

2.1 GDM Likelihood Function

GDM considers fitting the nonlinearity of compositional dissimilarity between sites u and v as the I-spline function of environmental variables which denoted by η_{uv} as follows (1)

$$\eta_{uv} = \beta_0 + \sum_{j=1}^p \sum_{l=1}^{k_j} \beta_{jl} |I_{jl}(x_{ju}) - I_{jl}(x_{jv})| \quad (1)$$

where p is the number of environmental variables, k_j is the number of spline knots for j -th variable, and $I_{jl}(\cdot)$ represents the I-spline function of j -th variable and l -th knot. Then, suppose that $\tilde{x}_{uv,jl} = |I_{jl}(x_{ju}) - I_{jl}(x_{jv})|$, thus formula (1) can be written as formula (2):

$$\eta_{uv} = \beta_0 + \sum_{j=1}^p \sum_{l=1}^{k_j} \beta_{jl} \tilde{x}_{uv,jl} \quad (2)$$

Suppose there are n pair-sites, where $i = 1, 2, \dots, n$, $\boldsymbol{\beta}_t = (\beta_0, \beta_{11}, \dots, \beta_{pk_j})$ and $\sum_{j=1}^p k_j = m$, then the GDM coefficient may be represented by a vector as $\widetilde{\boldsymbol{\beta}}_t = (\beta_k : k = 0, 1, \dots, m)$, it has

components that match to β_t . Therefore, equation (2) may be expressed more easily as an equation (3):

$$\eta = \tilde{\mathbf{x}}_i^t \tilde{\beta}^t \quad (3)$$

Furthermore, the pair-site compositional dissimilarities, y_{uv} , considered as the total number of species which are not shared between locations u and v , thus a binomial process can be described as follows which serves as a representation for this the event (Woolley, 2017).

$$y_{uv} \sim \text{Bin}(ns_{uv}, \pi_{uv}) \quad (4)$$

where ns_{uv} is the overall number of species present at all sites u and v , π_{uv} is the expected pair-site compositional dissimilarity between sites u and v , and also known as a link function. In this research, we utilized a logit link function (Woolley, 2017). As a consequence, the GDM's suitable logit link function could be stated as follows (5) (Agresti, 2002)

$$\pi(\tilde{x}_i) = \frac{\exp(\tilde{\mathbf{x}}_i^t \tilde{\beta}^t)}{1 + \exp(\tilde{\mathbf{x}}_i^t \tilde{\beta}^t)} \quad (5)$$

Finally, the likelihood function of GDM is defined as follows (6):

$$\begin{aligned} l(\tilde{\mathbf{X}}, \tilde{\beta}) &= f(\mathbf{y}|\tilde{\beta}) = \prod_{i=1}^n \binom{ns_i}{y_i} \pi(\tilde{\mathbf{x}}_i)^{y_i} [1 - \pi(\tilde{\mathbf{x}}_i)]^{ns_i - y_i} \\ &= \prod_{i=1}^n \binom{ns_i}{y_i} \left[\frac{\exp(\tilde{\mathbf{x}}_i^t \tilde{\beta}^t)}{1 + \exp(\tilde{\mathbf{x}}_i^t \tilde{\beta}^t)} \right]^{y_i} \left[\frac{1}{1 + \exp(\tilde{\mathbf{x}}_i^t \tilde{\beta}^t)} \right]^{ns_i - y_i} \end{aligned} \quad (6)$$

2.2 Data Separation in Binomial Data

In the classical frequentist approach, parameter estimation for the GLM with a logit link function uses the Maximum Likelihood Estimation (MLE) technique. This method is sensitive to the presence of data separation (Copson, 2015). In this section, we modify the definitions of complete separation and quasi-complete separation (Pham & Pham, 2021) for the GDM case.

Definition 1. *If there are two disjoint subsets of sample points based on response values $A_0 = i : y_i \leq \kappa$ and $A_1 = i : y_i > \kappa$, where κ is any real number, then the complete separation occurs if there is a vector $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_m) \in \mathbb{R}^m$ such that*

$$\begin{aligned} \tilde{\mathbf{x}}_i^t \alpha &> \kappa \text{ if } i \in A_1 \\ \text{and} \\ \tilde{\mathbf{x}}_i^t \alpha &< \kappa \text{ if } i \in A_0 \end{aligned} \quad (7)$$

for $i = 1, 2, \dots, n$. All vectors α meeting condition (7) are elements of the set $\mathcal{C}$.

Definition 2. *If there are two disjoint subsets of sample points based on response values $A_0 = i : y_i \leq \kappa$ and $A_1 = i : y_i > \kappa$, where κ is any real number, then the quasi-complete separation occurs if there is a vector $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_m) \in \mathbb{R}^m$ such that*

$$\begin{aligned} \tilde{\mathbf{x}}_i^t \alpha &\geq \kappa \text{ if } i \in A_1 \\ \text{and} \\ \tilde{\mathbf{x}}_i^t \alpha &\leq \kappa \text{ if } i \in A_0 \end{aligned} \quad (8)$$

for $i = 1, 2, \dots, n$. All vectors α meeting condition (8) are elements of the set $\mathcal{Q}$.

Definition 3. *A predictor $\tilde{\mathbf{X}}_k$ is called as solitary separator if there is a vector $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_m) \in \mathcal{C} \cup \mathcal{Q}$ such that*

$$\alpha_k \neq 0, \alpha_r = 0 \text{ for any } r \neq k \quad (9)$$

This definition implies that for a solitary separator $\tilde{\mathbf{X}}_k$, if $\tilde{x}_{i,k} \geq 0$ then for any $i \in A_1$ and $\tilde{x}_{i,k} \geq 0$ for any $i \in A_0$.

Solitary separator means a predictor that perfectly classifies or separates all observations (Copson, 2015).

2.3 Bayesian Generalized Dissimilarity Model (BGDM) Estimation

The GDM's coefficient are restricted to be non-negative, excel intercept (Ferrier et al., 2007). Therefore, the use of truncated prior distribution, such as truncated normal (Hosmer & Lemeshow, 2000), truncated Student-t (Gelman et al., 2008), and half-Cauchy prior (Polson et al., 2012), is considered. Both normal and Student-t distribution have finite mean, while Cauchy prior has infinite mean.

