RT&A, No 1 (82)
Volume 20, March 2025

BAYESIAN GLM: A NON-INFORMATIVE APPROACH FOR PARAMETER ESTIMATION IN EXPONENTIAL DISPERSION REGRESSION MODELS

[1*]Ibrahim Sadok & [2]Mourad Zribi

•

[1*]Department of Mathematics and Computer Science, Faculty of Exact Sciences,
University of Bechar, Algeria
[2]Laboratoire d'Informatique Signal et Image de la Côte d'Opale
ULCO, 50 Rue Ferdinand Buisson BP 719, 62228 Calais Cedex, France
ibrahim.sadok@univ-bechar.dz

Abstract

This paper proposes a novel Bayesian approach to parameter estimation in exponential dispersion regression models (EDRM). By employing a non-informative prior distribution, we offer a flexible and robust framework that avoids the need for subjective prior specification. To efficiently sample from the posterior distribution, we develop an importance-sampling algorithm tailored to the EDRM. Through a real-world data analysis, we demonstrate the efficacy of our proposed method in providing accurate and reliable parameter estimates. This research contributes to the advancement of Bayesian statistical modeling techniques and offers valuable insights for practitioners in various fields.

Keywords: Generalized linear models, Bayesian method, Multivariate exponential dispersion, Non-informative prior distribution, Real-world data analysis

1. Introduction

Statistical modelling plays a crucial role in decision-making as it enables the representation of relationships between variables, whether they are linear or non-linear. To establish the connections between observed responses, y_i, and corresponding covariates, x_i, regression models are initially developed. The Ordinary Least Squares (OLS) approach is frequently utilized to estimate unknown parameters, β_i, under the assumption that the response variable adheres to a normal distribution. However, in reality, the normality assumption for residuals may be violated, which leads to the consideration of other exponential family distributions (refer to [25, 26]). In such cases, the use of alternative approaches becomes necessary as OLS estimates may be inaccurate. The Generalized Linear Model (GLM), introduced by J. A. Nelder and R.W. Weddernburn [20], accommodates non-normal distributions of response variables that adhere to exponential family distributions including Poisson, binomial, negative binomial, inverse Gaussian, and gamma distributions. For additional information and practical examples, refer to [13, 14, 19].

The GLM has been extensively studied by researchers for several exponential family distributions. In the health sciences, these models have diverse applications, such as predicting the effect of animal age on dried eye lens weight [23], estimating the prevalence of renal failure based on various parameters (see [32]), modelling lifetime data (see [21]), and addressing transportation challenges (see [34]). As hydrological variables like rainfall and rain-off are inherently positive,

the gamma distribution has proven to be effective in hydrology, as it can model only positive values (as discussed in [18]).

In analytical applications, GLM is utilized to model the relationships between a large number of responses and a single set of predictor factors, as described in ([22]). The GLM typically involves several unknown parameters that are unique to the population. To estimate these parameters, two statistical procedures are commonly used, as noted by W. M. Bolstad and J. M. Curran [3]. The first approach is the conventional technique that relies on all the information obtained from the random sample. A common alternative approach to estimating unknown parameters in GLM is the Bayesian method, which incorporates prior information along with the data from a random sample. The posterior distribution is then derived by combining the likelihood function with the prior distribution, as outlined in standard Bayesian methodologies (see [6], [27, 28]). The choice of prior distribution can significantly influence the posterior estimates, particularly when informative priors are employed. However, it is crucial to note that the use of informative priors is subjective and can have a significant impact on the posterior distribution (see [9]). In situations where limited or no prior knowledge about the parameters is available, non-informative priors can be used. These priors are designed to have minimal influence on the posterior distribution, allowing the data to primarily drive the inference. Non-informative priors, also known as vague or weakly informative priors, are used when we aim to reflect ignorance or neutrality about the parameters prior to observing the data (see [1]).

J. O. Berger and D. Sun [2] conducted research on the types of non-informative prior distributions that could be utilized to enhance the accuracy of normal multivariate models. C. P. Robert *et, al.* [24] provided a comprehensive and contemporary review of Jeffrey's prior distribution.

A study by A. A. I. A. Iswari *et, al.* [14] involved a simple linear regression analysis and the computation of credible intervals for the regression parameters based on simulated data where the prior distribution was not known to the researchers. Additionally, an important contribution to the field is the multivariate multiple linear regression framework, which is a combination of modelling techniques utilizing a Jeffrey's prior distribution [30].

In this paper, we introduce a method for characterizing the GLM using exponential dispersion models, which we refer to as the Exponential Dispersion Regression Models (EDRM). The EDRM represents a rich subclass of the well-known exponential family. Furthermore, we examine the estimation of parameters in the EDRM through maximum likelihood estimation (MLE) and non-informative Bayesian estimation.

This paper is structured as follows: Section 2 details the exponential dispersion regression models. In Section 3, we describe the maximum likelihood estimation and non informative Jeffrey's prior for EDRM. Section 4 is dedicated to providing numerical illustrations, emphasizing the practical application and demonstration of the concepts discussed in earlier sections. Finally, in the last section, we present the conclusion and discussion.

2. Exponential Dispersion Regression Models (EDRM)

In this section, we first describe briefly the exponential dispersion models (EDM). Then, we discuss the EDRMs in details.

2.1. Exponential dispersion models

In the upcoming section, we will delve into the essential features of Exponential Dispersion Models (EDMs) - a noteworthy subset of the renowned exponential family of functions (see [15]). EDMs encompass distinguished distributions such as the inverse Gaussian, gamma, and the normal distribution, to name a few.

EDMs expand upon the concepts of Natural Exponential Families, which offer an extensive range of possibilities, as elucidated in [13]. The probability density function for EDMs is defined as:

$$f(y;\mu,\lambda) = e^{\lambda[\theta y - K_\nu(\theta)]} c(y,\lambda), \quad y \in \mathbb{R} \tag{1}$$

where λ (dispersion parameter) and θ (canonical parameter), with domain $(\lambda, \theta) \in \Lambda \times \Theta \subseteq \mathbb{R}_+ \times \mathbb{R}$. In (1), $K_\nu(\theta) = \log \int_{\mathbb{R}} e^{\theta x} \nu(dx)$ (cumulant function) is a known function of a generating probability measure ν and $c(y, \lambda)$ is a normalizing constant that ensures (1) is a probability function.

