American Journal of Sciences and Engineering Research

E-ISSN– 2348–703X, Volume 8, Issue 5, 2025

Bayesian Inference in The Linear Regression Model

Théophile RABENANTENAINA[1], Parfait BEMARISIKA[2], André TOTOHASINA[3]

[1]*Mixed High School Besalampy, B.P.O, Besalampy 210, Madagascar.*
[2,3]*Higher Normal School for Technical Education, University of Antsiranana, B.P.O, Antsiranana 201, Madagascar.*

ABSTRACT: *This article studies linear regression in Bayesian inference because it provides greater flexibility and objectivity when analyzing statistical data. We have attempted to explain the concepts used to study Bayesian regression (prior, posterior, likelihood, MCMC simulation methods, etc.). Numerical applications and simulations have illustrated the methodology in different models for calculating the posterior in multiple Bayesian regression.*

Keywords: *Bayesian inference; prior distribution; posterior distribution; Bayesian linear regression.*

I. INTRODUCTION

Linear regression analysis is one of the most commonly used in statistical methods for modeling cross-sectional data. Building a regression model involves estimating the parameters of that model. This allows us to obtain the regression coefficient for each independent variable. Several methods can be used to estimate these parameters. The method most frequently used by researchers is the frequentist/classical method, which uses OLS (Ordinary Least Squares) or MLE (Maximum Likelihood Estimation). The OLS method, also known as the least squares method, consists of minimizing the number of errors in the regression equation. The parameters of the regression model are thus obtained by minimizing the error function of the equation. The MLE method, on the other hand, consists of minimizing the probability density function of a given data set. When using these two methods, there are classical assumptions that must be satisfied based on the results of the regression modelling. These assumptions include error independence, identity, and normal distribution. In practice, regression coefficients are often assumed to be constant. However, in theory, models with one or more varying (i.e., non-constant) parameters are suggested. The parameter of interest in this analysis is the breakpoint, which indicates where and when the change occurs.

In addition to these two methods, other methods can be used to estimate the parameters of the regression model, such as the Bayesian approach. The difference between frequentist and Bayesian methods lies in the perspective adopted regarding the parameters. The Bayesian approach considers parameters as random variables, meaning that their value is not unique, unlike the frequentist approach. In this regard, we can say that the latter method is the best of the three. In this article, we will study and implement Bayesian linear regression models on different datasets. To do this, we will use Bayes' theorem. We will begin by examining some basic concepts related to Bayesian linear regression.

II. BAYESIAN ESTIMATION OF A LINEAR REGRESSION

Bayesian analysis, as developed by Bayes (1763) and Laplace (1995), begins with an examination of a given situation and the identification of an uncertainty pertaining to an unknown parameter θ. This uncertainty is then quantified through the application of probabilistic distributions, utilizing fundamental principles of probability

 Received-13-08-2025 Accepted- 30-08-2025

calculus. The uncertainty about θ is modeled in the form of a distribution, known as prior distribution, which provides information about θ taken as a constant. This prior distribution is updated by extracting information from the observations of the variable X, to obtain another master distribution known as the posterior distribution. In this section, we will apply the Bayesian approach to regression models, as first initiated by Jeffreys (1939).

2.1. Linear regression model (Koch, 2007)

In a population, we wish to predict the values of a quantitative variable $y = (y_1, y_2, \dots, y_n)'$ from the values of $p-1$. Other variables $X_1, \dots, X_{p-1}$. This is equivalent to explaining variations in y from those in $X_1, \dots, X_{p-1}$. We then say that we wish to explain y from $X_1, \dots, X_{p-1}$; so y is called the "variable to be explained" and $X_1, \dots, X_{p-1}$ are called the "explanatory variables". The data available are n observations of $(y, X_1, \dots, X_{p-1})$ noted $(y_1, x_{1,1}, \dots, x_{1,p-1}), \dots, (y_n, x_{n,1}, \dots, x_{n,p-1})$. The linear regression model is written:

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots + \beta_{p-1} x_{i,p-1} + \varepsilon_i, \quad i = 1, \dots, n. \qquad (1)$$

The regression model can be written in matrix from:

$$y = X\beta + \varepsilon, \qquad (2)$$

where

- $X = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,p-1} \\ 1 & x_{2,1} & \dots & x_{2,p-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,p-1} \end{bmatrix}$ is the explanatory variable matrix $(n \times p)$;
- $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}$ is the vector of unknown regression model parameters $(p \times 1)$;
- $\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$ is the error vector $(n \times 1)$;

With $X\beta = E(y|\beta)$ and $D(y|\sigma^2) = D(\varepsilon|\beta, \sigma^2) = \sigma^2 P^{-1}$ the variance-covariance matrix of y and P the known positive-definite observation weight matrix. Therefore, $y|\beta, \sigma^2 \sim \mathcal{N}(X\beta, \sigma^2 P^{-1})$.

NB: Before deciding that $y = X\beta + \varepsilon$, we first had to check that the scatterplot $\{(y_i, x_{i,1}, \dots, x_{i,p-1}),\ i = 1, \dots, n\}$ showed linear regression.

Thus, the likelihood function of y knowing β and σ^2 is defined as follows:

$$f(y|\beta, \sigma^2) = (2\pi\sigma^2)^{-n/2} (\det(P))^{1/2} exp\left\{-\frac{1}{2\sigma^2}(y - X\beta)'P(y - X\beta)\right\}. \qquad (3)$$

In order to make an inference, estimate of β are obtained by maximizing the likelihood function or its logarithm. For mathematical convenience, when the values of the parameters that maximize the two functions are the same, the ln −likelihood is commonly used, given by

$$\begin{aligned} \ln\left(f(y|\beta, \sigma^2)\right) &= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) + \frac{1}{2}\ln(\det(P)) - \frac{1}{2\sigma^2}(y - X\beta)'P(y - X\beta) \\ &= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) + \frac{1}{2}\ln(\det(P)) - \frac{1}{2\sigma^2}(y'Py - 2\beta'X'Py + \beta'X'PX\beta) \qquad (4) \end{aligned}$$

Let's derive (4) with respect to the variable β and we have: $\frac{\partial \ln\left(f(y|\beta,\sigma^2)\right)}{\partial \beta} = -\frac{1}{2\sigma^2}(-2X'Py + X'PX\beta) = 0$. And we obtain $\hat{\beta} = (X'PX)^{-1}X'Py$, the maximum likelihood estimator of β. Similarly, for the parameter σ^2:

$$\frac{\partial \ln\left(f(y|\beta, \sigma^2)\right)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}(y - X\beta)'P(y - X\beta) = 0. \qquad (5)$$

Let $\hat{\sigma}^2 = \frac{1}{n}(y - X\beta)'P(y - X\beta)$, the maximum likelihood estimator of σ^2.

2.2. Noninformative prior

In the Bayesian approach, it is necessary and crucial to determine the prior distribution, denoted by $\pi(\theta)$, of the

parameter θ. However, in practice, prior information is often inadequate, which complicates and impedes the selection of an appropriate prior distribution. In scenarios where resources or time are limited, researchers may find themselves unable to construct an accurate prior, compelling them to rely on the partial information provided by the model data. This reliance on partial information leads to the adoption of an noninformative prior distribution. If nothing is known in advance about the unknown parameter θ, it can take on values between $-\infty$ and $-\infty$. Its noninformative prior $\pi(\theta)$ is then assumed to be

$$\pi(\theta) \propto const \;\; for \;\; -\infty < \theta < +\infty, \qquad (6)$$

where $const$ denotes a constant. The density is improper density function, since with $\int_{-\infty}^{+\infty} \pi(\theta) \neq 1$. So, if an unknown parameter like the variance σ^2 can only take on values between 0 and $+\infty$, we set