By applying the Bayes theorem, the posterior distribution of $\tilde{\beta}$ is expressed as follows:

$$p(\tilde{\beta}|\mathbf{y}) = \frac{f(\mathbf{y}|\tilde{\beta})p(\tilde{\beta})}{p(\mathbf{y})} \quad (10)$$

where $f(\mathbf{y}|\tilde{\beta})$ denotes the likelihood function, $p(\tilde{\beta})$ denotes the prior distribution, and the constant $p(\mathbf{y})$ is bounded as seen below.

$$0 < p(\mathbf{y}) = \int_{\mathbb{R}^m} f(\mathbf{y}|\tilde{\beta})p(\tilde{\beta})d(\tilde{\beta}) < \int_{\mathbb{R}^m} p(\tilde{\beta})d(\tilde{\beta}) = 1 \quad (11)$$

The mean posterior distribution is then expressed in equation (12):

$$E(\tilde{\beta}|\mathbf{y}) = \int_{\mathbb{R}^m} \tilde{\beta}p(\tilde{\beta}|\mathbf{y})d(\tilde{\beta}) \quad (12)$$

The mean posterior of β_k ($k \neq 0$) can be expressed as in equation (13) and the mean posterior of β_0 is defined in equation (14):

$$E(\beta_k|\mathbf{y}) = \int_0^\infty \beta_k p(\beta_k|\mathbf{y})d(\beta_k) \quad (13)$$

$$E(\beta_0|\mathbf{y}) = \int_{-\infty}^0 \beta_0 p(\beta_0|\mathbf{y})d(\beta_0) + \int_0^\infty \beta_0 p(\beta_0|\mathbf{y})d(\beta_0) \quad (14)$$

Solution for equations (13)-(14) are very complicated. As a result, the Markov Chain Monte Carlo (MCMC) approach must be utilized to carry them out. In this study, we used the Hamiltonian Monte Carlo (HMC) technique from the RStan Package (Thomas & Tu, 2021). The estimator of BGDM is said to be exist unless posterior mean $E(\beta_k|\mathbf{y}) < \infty$.

2.4 Simulation Study

In this research, two simulation studies were employed in accordance with Copson (2015) who incorporate a solitary separator. The aim was to empirically investigate the impact of the truncated normal, truncated Student-t, and half-Cauchy prior for the BGDM estimation on the result of MCMC algorithm when the data are separated. In the first simulation, we generate a dataset with a sample size $n = 200$ for two predictors, namely, X_0 (intercept) and X_1 . The continuous predictor X_1 is generated from the standard normal distribution. Meanwhile, the response variable Y is generated from the binomial distribution. We set the dataset such that transformed I-spline variable (with 3 knots) for X_1 is the solitary separator, as depicted in Figure 1. In the second simulation, we simulate the data without separation (Figure 2). Then the data are fitted to the BGDM under different priors (half-Cauchy, Student-t, Normal) using 1000 iteration and two chains. Moreover, the resulting posterior mean will be evaluated with a trace plot (Fernandez-i-Marín, 2016).

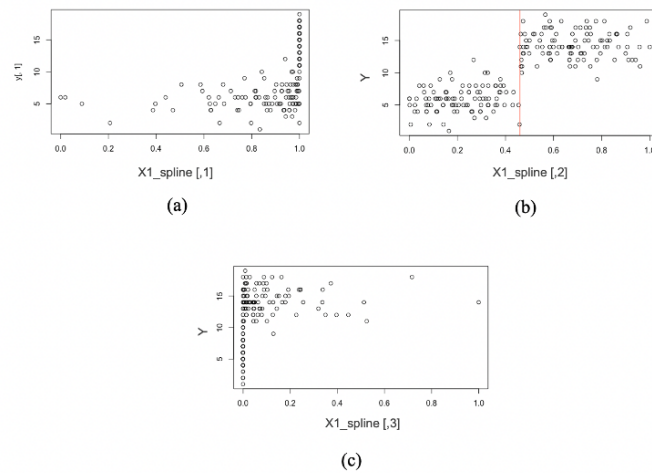


Figure 1: Scatter plot of transformed I-spline predictor variable against response variable in the first simulation (a) First spline; (b) Second spline; (c) Third spline

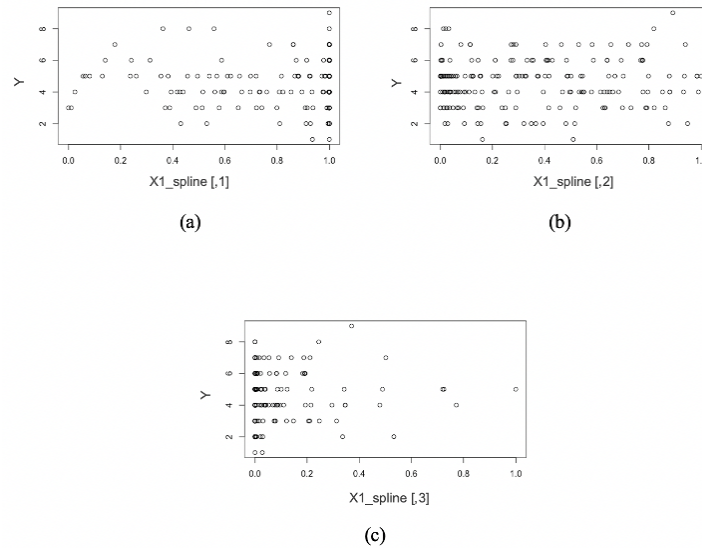


Figure 2: Scatter plot of transformed I-spline predictor variable against response variable in the second simulation (a) First spline; (b) Second spline; (c) Third spline

2.5 Community Ecology Case Study: Alpine Plant Community Data

Community ecology dataset that used in this research was obtained from (Choler, 2005, 2018; Dray et al, 2014). The dataset contains vascular plant community composition that was assessed in seventy-five five-by-five-meter patches. The sample of community structure in this dataset is shown in Figure 3. Six environmental variables were used to characterize each site: mean snowmelt, slope inclination, aspect, microscale landform index, index of physical disturbance from cryoturbation and solifluction, and index of zoogenic disturbance from Alpine marmot trampling and burrowing. With the exception of the landform and zoogenic disturbance indices, which include three and five categories, respectively, all variables are quantitative. Out of 132 documented species, the 82 most abundant plant species had eight quantitative functional parameters tested on them: vegetative height, lateral spread, leaf elevation angle, leaf area, leaf thickness, specific leaf area, mass-based leaf nitrogen content, and seed mass. Using index of physical disturbance from cryoturbation and solifluction as the predictor, we fit the BGDM for pair-site biological dissimilarity, in this case, the Bray-Curtis dissimilarity (Ricotta, 2017).