For EDMs we have some well-known relations, if $Y \sim f(.; \mu, \lambda)$, then $\mu = \mathbb{E}(Y) = K'_\nu(\theta)$ is the expectation of (1) due to the relationship/map between θ and μ. The variance of (1) is $Var(Y) = \frac{1}{\lambda} V(\mu)$ and $V(\mu)$ being the variance function which uniquely corresponds to an exponential dispersion model. Define $\psi_\nu(\mu) = (K'_\nu(\theta))^{-1}$ and $V(\mu) = K''_\nu(\psi_\nu(\mu))$. It can also be shown that when the functions $K_\nu(.)$ and $c(.,.)$ as well as ψ_ν are fixed, the subfamily arising by taking different θ consists of elements that are all Esscher-transforms of each other. A family with K_ν, c and θ fixed and varying ψ_ν can be generated by the operation of taking sample means. For further information, we refer the reader to [7]).

In Table 1, we present necessary details of absolutely continuous PDFs of the EDM family specifying the normalizing constant ($c(y, \lambda)$), the cumulant function (K_ν), canonical parameter (θ), dispersion parameter (λ), mean (K'_ν), inverse function of the mean (ψ_ν) and variance function (V) of each distribution (see [33, 16]).

Table 1: *Examples of some absolutely continuous PDF of EDMs.*

	Gaussian	**Gamma**	**Inverse Gaussian**	**Laplace**
$c(y, \lambda)$	$\frac{\sqrt{\lambda}}{\sqrt{2\pi}} e^{-\frac{\lambda y^2}{2}}$	$\frac{\lambda^\lambda y^{\lambda-1}}{\Gamma(\lambda)}$	$\frac{\sqrt{\lambda}}{\sqrt{2\pi}} y^{-\frac{3}{2}} e^{-\frac{\lambda}{2y}}$	$\frac{\lambda e^{\lambda y}}{\Gamma(\lambda)^2} \int_{\lambda y}^{+\infty} e^{-2t} t^{\lambda-1} (t - \lambda y)^{\lambda-1} dt$
K_ν	$\frac{\theta^2}{2}$	$-\log(-\theta)$	$-\sqrt{-2\theta}$	$-\log(1-\theta^2)$
K'_ν	θ	$-\frac{1}{\theta}$	$(-2\theta)^{-1/2}$	$\frac{2\theta}{1-\theta^2}$
ψ_ν	μ	$-\frac{1}{\mu}$	$-\frac{1}{2\mu^2}$	$\frac{\left\lvert\sqrt{1+\mu^2}-1\right\rvert}{\mu}$
V	1	μ^2	μ^3	$\frac{\left\lvert\sqrt{1+\mu^2}-1\right\rvert}{\mu^2\sqrt{1+\mu^2}}$

2.2. Description of EDRM

J. A. Nelder and R. W. Wedderburn [20] introduced Generalized Linear Models (GLMs) as a unified framework for handling a variety of commonly used statistical models, including multiple linear regression and log-linear models, for both normally and non-normally distributed data. GLMs are highly versatile, making them suitable for a wide range of models in actuarial statistics, while being structured in a way that allows a single algorithm to be used for maximum likelihood estimation across all of these models (see [19]). GLMs are more helpful in actuarial statistics than ordinary multiple regression, since apart from normal distributions, GLMs explicitly allow Poisson, binomial, gamma and some other useful error distributions. Also, GLMs allow linearity on other scales than the identity scale (logarithmic, logit, probit, reciprocal and others).

The GLMs are described by a link and variance functions further with the selection of a response variable and a collection of explanatory variables. The link function transforms the mean $\mu = (\mu_1, ..., \mu_N)^\top$ of the response variable $Y = (y_1, ..., y_N)^\top$ into a scale where the model is linear. In fact, each response variable y_i is assumed to follow its own regression model, so that

$$y_i = \beta_0 + \sum_{j=1}^{q} \beta_j x_{i,j} + \varepsilon_i \quad i = 1, ..., N, \tag{2}$$

where $x_i = (x_{i1}, x_{i2}, ..., x_{iq})^\top$ be a vector of predictors or covariates, and $\beta = (\beta_1, \beta_2, ..., \beta_q)$ is a vector of unknown parameters of adjacent regression coefficients, and ε_i is a random variables with mean zero and variance σ^2.

In a GLM, we assume that the response variable follows a distribution from the EDMS, satisfying:

$$\psi(\mu_i) = \beta_0 + \beta_1 x_{i1} + ... + \beta_q x_{iq}, \quad i = 1, ..., N \tag{3}$$

where ψ denotes the link function. For simplicity, equation (3) may be also written as $\psi(\mu_i) = x_i^\top \beta = \eta_i$. In what follows, we analyse a data set, $(y_i; x_i) \quad i = 1, ..., N$ with each y_i follows the density function $f(.; \mu, \lambda)$ described in equation (1) with mean and variance as follows

$$\mathbb{E}(\log y_i) = \mu_i = K'_\nu(x_i^\top \beta) = \psi_\nu^{-1}(x_i^\top \beta)$$
$$Var(\log y_i) = \frac{1}{\lambda} K''_\nu(x_i^\top \beta) = \frac{1}{\lambda} V(\mu_i)$$

The multivariate extension relies on the deviance residual, denoted as $r(y, \mu) = \pm\sqrt{d(y, \mu)}$, where $\pm = \text{sgn}(y - \mu)$ and the function d is assumed to be a unit deviance with the property $d(\mu, \mu) = 0$ for $\mu \in \Omega$ (an interval), and $d(y, \mu) > 0$ for $y = \mu$. It is assumed throughout that $d(, \mu)$ is continuous and strictly monotone on each side of μ, implying that $r(, \mu)$ is strictly increasing for each $\mu \in \Omega$.