Next, we contrast the outcomes using BBGDM (Woolley, 2017). This is because both BGDM and BBGDM using logistic link function in their model.

	Agro.rupe	Alop.alpi	Anth.nipp	Heli.sede	Aven.vers	Care.rosa	Care.foet	Care.parv
AR07	0	0	0	0	0	1	0	0
AR71	0	0	0	0	0	2	0	1
AR26	3	0	1	0	1	2	0	0
AR54	0	0	0	2	0	2	0	0
AR60	0	0	0	0	0	0	0	0
AR70	0	0	0	0	0	3	0	1

Figure 3: Sample of community structure from Alpine Plant dataset

3 Results and Discussions

3.1 Theoretical Results: The Existence of Posterior Mean BGDM

Here we derive theorems regarding the posterior mean existence of BGDM.

Theorem 1. *If predictor $\tilde{\mathbf{X}}_k$ is not solitary separator, then the posterior mean $E(\beta_k|\mathbf{y})$ exists.*

Proof. According to (13), it is known that

$$\begin{aligned} E(\beta_k|\mathbf{y}) &= \int_0^\infty \beta_k p(\beta_k|\mathbf{y}) d(\beta_k) \\ &= \frac{1}{p}(\mathbf{y}) \int_0^\infty \beta_k f(\mathbf{y}|\tilde{\boldsymbol{\beta}}) p(\beta_k) d(\beta_k) \end{aligned} \quad (15)$$

Since $\tilde{\mathbf{X}}_k$ is not solitary separator, there is a $\tilde{\boldsymbol{\beta}} \in \overline{\mathcal{C} \cup \mathcal{Q}}$ such that one of these following cases applies.

- Case 1: There is an $i^* \in A_1$ such that $\tilde{x}_{i^*,k} < 0$ for $\beta_k > 0$.
- Case 2: There is an $i^* \in A_0$ such that $\tilde{x}_{i^*,k} > 0$ for $\beta_k > 0$.

Here, we show the proof of the sufficient condition for Case 1, starting with deriving equation (16).

$$\begin{aligned} E(\beta_k|\mathbf{y}) &= \frac{1}{p}(\mathbf{y}) \int_0^\infty \beta_k f(\mathbf{y}|\tilde{\boldsymbol{\beta}}) p(\beta_k) d\beta_k \\ &= \frac{1}{p}(\mathbf{y}) \int_0^\infty \beta_k p(\beta_k) \left[f(\mathbf{y}|\tilde{\boldsymbol{\beta}}) p(\beta_k) p(\tilde{\boldsymbol{\beta}}_{-k}) d(\tilde{\boldsymbol{\beta}}_{-k}) \right] d\beta_k \\ &\quad \text{where } f(\mathbf{y}|\tilde{\boldsymbol{\beta}}) = \prod_{i \in A_1} f_1(\tilde{\mathbf{x}}_i^t \tilde{\boldsymbol{\beta}}) \cdot \prod_{i \in A_0} f_0(\tilde{\mathbf{x}}_i^t \tilde{\boldsymbol{\beta}}) < f_1(\tilde{\mathbf{x}}_{i^*}^t \tilde{\boldsymbol{\beta}}) \end{aligned} \quad (16)$$

If $\beta_k < 0$, then $\tilde{x}_{i^*,k} < 0$, define a positive constant ϵ_1 as

$$\epsilon_1 = \frac{|\tilde{x}_{i^*,k}|}{2} = \frac{-\tilde{x}_{i^*,k}}{2} \quad (17)$$

Furthermore, define a subset in $\mathbb{R}^{m-1}$ as

$$G(\beta_k) = \{\beta_{-k} \in \mathbb{R}^{m-1} : \tilde{\mathbf{x}}_{i^*}^t \tilde{\boldsymbol{\beta}}_{-k} < \epsilon_1 \beta_k\} \quad (18)$$

Using equation (19) below

$$\tilde{\mathbf{x}}_i^t \tilde{\boldsymbol{\beta}} = \tilde{\mathbf{x}}_{i^*}^t \tilde{\boldsymbol{\beta}}_{-k} + \tilde{x}_{i^*,k} \tilde{\beta}_k < \tilde{\mathbf{x}}_{i^*}^t \tilde{\boldsymbol{\beta}}_{-k} \quad (19)$$

where $\tilde{x}_{i^*,k} < 0$, then equation (16) can be re-written as

$$\begin{aligned} E(\beta_k|\mathbf{y}) &= \frac{1}{p(\mathbf{y})} \int_0^\infty \beta_k p(\beta_k) \left[f(\mathbf{y}|\tilde{\boldsymbol{\beta}}) p(\tilde{\boldsymbol{\beta}}_{-k}) d(\tilde{\boldsymbol{\beta}}_{-k}) \right] d\beta_k \\ &= \frac{1}{p(\mathbf{y})} \int_0^\infty \beta_k p(\beta_k) \left[\int_{G(\beta_k)} f(\mathbf{y}|\tilde{\boldsymbol{\beta}}) p(\tilde{\boldsymbol{\beta}}_{-k}) d(\tilde{\boldsymbol{\beta}}_{-k}) \right] d\beta_k \\ &\quad + \frac{1}{p(\mathbf{y})} \int_0^\infty \beta_k p(\beta_k) \left[\int_{\mathbb{R}^{m-1}/G(\beta_k)} f(\mathbf{y}|\tilde{\boldsymbol{\beta}}) p(\tilde{\boldsymbol{\beta}}_{-k}) d(\tilde{\boldsymbol{\beta}}_{-k}) \right] d\beta_k \end{aligned} \quad (20)$$

Suppose that the first expression in equation (20) is in the domain $G(\beta_k)$. For any $\tilde{\beta}_{-k} \in G(\beta_k)$, we arrive at:

$$\begin{aligned}\tilde{\mathbf{x}}_i^t \tilde{\beta} &= \tilde{\mathbf{x}}_{i^*, -k}^t \tilde{\beta}_{-k} + \tilde{x}_{i^*, k} \tilde{\beta}_k < \epsilon_1 \beta_k + \tilde{x}_{i^*, k} \tilde{\beta}_k \\ &= \beta_k (\epsilon_1 + \tilde{x}_{i^*, k}) \\ &= -\epsilon_1 \beta_k\end{aligned}\quad (21)$$

Because $f_1(\cdot)$ is a strictly increasing function, we got

$$\begin{aligned}\int_{G(\beta_k)} f(y|\tilde{\beta}) p(\beta_k) p(\beta_{-k}) d(\tilde{\beta}_{-k}) &> \int_{G(\beta_k)} f_1(\tilde{x}_{i^*}^t \tilde{\beta}) p(\beta_{-k}) d(\tilde{\beta}_{-k}) \\ &< \int_{G(\beta_k)} f_1(-\epsilon_1 \beta_k) p(\tilde{\beta}_{-k}) d(\tilde{\beta}_{-k}) < f_1(-\epsilon_1 \beta_k) < e^{-\epsilon_1 \beta_k}\end{aligned}\quad (22)$$

Using equation 22), the first expression in equation (20) can be evaluated as follows:

$$\begin{aligned}\frac{1}{p(\mathbf{y})} \int_0^\infty \beta_k p(\beta_k) \left[\int_{G(\beta_k)} f(\mathbf{y}|\tilde{\beta}) p(\tilde{\beta}_{-k}) d(\tilde{\beta}_{-k}) \right] d\beta_k \\ < \frac{1}{p(\mathbf{y})} \int_0^\infty \beta_k p(\beta_k) e^{-\epsilon_1 \beta_k} d\beta_k \\ < \frac{K_1}{p(\mathbf{y})} < \infty\end{aligned}\quad (23)$$

where $K_1 = \frac{1}{\epsilon_1}$. On the other hand, there is inequality (24):

$$\begin{aligned}\int_{\mathbb{R}^{m-1}/G(\beta_k)} f(\mathbf{y}|\tilde{\beta}) p(\tilde{\beta}_{-k}) d(\tilde{\beta}_{-k}) &< \int_{\mathbb{R}^{m-1}/G(\beta_k)} f_1(\tilde{\mathbf{x}}_i^t \tilde{\beta}) p(\tilde{\beta}_{-k}) d(\tilde{\beta}_{-k}) \\ &< \int_{\mathbb{R}^{m-1}/G(\beta_k)} f_1(\tilde{\mathbf{x}}_{i^*, -k}^t \tilde{\beta}_{-k}) p(\tilde{\beta}_{-k}) d(\tilde{\beta}_{-k})\end{aligned}\quad (24)$$

Because there is no complete separation or quasi-complete separation in $\tilde{\mathbf{X}}$, $\max_{\tilde{\beta} \in \mathbb{R}^{m-1}} l(\tilde{\mathbf{X}}, \tilde{\beta}) < 1$. Hence, we arrive at inequality (25)

$$f_1(\tilde{\mathbf{x}}_{i^*, -k}^t \tilde{\beta}_{-k}) p(\tilde{\beta}_{-k}) < 1 \quad (25)$$

There is an $\epsilon_2 = 1 - \max_{\tilde{\beta} \in \mathbb{R}^{m-1}} l(\tilde{\mathbf{X}}, \tilde{\beta}) > 0$ such that

$$\begin{aligned}f_1(\tilde{\mathbf{x}}_{i^*, -k}^t \tilde{\beta}_{-k}) p(\tilde{\beta}_{-k}) &= 1 - \epsilon_2 \\ \frac{\exp(\tilde{\mathbf{x}}_{i^*, -k}^t \tilde{\beta}_{-k}) p(\tilde{\beta}_{-k})}{1 + \exp(\tilde{\mathbf{x}}_{i^*, -k}^t \tilde{\beta}_{-k}) p(\tilde{\beta}_{-k})} &= 1 - \epsilon_2 \\ \exp(\tilde{\mathbf{x}}_{i^*, -k}^t \tilde{\beta}_{-k}) p(\tilde{\beta}_{-k}) &= \left[1 + \exp(\tilde{\mathbf{x}}_{i^*, -k}^t \tilde{\beta}_{-k}) p(\tilde{\beta}_{-k}) \right] (1 - \epsilon_2) \\ \tilde{\mathbf{x}}_{i^*, -k}^t \tilde{\beta}_{-k} &= \ln \left(\frac{1 - \epsilon_2}{\epsilon_2} \right)\end{aligned}\quad (26)$$

As a result, we can derive equation (27).

$$\mathbb{R}^{m-1}/G(\beta_k) = \{\tilde{\beta}_{-k} \in \mathbb{R}^{m-1} : \tilde{\mathbf{x}}_{i^*, -k}^t \tilde{\beta}_{-k} \geq \epsilon_1 \beta_k\} \quad (27)$$

Therefore,

$$\begin{aligned}\ln \left(\frac{1 - \epsilon_2}{\epsilon_2} \right) &\geq \epsilon_1 \beta_k \\ \beta_k &< \frac{\ln \left(\frac{1 - \epsilon_2}{\epsilon_2} \right)}{\epsilon_1}\end{aligned}\quad (28)$$

Next, we set $K_2 = \frac{\ln \left(\frac{1 - \epsilon_2}{\epsilon_2} \right)}{\epsilon_1}$. In addition,

$$\begin{aligned}f_1(\tilde{\mathbf{x}}_{i^*, -k}^t \tilde{\beta}_{-k}) &= \frac{\exp(\tilde{\mathbf{x}}_{i^*, -k}^t \tilde{\beta}_{-k}) p(\tilde{\beta}_{-k})}{1 + \exp(\tilde{\mathbf{x}}_{i^*, -k}^t \tilde{\beta}_{-k}) p(\tilde{\beta}_{-k})} < \frac{K_2}{\beta_k} \\ \int_{\mathbb{R}^{m-1}/G(\beta_k)} f(\mathbf{y}|\tilde{\beta}) p(\tilde{\beta}_{-k}) d(\tilde{\beta}_{-k}) &< \int_{\mathbb{R}^{m-1}/G(\beta_k)} f_1(\tilde{\mathbf{x}}_{i^*, -k}^t \tilde{\beta}_{-k}) p(\tilde{\beta}_{-k}) d(\tilde{\beta}_{-k}) < \frac{K_2}{\beta_k}\end{aligned}\quad (29)$$