Let us consider the vector of deviance residuals as $r(y, \mu) = [r(y_1, \mu_1), ..., r(y_N, \mu_N)]^\top$, where y_i and μ_i denote the elements of the N-vectors y and μ, respectively. Given a symmetric positive-definite $N \times N$ matrix Σ, the scaled deviance is defined as the following quadratic form in the vector of deviance residuals:

$$D(y, \mu, \Sigma) = r^\top(y, \mu)\Sigma^{-1} r(y, \mu) = \text{tr}\left[\Sigma^{-1} r(Y, \mu) r^\top(Y, \mu)\right].$$

Following the approach of Jorgensen (1999) and Jorgensen and Lauritzen (2000), a multivariate dispersion model is defined as:

$$f(y; \mu, \Sigma) = a(y, \Sigma) \exp\left[-\frac{1}{2} D(y, \mu, \Sigma)\right] \qquad \text{for } y \in \mathbb{R}^N \tag{4}$$

where $a(y, \Sigma)$ is a suitable function ensuring that (4) is a probability density function on $\mathbb{R}^N$.

3. Parameter estimation of EDRM

3.1. Maximum likelihood estimation

Two types of estimation procedures are known. One is the point estimation and another is interval estimation or confidence interval [29]. Here we mainly focus on point estimation of the parameters associated with a distribution function. This refers to point estimation, where the goal is to approximate an unknown parameter using sample data. Let us consider a random sample $Y = (y_1, y_2, ..., y_N)$ follows $f(., \mu, \lambda)$, where μ and λ are known. In most cases, there are two different approaches for obtaining a point estimator for unknown parameter. Namely classical method and decision theoretic approach. In this section, we focus the estimation of parameter β by the method of maximum likelihood. To estimate this parameter, we fix an underlying exponential dispersion model and common dispersion parameter $\lambda > 0$, but allow each sample y_i its own natural parameter $\theta_i = K'^{-1}(\mu_i)$. Our goal is to estimate the means $\mu_i = \mathbb{E}(y_i)$ for $i = 1, ..., N$ by the maximum likelihood method. Thus we can first estimate the coefficients β, and then use these estimates to argue that $\mu_i = \psi_\nu^{-1}(x_i^\top \beta)$ and the hypothesis of the components independence of response variable y_i, the likelihood function is given by

$$L(\beta, \lambda) = \prod_{i=1}^{N} c(y_i, \lambda) e^{\lambda\left[y_i (x_i^\top \beta)_i - K_\nu((x_i^\top \beta)_i)\right]}. \tag{5}$$

The log-likelihood function is written as

$$\mathcal{L}(\beta, \lambda) = \sum_{i=1}^{N} \left\{\lambda\left[y_i (x_i^\top \beta)_i - K_\nu((x_i^\top \beta)_i)\right] + \log c(y_i, \lambda)\right\} = \sum_{i=1}^{N} l_i(\beta, \lambda) \tag{6}$$

where $l_i(\beta,\lambda) = \left\{\lambda\left[y_i(x_i^\top\beta)_i - K_\nu((x_i^\top\beta)_i)\right] + \log c(y_i,\lambda)\right\}$. It should be to note that $\beta \mapsto \mathcal{L}(\beta,\lambda)$ is a strictly concave function. Therefore, to obtain the maximum likelihood estimator (MLE) of β, we derive the log-likelihood with respect to different components of the vector $\beta = (\beta_0,...,\beta_q)$ and we need to solve the likelihood equations

$$\frac{\partial\mathcal{L}}{\partial\beta_j}(\beta,\lambda) = \sum_{i=1}^{N}\frac{\partial l_i}{\partial\beta_j}(\beta,\lambda) = 0, \quad j = 1,...,q.$$

Since $\dfrac{d\eta_i}{d\mu_i} = \psi'(\mu_i)$, $\dfrac{d\theta_i}{d\mu_i} = 1/V(\mu_i)$ and by the chain rule for differentiation with respect to β_j, we have

$$\begin{aligned}
\frac{\partial l_i}{\partial\beta_j} &= \frac{dl_i}{d\eta_i}\frac{\partial\eta_i}{\partial\beta_j}\\
&= \frac{dl_i}{d\theta_i}\frac{d\theta_i}{d\mu_i}\frac{d\mu_i}{d\eta_i}\frac{\partial\eta_i}{\partial\beta_j}\\
&= \frac{dl_i}{d\theta_i}\left(\frac{d\mu_i}{d\theta_i}\right)^{-1}\left(\frac{d\eta_i}{d\mu_i}\right)^{-1}\frac{\partial\eta_i}{\partial\beta_j}\\
&= \lambda\left(y_i - K'(\theta_i)\right)\left(K''(\theta_i)\right)^{-1}\left(\psi'(\mu_i)\right)^{-1}x_{ij}\\
&= \frac{\lambda\left(y_i-\mu_i\right)x_{ij}}{V(\mu_i)\psi'(\mu_i)}.
\end{aligned}$$

Thus the likelihood equations we have to solve for the MLE of β are

$$\sum_{i=1}^{N}\frac{\lambda\left(y_i-\mu_i\right)x_{ij}}{V(\mu_i)\psi'(\mu_i)} = 0 \quad j = 1,...,q. \tag{7}$$

where μ is the mean vector with N components.

Note that the derivative of the log-likelihood with respect to β does not depend on the dispersion parameter λ.