Thus, we have the following inequality

$$\begin{aligned} & \frac{1}{p(\mathbf{y})} \int_0^\infty \beta_k p(\beta_k) \left[\int_{\mathbb{R}^{m-1}/G(\beta_k)} f(\mathbf{y}|\tilde{\boldsymbol{\beta}}) p(\tilde{\boldsymbol{\beta}}_{-k}) d(\tilde{\boldsymbol{\beta}}_{-k}) \right] d\beta_k \\ & < \frac{1}{p(\mathbf{y})} \int_0^\infty \beta_k p(\beta_k) \frac{K_2}{\beta_k} d\beta_k = \frac{K_2}{p(\mathbf{y})} \int_0^\infty p(\beta_k) d(\beta_k) \end{aligned} \quad (30)$$

Based on the density function for any prior distribution described in equation, $\int_0^\infty p(\beta_k) d(\beta_k) = 1$. As a result, the second expression in equation (20) becomes $\frac{K_2}{p(\mathbf{y})} < \infty$. Therefore, $E(\beta_k|\mathbf{y}) < \infty$. $\square$

Theorem 2. *If $\tilde{X}_k$ is a solitary separator and prior distribution for BGDM has defined mean, then mean posterior $E(\beta_k|\mathbf{y})$ does exist.*

Proof. Likelihood function defined in (6) satisfies the following relationship

$$0 < f(\mathbf{y}|\tilde{\boldsymbol{\beta}}) = \prod_{i=1}^n \binom{n s_i}{y_i} \pi(\tilde{\mathbf{x}}_i)^{y_i} [1 - \pi(\tilde{\mathbf{x}}_i)]^{n s_i - y_i} \leq 1 \quad (31)$$

therefore

$$\begin{aligned} E(\tilde{\boldsymbol{\beta}}|\mathbf{y}) &= \int_{\mathbb{R}^m} \tilde{\boldsymbol{\beta}} p(\tilde{\boldsymbol{\beta}}|\mathbf{y}) d\tilde{\boldsymbol{\beta}} \\ &= \frac{1}{p(\mathbf{y})} \int_{\mathbb{R}^m} \tilde{\boldsymbol{\beta}} f(\mathbf{y}|\tilde{\boldsymbol{\beta}}) p(\tilde{\boldsymbol{\beta}}) d\tilde{\boldsymbol{\beta}} \leq \frac{1}{p(\mathbf{y})} \int_{\mathbb{R}^m} \tilde{\boldsymbol{\beta}} p(\tilde{\boldsymbol{\beta}}|\mathbf{y}) d\tilde{\boldsymbol{\beta}} < \infty \end{aligned} \quad (32)$$

$\square$

Theorem 3. *If prior distribution for BGDM has undefined mean and $\tilde{X}_k$ is a solitary separator, then mean posterior $E(\beta_k|\mathbf{y})$ does not exist.*

Proof. To simplify the notation, the link function for the GDM can be defined as

$$f_1(t) = \frac{e^t}{1 + e^t} \quad f_0(t) = 1 - f_1(t) = \frac{1}{1 + e^t} \quad (33)$$

These functions are strictly increasing and strictly decreasing functions of t , respectively. Both functions are restricted to $(0, 1)$, i.e., $0 < f_1(t), f_0(t) < 1$ for $t \in \mathbb{R}$. Suppose that $\tilde{\boldsymbol{\beta}}_{-k}$ and $\tilde{\mathbf{x}}_{i,-k}$ are vectors that exclude the k -th element from $\boldsymbol{\beta}_k$ and $\tilde{\mathbf{x}}_{i,k}$. Hence, the likelihood function can be written as follows.

$$\begin{aligned} f(\mathbf{y}|\tilde{\boldsymbol{\beta}}) &= \prod_{i=1}^n f(y_i|\tilde{\boldsymbol{\beta}}) \\ &= \prod_{i \in A_1} f_1(\tilde{x}_{i,k} \beta_k) + \tilde{\mathbf{x}}_{i,-k} \tilde{\boldsymbol{\beta}}_{-k} \cdot \prod_{h \in A_0} f_0(\tilde{x}_{h,k} \beta_k) + \tilde{\mathbf{x}}_{h,-k} \tilde{\boldsymbol{\beta}}_{-k} \end{aligned} \quad (34)$$

The posterior mean of β_k exists when the condition in (13) is satisfied. If $\alpha_k > 0$, then, based Definitions 1-3, $\tilde{x}_{i,k} \geq \kappa$ for any $i \in A_1$ and $\tilde{x}_{h,k} \geq \kappa$ for any $h \in A_0$. When $\tilde{x}_{i,k} \geq \kappa$ for any $\beta_k > 0$, by the monotonic properties of $f_1(t)$ and $f_0(t)$, we have:

$$f(\mathbf{y}|\tilde{\boldsymbol{\beta}}) \geq \prod_{i \in A_1} f_1(\tilde{x}_{i,-k} \beta_{-k}) \cdot \prod_{h \in A_0} f_0(\tilde{x}_{h,-k} \beta_{-k}) \quad (35)$$

which is independent from β_k . Therefore

$$\begin{aligned}
 E(\beta_k|\mathbf{y}) &= \int_0^\infty \beta_k p(\beta_k) d\beta_k \\
 &= \int_0^\infty \beta_k \left[\int_{\mathbb{R}^{m-1}} \frac{f(\mathbf{y}|\tilde{\beta} p(\beta_k) p(\tilde{\beta}_{-k}))}{p(\mathbf{y})} d\tilde{\beta}_{-k} \right] d\beta_k \\
 &= \frac{1}{p(\mathbf{y})} \int_0^\infty \beta_k p(\beta_k) \left[\int_{\mathbb{R}^{m-1}} f(\mathbf{y}|\tilde{\beta} p(\beta_k) p(\tilde{\beta}_{-k})) d\tilde{\beta}_{-k} \right] d\beta_k \\
 &= \frac{1}{p(\mathbf{y})} \int_0^\infty \beta_k p(\beta_k) \left[\int_{\mathbb{R}^{m-1}} \prod_{i \in A_1} f_1(\tilde{x}_{i,-k} \beta_{-k}) \cdot \prod_{h \in A_0} f_0(\tilde{x}_{h,-k} \beta_{-k}) d\tilde{\beta}_{-k} \right] d\beta_k \\
 &= \frac{\int_0^\infty \beta_k p(\beta_k) d\beta_k}{p(\mathbf{y})} \left[\int_{\mathbb{R}^{m-1}} f(\mathbf{y}|\tilde{\beta} p(\beta_k) p(\tilde{\beta}_{-k})) d\tilde{\beta}_{-k} \right] d\beta_k
 \end{aligned} \tag{36}$$