Remark 1. When the response variable y_i is normally distributed with mean μ_i and dispersion parameter λ, we have $\psi(\mu_i) = \mu_i = x_i^\top\beta$. Hence, $\psi'(\mu_i) = \frac{d\psi(\mu_i)}{d\mu_i} = 1$; also $V(\mu_i) = 1$ and $\lambda = 1/\sigma^2$. Therefore, the equation (7) becomes

$$\frac{1}{\sigma^2}\sum_{i=1}^{N}\left(y_i - x_i^\top\beta\right)x_{ij} = 0.$$

By ignoring the factor $1/\sigma^2$, the equation reduce to the Normal equation of least squares

$$X^\top\left(Y - X\beta\right) = 0 \text{ or quivalently } \widehat{\beta} = \left(X^\top X\right)^{-1}X^\top Y. \tag{8}$$

3.2. Non-informative Bayesian estimation

In this section, we treat the prior of the dispersion parameter λ as known constant. In practice, it is unknown for most cases and it is necessary for us to specify reasonable values. Now, let us focus on specifying the value of λ. In the exponential family, for some distributions, λ is constant, for example, Poisson, Exponential, Bernoulli, Binomial, and Negative Binomial distributions; for other distributions, like Poisson and binomial with over-dispersion, or Normal, gamma, inverse Gaussian, λ is unknown and one may proceed as before with λ replaced by an estimate $\widehat{\lambda}$.

Jeffrey's prior based on the observed Fisher information matrix. Because it is locally uniform, it is a non-informative prior. It is a useful prior because it does not change much over the region in which the likelihood is significant and does not have large values outside that range the

local uniformity property. The Jeffrey's prior is justified on the grounds of its invariance under parametrization according to S. K. Sinha [31].

In this step, we assess Jeffrey's prior regression for the GLM class and the associated prior and posterior distribution characteristics. Our key subject is the case where λ is unknown. The Jeffrey's prior can be an enticing one for the normal linear regression model as it corresponds to tractable posterior distributions. See G. E. Box and G. C. Tiao [4] for more details about the usage of the Jeffrey's prior in this model. In addition, D. M. Eaves [8] assumed Jeffrey's prior to a non-linear phenomenon and obtained tractable posteriors through linearisation of the non-linear model. However, Jeffrey's prior can indicate that it leads to proper posteriors in many GLMs. The two theorems 1 and 2 are given by [13] below that help to evaluate and establish the propriety of the posterior distribution under Jeffrey's prior by giving (i) sufficient and (ii) necessary and sufficient conditions for the propriety of the posterior and prior distributions, respectively. The two theorems 1 and 2 discuss also the existence of joint moments.

Theorem 1. Suppose the likelihood and Jeffrey's prior for β are as above. Additionally, assume that X is of full rank and the likelihood of β is bounded above. Then, a sufficient condition for the existence of the posterior moment generating function of β for any GLM is that the integral

$$\int_S e^{\psi_\nu \theta^{-1}(r)+\phi^{-1}w(yr-K_\nu(r))}\left(\frac{d^2K_\nu(r)}{dr^2}\right)^{\frac{1}{2}}dr \tag{9}$$

is finite, where $\lambda(\phi)=\frac{\phi}{w}$ and for ψ_ν in some open neighbourhood about 0. Here S denotes the parameter space for the canonical parameter θ.

Theorem 2. A necessary and sufficient condition for existence of moment generating function of Jeffrey's prior for any GLM is that the integral

$$\int_S e^{\psi_\nu \theta^{-1}(r)}\left(\frac{d^2K_\nu(r)}{dr^2}\right)^{\frac{1}{2}}dr \tag{10}$$

is finite for ψ_ν in some open neighbourhood about 0.

It should be to note that tractable posteriors for GLM's with Jeffrey's prior are valid for certain examples only in very particular cases, and closed type outcomes in general are not available. Let us assume now $y_1,...,y_N$ are independent observations from a GLM. Jeffrey's joint prior for (λ,β) is given by $\pi(\lambda,\beta)=|I(\lambda,\beta)|^{\frac{1}{2}}$, where $I(\lambda,\beta)$ is the Fisher information matrix of (λ,β). From equation (1), it is obvious that $I(\lambda,\beta)$ is diagonal block in λ and β, with the form

$$\begin{pmatrix} -\sum_{i=1}^N \frac{2K'_\nu(\theta_i)\theta_i-K_\nu(\theta_i)}{\lambda^3}+\mathbb{E}(\ddot{c}(y_i,\lambda)) & 0 \\ 0 & \frac{X^\top V(\beta)\Delta^2(\beta)X}{\lambda} \end{pmatrix},$$

where $\ddot{c}(y_i,\lambda)=\frac{\partial^2 c(y_i,\lambda)}{\partial\lambda^2}$, $V(\beta)$ and $\Delta(\beta)$ are $N\times N$ diagonal matrices defined as $V(\beta)=Diag\left(K''_\nu\left(\psi(x_1^\top\beta)\right),...,K''_\nu\left(\psi(x_N^\top\beta)\right)\right)$ and $\Delta(\beta)=Diag\left(\psi'(x_1^\top\beta),...,\psi'(x_N^\top\beta)\right)$. Therefore, Jeffrey's prior for (λ,β) is driven as follows

$$\pi(\lambda,\beta)\propto\left(-\sum_{i=1}^N\frac{2K'_\nu(\theta_i)\theta_i-K_\nu(\theta_i)}{\lambda^3}+\mathbb{E}(\ddot{c}(y_i,\lambda))\right)^{\frac{1}{2}}\lambda^{-\frac{q}{2}}\left|X^\top V(\beta)\Delta^2(\beta)X\right|^{\frac{1}{2}}. \tag{11}$$

We can see from equation (11) that in fact, λ and β under Jeffrey's prior are not independent. In the normal linear model of regression: $\pi(\lambda,\beta)\propto\lambda^{-\frac{(q+2)}{2}}$ and in this case, λ and β are independent. On the contrary, for the gamma model λ and β are not. In fact, Jeffreys joint prior for λ and β may be quite difficult to analyse for many GLMs, as well as the gamma model, which may not be feasible for number calculation. Conversely, we can find the following joint prior to (λ,β): suppose that λ and β are independent, choose a (possibly informative) Jeffrey's prior to λ

and $\frac{X^\top V(\beta)\Delta^2(\beta)X}{\lambda}$ as Jeffrey's prior to β. This is one method for choosing an analytically feasible joint prior (λ, β), which simultaneously enhances Jeffrey's prior.

Maximum a-posteriori probability (MAP) estimate, like the maximum likelihood method, is a method that can be used to estimate a number of unknown parameters, such as the parameters of a probability density, related to a sample given. This method is closely linked to the maximum likelihood but differs from it however by the possibility of taking into account a non-informative a prior on the parameters to be estimated.

In MAP estimation, the model parameters are obtained by solving

$$\widehat{\underline{\theta}}_{MAP} = \underset{\underline{\theta}}{\operatorname{argmax}}\, \mathcal{L}(\underline{\theta}) + \log \pi(\underline{\theta}) \tag{12}$$

where $\underline{\theta} = (\lambda, \beta)$. Our goal is to solve the maximization in (12) when Jeffrey's prior is considered. According to Jeffrey's prior, the probability of the prior is proportional to the square root of the determinant of the Fisher information matrix I:

$$\pi(\underline{\theta}) \propto \sqrt{\det I(\lambda, \beta)} \tag{13}$$

Substituting (11) into (12) and removing terms which are independent of $\underline{\theta}$, we obtain

$$\widehat{\underline{\theta}} = \underset{\underline{\theta}}{\operatorname{argmax}}\, \mathcal{L}(\underline{\theta}) + \frac{1}{2}\log\left(-\sum_{i=1}^{N}\frac{2K_\nu'(\theta_i)\theta_i - K_\nu(\theta_i)}{\lambda^3} + \mathbb{E}(\ddot{c}(y_i,\lambda))\right) - \frac{q}{2}\log(\lambda) \tag{14}$$

$$+ \frac{1}{2}\log\left|X^\top V(\beta)\Delta^2(\beta)X\right| \tag{15}$$

Proposition 1. The coefficients regression estimates under Jeffrey's prior can be presented as

$$\widehat{\beta}^{(l+1)} = \widehat{\beta}^{(l)} + \lambda^{(l)} X^\top (Y - \mu) + \frac{X^\top \psi''(X^\top \beta(l)) X^\top X}{X^\top \psi'(X^\top \beta^{(l)}) X}$$

$$\widehat{\lambda}^{(l+1)} = \widehat{\lambda}^{(l)} + \sum_{i=1}^{N}\frac{1}{c(y_i,\lambda^{(l)})}\frac{\partial c(y_i,\lambda^{(l)})}{\partial \lambda} + \sum_{i=1}^{N} y_i (x_i^\top \beta^{(l)})_i - K_\nu(x_i^\top \beta^{(l)}) - \frac{q+2}{2\lambda}$$

where λ represents the dispersion parameter.

Proof.

In order to estimate the coefficients regression, we adopt the Gradient ascent approach (see [5]). Then, the coefficients regression and the dispersion parameter can be calculated respectively, as

$$\widehat{\beta}^{(l+1)} = \widehat{\beta}^{(l)} + \frac{\partial \log L(\beta^{(l)}, \lambda^{(l)})}{\partial \beta}$$

$$\widehat{\lambda}^{(l+1)} = \widehat{\lambda}^{(l)} + \frac{\partial \log L(\beta^{(l)}, \lambda^{(l)})}{\partial \lambda}$$

According to (14), the derivative of $\log l(\underline{\theta})$ is

$$\frac{d \log l(\underline{\theta})}{d\underline{\theta}} = \frac{d\mathcal{L}(\underline{\theta})}{d\underline{\theta}} + \frac{d \log \pi(\underline{\theta})}{d\underline{\theta}}$$

where the derivative of $\log l(\underline{\theta})$ with respect to β and λ

is obtained as follows

$$\frac{d\mathcal{L}(\underline{\theta})}{d\beta} = \lambda X^\top (Y - \mu) \tag{16}$$

$$\frac{d\mathcal{L}(\underline{\theta})}{d\lambda} = \sum_{i=1}^{N}\frac{1}{c(y_i,\lambda)}\frac{\partial c(y_i,\lambda)}{\partial \lambda} + \sum_{i=1}^{N} y_i (x_i^\top \beta)_i - K_\nu(x_i^\top \beta) \tag{17}$$

The derivative of $\log \pi(\underline{\theta})$ with respect to β and λ is given by

$$\frac{\partial \log \pi(\underline{\theta})}{\partial \beta} = \frac{\partial}{\partial \beta} \log \frac{X^\top V(\beta)\Delta^2(\beta)X}{\lambda} = \frac{X^\top \psi''(X^\top \beta)X^\top X}{X^\top \psi'(X^\top \beta)X}$$
$$\frac{\partial \log \pi(\underline{\theta})}{\partial \lambda} = \frac{\partial}{\partial \lambda} \log \lambda^{-\frac{(q+2)}{2}} = -\frac{q+2}{2\lambda}$$

From equations (16), (17) and the derivative of $\log \pi(\underline{\theta})$, the specific update for estimating $\widehat{\beta}^{(l+1)}$ and $\widehat{\lambda}^{(l+1)}$ in linear regression can be expressed as

$$\widehat{\beta}^{(l+1)} = \widehat{\beta}^{(l)} + \lambda^{(l)} X^\top (Y - \mu) + \frac{X^\top \psi''(X^\top \beta^{(l)})X^\top X}{X^\top \psi'(X^\top \beta^{(l)})X}$$
$$\widehat{\lambda}^{(l+1)} = \widehat{\lambda}^{(l)} + \sum_{i=1}^{N} \frac{1}{c(y_i, \lambda^{(l)})} \frac{\partial c(y_i, \lambda^{(l)})}{\partial \lambda} + \sum_{i=1}^{N} y_i (x_i^\top \beta^{(l)}) - K_\nu(x_i^\top \beta^{(l)}) - \frac{q+2}{2\lambda^{(l)}}$$

■

Remark 2. In order to estimate the coefficients regression and dispersion parameter, we proposed an iterative algorithm based on the Gibbs-sampling one. The Gibbs sampling algorithm was introduced by S. Geman and D. Geman [10]. We start by setting the coefficients regression and dispersion parameter to its initial values $\beta^{(0)}$ and $\lambda^{(0)}$, respectively. This process continues until "convergence" (i.e; $\left|\beta^{(l+1)} - \beta^{(l)}\right| < \epsilon_\beta$ and $\left|\lambda^{(l+1)} - \lambda^{(l)}\right| < \epsilon_\lambda$) for ϵ_β and ϵ_λ are small enough.