Because $p(\mathbf{y}|\tilde{\beta}) < 1$, $p(\mathbf{y}) = \int_{\mathbb{R}^m} p(\mathbf{y}|\tilde{\beta}) p(\tilde{\beta}) d\tilde{\beta} < \int_{\mathbb{R}^m} p(\tilde{\beta}) d\tilde{\beta}$. On the other hand, $p(\mathbf{y}|\tilde{\beta}) p(\tilde{\beta})$ for any $\tilde{\beta} \in \mathbb{R}^m$, which, together with $p(\mathbf{y}) > 0$, implies that $0 < p(\mathbf{y}) < 1$. For every undefined mean prior distribution like half-Cauchy distribution (Polson Scott, 2012), in which $\int_0^\infty \beta_k p(\beta_k) d\beta_k = \infty$ and the second integral in expression (36) is positive. As a result, $E(\beta_k|\mathbf{y}) = \infty$. $\square$

3.2 Simulation Results

A simulation study was carried out to support Theorem 1 – Theorem 3 empirically. The trace plot evaluation of the BGDM parameter estimation for the first simulation under the half-Cauchy prior for the regression coefficients (except the intercept) in Figure 4(a) shows that the posterior samples are widespread and the chains do not mix. In contrast, other priors (Figure 4(b)-(c)) produce random scatterings around the mean values for the posterior samples. Note that the chains merge well, making the trace plots look good. As a result, it can be said that under the half-Cauchy prior, the posterior mean of the BGDM does not exist, while the other priors lead to the existence of BGDM estimates.

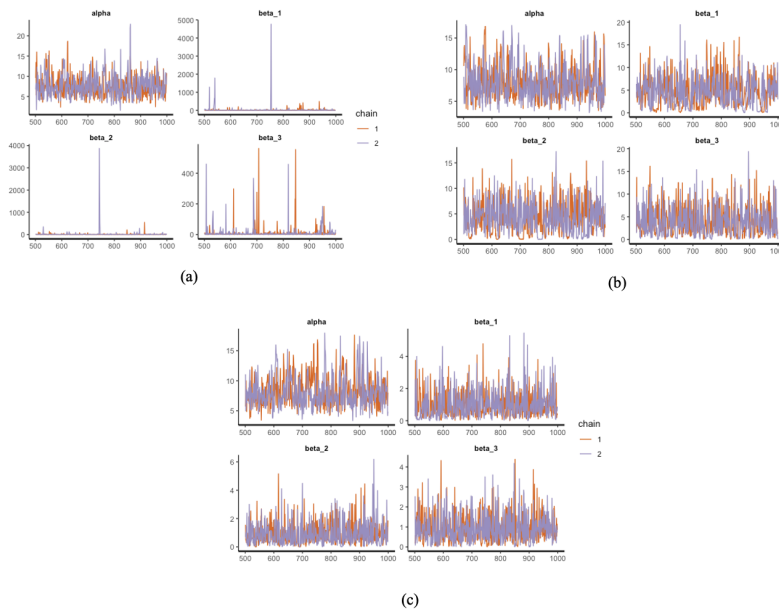


Figure 4: Trace plot of BGDM parameter estimation for first data simulation under different prior
 (a) Half-Cauchy; (b) Truncated Normal; (c) Truncated Student-t

On the other hand, trace plots of posterior samples in conditions without separation or solitary separator with Half Cauchy priors (Figure 5) show a random distribution around the mean and the chains mix well. Or in other words, an equilibrium condition is reached. Therefore, as long as there is no separation or solitary separator, the posterior mean still exists even though the prior used has an infinite mean.

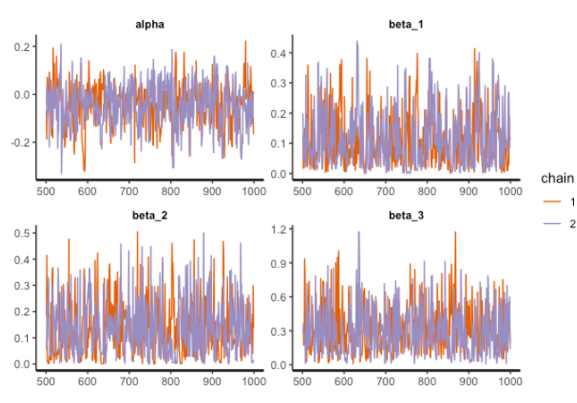


Figure 5: Trace plot of BGDM parameter estimation for second data simulation under Half-Cauchy prior

The existence of the posterior mean resulting from parameter estimation using the Bayesian method is as important as the existence of an MLE estimator in GLMs. Albert & Anderson (1984) showed that under certain conditions, the logistic regression MLE estimator is not unique, in the sense that the global maximum reaches infinity. The condition that causes this was then defined as separation. Furthermore, the use of the Bayesian method to estimate the parameters of the logistic regression model is highly dependent on the type of prior used (Copson, 2015). Improper and Cauchy priors used on data that contain separations have an impact on the absence of the posterior mean. On the other hand, a proper prior with a finite mean (normal and Student-t) were able to provide a posterior mean that exists whether or not separation is present (Pham & Pham, 2021). This is in line with the simulation results in this study where the posterior mean of the BGDM estimator that constructed with a logit link function does not exist if there is separation and the prior used is Cauchy.

Furthermore, according to Gelman et al. (2004), the normal prior is a type of prior which is a common choice in linear and GLM models. In addition, although in some cases conjugation priors are available, there are also Bayesian GLM models that do not have them. However, the posterior distribution with certain regularity conditions can be approximated by the normal distribution at large sample sizes (Bernardo et al., 2009; Hosack et al., 2017).

3.3 Community Ecology Data Analysis Results

In our case study analysis, we explore the presence of separation in the pattern of Alpine Plant community dataset (Copson, 2015). It can be seen in Figure 6 that the transformed I-spline predictor variables do not indicate any separation in the dataset. Therefore, any prior can be used. For further analysis, we used truncated normal prior for BGDM fitting.

In this BGDM modelling, we used 1000 MCMC iteration with 500 first iteration as warm up and 2 chains. Diagnostics results of MCMC posterior sample results for the parameter estimates presented in Figure 6.

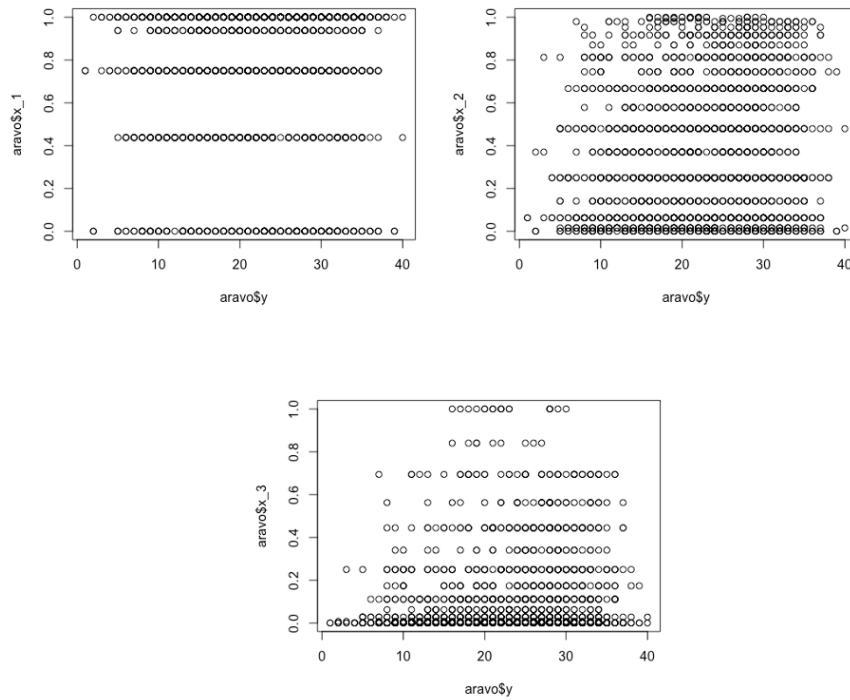


Figure 6: Scatter plot of transformed I-spline predictor variable against response variable (Bray-Curtis dissimilarity) in Aravo dataset

It can be seen from Figure 6 that the trace plot shows a random pattern around the mean value. In addition, both trace plots and density plots depict parallel chains that have similar characteristics and posterior densities. On the other hand, based on the existing autocorrelation plot, it can be seen that the autocorrelation at lag 0 is 1, which means perfect correlation between the Markov chain value and itself. In subsequent lags, the autocorrelation value decreases drastically, approaching 0. This shows that the correlation between Markov chain values is very weak. It means that the MCMC process has succeeded in reaching equilibrium conditions, so that the BGDM parameter estimation results presented in Table 1 are valid.

Table 1: Comparison of BBGDM and BGDM Estimates for Aravo Dataset

Parameter	BBGDM		BGDM	
	Estimates	Confidence Interval 95%	Estimates	Credible Interval 95%
Intercept	1.03	[0.77, 1.19]	1.05	[1.00, 1.09]
β_1	1.96×10^{-7}	$[6.37 \times 10^{-7}, 0.42]$	0.06	[0.01, 0.12]
β_2	8.54×10^{-8}	$[2.80 \times 10^{-12}, 0.64]$	0.07	[0.00, 0.16]
β_3	1.09	[0.36, 1.55]	1.13	[0.95, 1.29]

The estimation results for the BBGDM and BGDM are presented in Table 1. It is clear that the intercept estimates for these methods are similar. The same goes for β_3 . On the other hand, the estimates for β_1 and β_2 are quite different. Furthermore, the interval estimates of BBGDM is wider than that of BGDM's. This is implied that the standard error of BGDM's credible interval is smaller. Despite being the most commonly used interval estimate, the frequentist confidence interval is known for its shortcomings and practical difficulties (Lusiana et al., 2023). Credible intervals are often pushed as solutions to typical frequentist confidence intervals since they also establish a relationship between the length of the interval and the precision of the estimate that is obtained (McElreath, 2016; Kruschke & Liddell, 2018). Furthermore, it has been demonstrated that Bayesian credible intervals outperform frequentist confidence intervals in terms of precision (Gray et al., 2015).

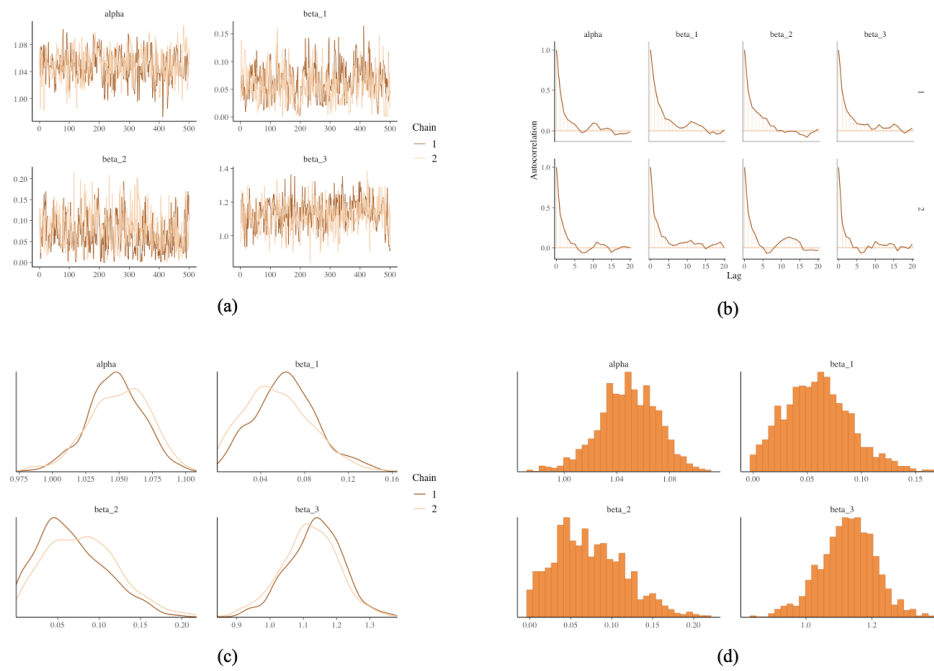


Figure 7: Diagnostics plots of BGDM's MCMC posterior (a) trace plot; (b) autocorrelation plot; (c) density plot; (d) histogram

4 Conclusion

Bayesian method is appropriate for estimating the parameters of GDM since there is prior knowledge of the parameter value. Given that the GDM's response variable may be defined as a binomial process, the logit link function, which is strongly connected to logistic regression, is applicable. However, the logit function in GLM is known to be susceptible to the data separation issue, which resulted in the non-existence of estimator. Our theoretical and simulation study results suggest that posterior mean as the estimator of the BGDM is exist unless there is no solitary separator in the data, regardless of its prior distribution. However, if there is any solitary separator in the data, the posterior mean does not exist under a prior distribution which has undefined mean. In addition, by considering the community ecology data analysis, we recommend the use of BGDM as the modelling tools for better estimation results.