3.3. Application

The Bayesian approach is employed to estimate a multivariate multiple regression model using a non-informative Jeffrey's prior. The dataset consists of 100 synthetically generated observations, created from a multivariate normal distribution using Matlab software.

Table 2: *Parameter estimation*

	Parameter	Mean	Credible Interval
MLE	$\widehat{\beta}_0$	1.1319	(0.5497 1.7141)
	$\widehat{\beta}_1$	-0.9689	(−1.1301 − 0.8077)
	$\widehat{\beta}_2$	0.69819	(0.4964 0.9000)
	$\widehat{\lambda}$	2.8041	(−1.0375 6.6457)
		AIC = 212.8195	BIC= 220.2481
Jeffrey's prior	$\widehat{\beta}_0$	1.0462	(0.4899 1.6025)
	$\widehat{\beta}_1$	-0.9507	(−1.1102 − 0.7912)
	$\widehat{\beta}_2$	0.7129	(0.5167 0.9091)
	$\widehat{\lambda}$	2.7965	(−1.0451 6.6381)
		AIC =212.8161	BIC= 220.0107

The parameter estimates are presented in Table 2, along with their 95% credible intervals.

It is important to note that this analysis is based on a single simulated dataset. While both the Bayesian approach with Jeffrey's prior and the classical method yield similar results in this case, drawing strong conclusions from a single dataset is limited. A more rigorous comparison would require a complete Monte Carlo simulation study, where multiple datasets are generated and analysed to account for the variability introduced by the random generation of data.

Conducting such a simulation study, with several iterations, would provide a more comprehensive evaluation of the performance of both methods. Additionally, we observe that Jeffrey's prior has some advantages in terms of model fit, with slightly lower AIC and BIC values, and may offer more stable estimates when there is limited prior knowledge or small sample sizes. However, further validation through a more extensive simulation study is necessary to confirm these findings.

4. Numerical Illustration

Universal processes of rainfall involve data collection, data preprocessing and data selection,building a model using regression, and at the last validity check. The areal rainfall estimated by the rain gauges presents a great uncertainty where the network of rain gauges is sparse. From this approach, we can predict the rainfall of any future year using climatic factors. In our study, we have chosen an application concerning daily climatic data for the studied regions of northwestern Algeria. These data were extracted from the National Office of Meteorology (https://www.meteo.dz) and TuTiempo (https://en.tutiempo.net/climate), 2021.

For all existing rainfall stations (more than 10 climatic stations), the annual rainfall averages are available for a period varying from 35 to 40 complete years, from 1981 to 2021. Most of the stations are located in the plains and on the coasts, the number of stations decreases towards the south and in mountainous regions (more than 600m). The main available factors that depend on precipitation are temperature values, wind speed, station elevation, and station coordinates such as latitude and longitude (see Table 3).

Table 3: *Data of the topographic parameters at the study rainfall stations.*

Rainfall stations	Rainfall (mm)	Temperature (°C)	Wind speed (m/s)	Elevation (m)	Latitude (°)	Longitude (°)
Ain Sefra	0.5546	16.5686	5.4929	1200.55	32,76	-0.6
El-Bayadh	0.6944	15.4328	6.1949	1220.02	33,66	1
Mascara	1.4003	17.2871	5.2274	591.71	35.21	0.15
Mostaganem	1.2537	18.5080	5.2992	254.48	35,88	0.11
Oran	1.1661	18.6619	5.7160	127.31	35.63	-0.6
Relizene	1.3584	18.2210	5.2587	367.72	35.73	0.55
Saida	1.2480	15.9214	5.4670	928.27	34.86	0.15
Sidi Bel Abbes	1.3186	17.6557	5.3304	458.05	35.19	-0.64
Tiaret	1.4860	15.6170	5.6113	904.3	35.35	1.43
Tlemcen	1.2399	18.0853	5.3669	355.91	35.01	-1.46

In the present investigation, we have applied a multiple regression model, which has been elaborately described in Section 2.2, to the data set at hand. The purpose of this exercise is to unearth a predictable equation that can establish a link between rainfall and various climatic factors. In order to accomplish this goal, we have meticulously constructed a regression equation comprising five independent variables as follows:

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \beta_4 x_{i,4} + \beta_5 x_{i,5} + \varepsilon_i, \quad i = 1, ..., 10$$

where y_i is the response (predicted rainfall station), $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ are the regression coefficients, x_1 (Temperature), x_2 (Wind speed), x_3 (Elevation), x_4 (Latitude), x_5 (Longitude) are

the highly correlated climate indices (predictors); and ε is the residual term of the model.

Let us now proceed to the multiple regression analysis with the full climatic factors considered above and compare it with different types of model estimation using maximum likelihood and Jeffrey's prior estimations. Table 4 shows the regression comparison with the five independent variables. The main features of interest are the parameters estimates as well as their corresponding standard errors (SE Coeff) and the criteria of fitness (t-Stat and p-Value). the standard error of the regression coefficient is calculated as SE Coeff $= \frac{s.e._R}{sx\sqrt{N-1}}$, where $s.e._R$ is the standard deviation of the regression, sx is the standard deviation of x, and N is the sample size.

Moreover, the estimates are significantly similar and its associated standard error are close to zero, which justifies the very smaller over variation. Since the associated p-value is < 0.001, we reject the hypothesis in favour of the alternative hypothesis that at least one of the coefficients is not zero.