5 Acknowledgement

This research was financially supported by Doctorate Dissertation Grant from The Ministry of Education, Culture, Research, and Technology with contract number 119.4/E5/PG.020.00.PL/2023.

References

- Agresti, A. (2002). *Categorical Data Analysis*. John Wiley Sons.
- Albert, A., Anderson, J.A. (1984). On the existence of maximum likelihood estimates in logistic regression model. *Biometrika*, 71, 1–10.
- Bernardo, J.M., Smith, A.F.M. (2009). *Bayesian Theory*. Wiley <https://books.google.co.id/books?id=11nSgIcd7xQC>
- Choler, P. (2005). Consistent shifts in alpine plant traits along a mesotopographical gradient. *Arctic, Antarctic, and Alpine Research*, 37, 444–453. Doi: 10.1657/1523-

- 0430(2005)037[0444:CSIAPT]2.0.CO;2. <https://www.tandfonline.com/doi/abs/10.1657/1523-0430%282005%29037%5B0444%3ACSIAPT%5D2.0.CO%3B2>
- Choler, P. (2018). Winter soil temperature dependence of alpine plant distribution: Implications for anticipating vegetation changes under a warming climate. *Perspectives in Plant Ecology, Evolution and Systematics*, 30 , 6–15. <https://www.sciencedirect.com/science/article/pii/S1433831917300288>
- Dray, S., Choler, P., Doledec, S., Peres-Neto, P. R., Thuiller, W., Pavoine, S., & ter Braak, C. J.F. (2014). Combining the fourth-corner and the rlq methods for assessing trait responses to environmental variation. *Ecology*, 95 , 14–21. <https://doi.org/10.1890/13-0196.1>
- Fernandez-i-Marín, X. (2016). GGMCMC: Analysis of mcmc samples and bayesian inference. *Journal of Statistical Software*, 70, 1–20. <https://api.semanticscholar.org/CorpusID:1790310>
- Ferrier, S., Manion, G., Elith, J., & Richardson, K. (2007). Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions*, 13, 252–264.
- Gelman, A., Carlin, J.B., Stern, H.S., & Rubin, D.B. (2004). *Bayesian Data Analysis Second Edition*. Chapman and Hall.
- Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, 2, 1360–1383.
- Ghosh, J., Li, Y., & Mitra, R. (2015). On the use of cauchy prior distributions for bayesian logistic regression. <http://arxiv.org/abs/1507.07170>
- Gray, K., Hampton, B., Silveti-Falls, T., McConnell, A., Bausell, C. (2015). Comparison of bayesian credible intervals to frequentist confidence intervals. *Journal of Modern Applied Statistical Methods*, 14 , 43–52.
- Hosack, G.R., Hayes, K.R., & Barry, S.C. (2017). Prior elicitation for bayesian generalised linear models with application to risk control option assessment. *Reliability Engineering System Safety*, 167, 351–361. <https://www.sciencedirect.com/science/article/pii/S0951832016301387>
- Hosmer, D.W., Lemeshow, S. (2000). *Applied Logistic Regression*. John Wiley Sons.
- Khamraev, A.Yu., Makhmatkobilov, N.P. (2025). On the dynamics of a quasi-strictly non-Volterra cubic stochastic operator. *Journal of Contemporary Applied Mathematics*, 15(1), 1-9. <https://doi.org/10.62476/jcam.151.1>
- Kruschke, J.K., Liddell, T.M. (2018). The bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a bayesian perspective. *Psychonomic Bulletin Review*, 25, 178–206. <https://doi.org/10.3758/s13423-016-1221-4>
- Leitao, P.J., Schwieder, M., Suess, S., Catry, I., Milton, E.J., Moreira, F., Osborne, P.E., Pinto, M.J., van der Linden, S., & Hostert, P. (2015). Mapping beta diversity from space: Sparse generalised dissimilarity modelling (sgdm) for analysing high-dimensional data. *Methods in Ecology and Evolution*, 6, 764–771. <https://doi.org/10.1111/2041-210X.12378>.
- Lusiana, E.D., Astutik, S., Nurjannah, N., & Sambah, A.B. (2023). On the performance of bayesian generalized dissimilarity model estimator. *Mathematics and Statistics*, 11, 160–168. http://www.hrpub.org/journals/article_info.php?aid=12912

- McElreath, R. (2016). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. CRC Press.
- Mokany, K., Ware, C., Woolley, S. N., Ferrier, S., & Fitzpatrick, M. (2022). A working guide to harnessing generalized dissimilarity modelling for biodiversity analysis and conservation assessment. *Global Ecology and Biogeography*, 31, 802–821.
- Pham, H.T., Pham, H. (2021). On the existence of posterior mean for bayesian logistic regression. *Monte Carlo Methods and Applications*, 27(3), 277–288. <https://doi.org/10.1515/mcma-2021-2089>
- Polson, N.G., Scott, J.G. (2012). On the half-cauchy prior for a global scale parameter. *Bayesian Analysis*, 7, 887–902.
- Ricotta, C. (2017). Of beta diversity, variance, evenness, and dissimilarity. *Ecology and evolution*, 7, 4835–4843. <https://pubmed.ncbi.nlm.nih.gov/28690812><https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5496569/>
- Thomas, S., Tu, W. (2021). Learning hamiltonian monte carlo in r. *The American Statistician*, 75(4), 403–413.. Doi: 10.1080/00031305.2020.1865198. <https://doi.org/10.1080/00031305.2020.1865198>
- White, P.A., Frye, H.A., Slingsby, J.A., Jr, J.A.S., & Gelfand, A.E. (2024). Generative spatial generalized dissimilarity mixed modelling (spgdm): An enhanced approach to modelling beta diversity. *Methods in Ecology and Evolution*, 15, 214–226. <https://doi.org/10.1111/2041-210X.14259>
- Woolley, S.N., Foster, S.D., O’Hara, T.D., Wintle, B.A., & Dunstan, P.K. (2017). Characterising uncertainty in generalised dissimilarity models. *Methods in Ecology and Evolution*, 8, 985–995.