Table 4: *Multiple regression analysis with MLE and Jeffrey's prior approaches.*

	Predictor	Coeff $\widehat{\beta}$	SE Coeff $\widehat{\beta}$	t-Stat	p-Value
MLE	Intercept	5.14787×10^{-15}	0.03227271	1.59511×10^{-13}	0.99999999
	Temperature	-2.82471635	0.42179987	-6.69681646	0.00258658
	Wind speed	-0.81015953	0.10066482	-8.04808997	0.00129405
	Elevation	-4.06167602	0.77057323	-5.27097985	0.00620801
	Latitude	-1.22618013	0.44892198	-2.73138802	0.05236798
	Longitude	0.58214495	0.15687036	3.71099388	0.02063610
	RMSE= 0.102	R^2 =0.995	$R^2(adj)$ = 0.99	AIC=-14.42894643	
Jeffrey's prior	Intercept	5.59885×10^{-15}	0.03227271	1.73485×10^{-13}	0.99999999
	Temperature	-2.82465611	0.42179987	-6.69667364	0.00258679
	Wind speed	-0.81014627	0.10066482	-8.0479583	0.00129413
	Elevation	-4.06156466	0.77057324	-5.27083533	0.00620862
	Latitude	-1.22611567	0.44892198	-2.73124443	0.05237573
	Longitude	0.58212371	0.15687036	3.71085846	0.02063855
	RMSE= 0.10205529	R^2 =0.99536930	$R^2(adj)$ = 0.98958092	AIC=-14.42894637	

Upon examination of Table 4, we observe that the estimated values of β_i are strikingly alike and the associated standard deviations are nearly zero, which validates the incredibly low variance of the model. With a p-value greater than 0.001, we reject the null hypothesis and favour the alternative hypothesis that at least one of the coefficients is not zero.

Additionally, Table 4 displays the Root Mean Squared Errors (RMSE) for the variance of the residuals, calculated by both MLE and Jeffrey's prior estimation methods. This metric evaluates absolute fit of the model to the data, indicating the proximity of the observed data points to predicted values of the model. Interpreted as the standard deviation of the unexplained variance, the RMSE is expressed in the same units as the response variable. Lower RMSE values for both MLE and Jeffrey's prior estimation methods indicate better fit, supporting the equivalence of the two approaches. The RMSE serves as a reliable measure of the model's predictive accuracy, making it the most critical criterion for fit if the primary goal is prediction.

Table 4 indicates that other fitness criteria, such as R^2, $R^2(adj)$, and AIC, are reasonably similar for both MLE and Jeffrey's prior regression models. Figure 1 presents the estimates for each of the β_i coefficients of the regressions with five independent variables using both MLE and Jeffrey's prior regression models. It is evident that all the parameters converge rapidly to the MLE method as the number of iterations increases.

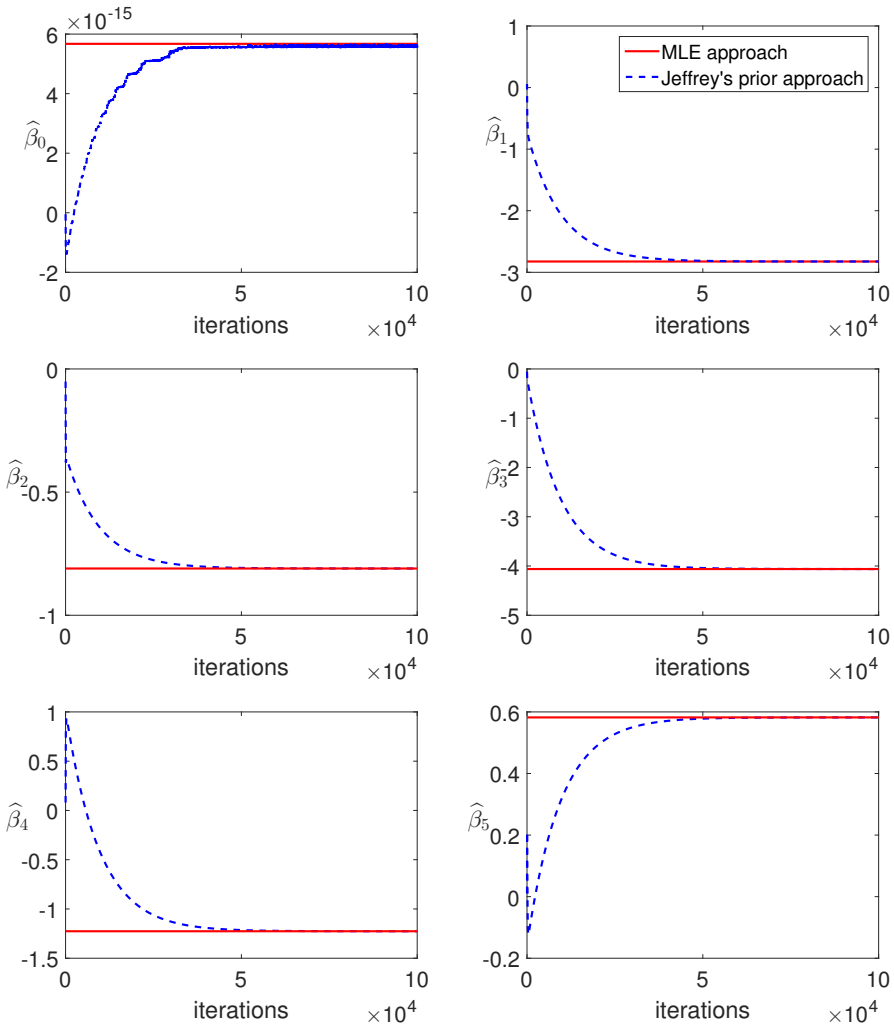

Figure 1: *Relationship between regression coefficient estimation and iteration number vs MLE method*

5. Conclusion and Discussion

In this paper, we apply the Bayesian estimation of EDRM parameters using Jeffrey's prior. We demonstrated that the maximum likelihood method can be used to accurately estimate this parameter. Additionally, we devised an iterative algorithm based on Jeffrey's prior to estimate the regression coefficients. The significance of rainfall on agriculture and global economies cannot be overstated, and accurate predictions of rainfall are essential for successful farming practices. While this model is currently the only one capable of predicting rain, it is not entirely precise due to the fluctuation of climatic variables. Although our study includes certain elements, other factors may also impact rainfall amounts. Nevertheless, our proposed technique shows promise, particularly when utilizing Jeffrey's prior. The results we obtained were comparable to those of the MLE method, validating the effectiveness of our approach.

References

[1] Berger, J. (2006). The case for objective Bayesian analysis.
[2] Berger, J. O., & Sun, D. (2006). Objective priors for a bivariate normal model with multivariate generalizations. Duke University ISDS Technical Report 2007, 6.
[3] Bolstad, W. M., & Curran, J. M. (2016). Introduction to Bayesian statistics. John Wiley & Sons.

[4] Box, G. E., & Tiao, G. C. (2011). Bayesian inference in statistical analysis. John Wiley & Sons.
[5] Boyd, S., & Vandenberghe, L. (2004). Convex optimization. Cambridge university press.
[6] Carlin, B. P., & Louis, T. A. (2008). Bayesian methods for data analysis. CRC press.
[7] Kaas, R., Goovaerts, M., & Dhaene, J. (2008). Modern actuarial risk theory. Springer Science & Business Media.
[8] Eaves, D. M. (1983). On Bayesian nonlinear regression with an enzyme example. Biometrika, 70(2), 373-379.
[9] Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). Bayesian data analysis. Chapman and Hall/CRC.
[10] Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Transactions on pattern analysis and machine intelligence, (6), 721-741.
[11] Guure, C. B., Ibrahim, N. A., & Adam, M. B. (2013). Bayesian Inference of the Weibull Model Based on Interval?Censored Survival Data. Computational and Mathematical Methods in Medicine, 2013(1), 849520.
[12] Hashimoto, S., & Koike, K. (2014). Non-informative prior with maximum divergence for nonregular distributions. International Journal of Applied & Experimental Mathematics.
[13] Ibrahim, J. G., & Laud, P. W. (1991). On Bayesian analysis of generalized linear models using Jeffreys's prior. Journal of the American Statistical Association, 86(416), 981-986.
[14] Iswari, A. A. I. A., Sumarjaya, I., & Srinadi, I. G. A. M. (2014). Analisis Regresi Bayes Linear Sederhana dengan Prior Noninformatif. E-Jurnal Matematika, 3(2).
[15] J?rgensen, B. (1987). Exponential dispersion models. Journal of the Royal Statistical Society Series B: Statistical Methodology, 49(2), 127-145.
[16] Jorgensen, B. (1997). The theory of dispersion models. CRC Press.
[17] Kass, R. E., & Wasserman, L. (1996). The selection of prior distributions by formal rules. Journal of the American statistical Association, 91(435), 1343-1370.
[18] Markovic, R. D. (1965). Probability functions of best fit to distributions of annual precipitation and runoff.
[19] McCullagh, P. (2019). Generalized linear models. Routledge.
[20] Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. Journal of the Royal Statistical Society Series A: Statistics in Society, 135(3), 370-384.
[21] Ortega, E. M., Cordeiro, G. M., Pascoa, M. A., & Couto, E. V. (2012). The log-exponentiated generalized gamma regression model for censored data. Journal of Statistical Computation and Simulation, 82(8), 1169-1189.
[22] Pituch, K. A., & Stevens, J. P. (2015). Applied multivariate statistics for the social sciences: Analyses with SAS and IBM's SPSS. Routledge.
[23] Ratkowsky, D. A. (1983). Nonlinear Regression Modelling. Marcel Decker. Inc., New York.
[24] Robert, C. P., Chopin, N., & Rousseau, J. (2009). Harold Jeffreys's theory of probability revisited. Statistical Science, 141-172.
[25] Sadok, I., Masmoudi, A., & Zribi, M. (2023). Integrating the EM algorithm with particle filter for image restoration with exponential dispersion noise. Communications in Statistics-Theory and Methods, 52(2), 446-462.
[26] Sadok, I., & Masmoudi, A. (2022). New parametrization of stochastic volatility models. Communications in Statistics-Theory and Methods, 51(7), 1936-1953.
[27] Sadok, I., Zribi, M., & Masmoudi, A. (2023). Non-informative Bayesian estimation in dispersion models. Hacettepe Journal of Mathematics and Statistics, 53(1), 251-268.
[28] Sadok, I. (2023). Non-informative Bayesian dispersion particle filter. Journal of Innovative Applied Mathematics and Computational Sciences, 3(2), 173-189.
[29] Sadok, I. (2024). ON THE Q-RAYLEIGH DISTRIBUTION AND ITS APPLICATIONS. Reliability: Theory & Applications, 19(1 (77)), 588-602.
[30] Saputro, D. R. S., Amalia, F., Widyaningsih, P., & Affan, R. C. (2018, May). Parameter estimation of multivariate multiple regression model using bayesian with non-informative

Jeffreys' prior distribution. In Journal of Physics: Conference Series (Vol. 1022, No. 1, p. 012002). IOP Publishing.

[31] Sinha, S. K. (1986). Bayes estimation of the reliability function and hazard rate of a Weibull failure time distribution. Trabajos de Estadistica, 1(2), 47-56.

[32] Staniswalis, J. G. (2006). On fitting generalized non-linear models with varying coefficients. Computational statistics & data analysis, 50(7), 1818-1839.

[33] Tweedie, M. C. (1984, December). An index which distinguishes between some important exponential families. In Statistics: Applications and new directions: Proc. Indian statistical institute golden Jubilee International conference (Vol. 579, pp. 579-604).

[34] Wood, G. R. (2005). Confidence and prediction intervals for generalised linear accident models. Accident Analysis & Prevention, 37(2), 267-273.

[35] Yang, R., & Berger, J. O. (1996). A catalog of noninformative priors (Vol. 1018). Durham, NC, USA: Institute of Statistics and Decision Sciences, Duke University.