$$\theta = \ln \sigma^2 \qquad (7)$$

and again $\pi(\theta) \propto const \;\; for \;\; -\infty < \theta < +\infty$. By the transformation of θ to σ^2 with $\frac{d\theta}{d\sigma^2} = \frac{1}{\sigma^2}$ from (7), see for instance Koch, (1999), the noninformative prior for the variance σ^2 follows by

$$\pi(\sigma^2) \propto \frac{1}{\sigma^2} \; for \; 0 < \sigma^2 < +\infty. \qquad (8)$$

Very often, it is more convenient to introduce the weight or precision parameter τ instead of σ^2 with $\tau = 1/\sigma^2$. By transforming of σ^2 to τ with $\frac{d\sigma^2}{d\tau} = -\frac{1}{\tau}$ the noninformative prior density function for τ follows instead of (8) by

$$\pi(\tau) \propto \frac{1}{\tau} \; for \; 0 < \tau < +\infty \qquad (9)$$

The noninformative prior density function (6) which is determined by a constant is selected for the vector β of unknown parameters and the noninformative prior density function (9), which is proportional to $1/\tau$, for the weight parameter τ. Thus, we have the joint prior distribution between the parameters β and τ: $\pi(\beta, \tau) = \pi(\beta)\pi(\tau) \propto \frac{1}{\tau}$. Subsequently, using Bayes' theorem (Ando, 2010) we obtain the posterior joint distribution $\pi(\beta, \tau|y)$:

$$\pi(\beta, \tau|y) \propto (2\pi)^{-n/2}\tau^{n/2} exp\left\{-\frac{\tau}{2}(y - X\beta)'P(y - X\beta)\right\}$$

$$\propto \tau^{n/2-1} exp\left\{-\frac{\tau}{2}(y - X\beta)'P(y - X\beta)\right\} \qquad (10)$$

The exponent of the term in (10) can be written as follows:

$$\begin{aligned}(y - X\beta)'P(y - X\beta) &= y'Py - 2\beta'X'Py + \beta'X'PX\beta \\ &= y'Py - 2(\mu^*)'X'Py + (\mu^*)'X'PX\mu^* + (\beta - \mu^*)'X'PX(\beta - \mu^*) \\ &= (y - X\mu^*)'P(y - X\mu^*) + (\beta - \mu^*)'X'PX(\beta - \mu^*) \qquad (11)\end{aligned}$$

with $\mu^* = (X'PX)^{-1}X'Py$.

Remark 2.1. *Let Y be an $m \times 1$ random vector and X a random variable. Assume that $Y, X \sim NG(\mu, V, a, b)$, so the joint density function $f(y, x|\mu, V, a, b)$ is written as (Koch, 2007):*

$$f(y, x|\mu, Va, b) = (2\pi)^{m/2}(det\, V)^{-1/2}a^b\left(\Gamma(b)\right)^{-1}$$

$$\times x^{-m/2+b-1} exp\left\{-\frac{x}{2}[2a + (y - \mu)^T V^{-1}(y - \mu)]\right\} \qquad (12)$$

with $a > 0$, $b > 0$, $0 < x < +\infty$ and $-\infty < y_i < +\infty$.

If the random variables Y and X are distributed according to the normal-gamma (NG) distribution, $Y, X \sim NG(\mu, V, a, b)$, then the random vector Y has a marginal distribution that can be expressed as a t −multivariate distribution, also known as a multivariate Student:

$$Y \sim t(\mu, aV/b, 2b) \qquad (13)$$

and the random variable X has marginal distribution that is the gamma distribution:

$$X \sim G(a, b) \qquad (14)$$

By substituting in (10) the expression in (11) and comparing with (12) the result obtained with $\frac{n}{2} - 1 = \frac{p}{2} + \frac{(n-p)}{2} - 1$ thus the posterior density function (10) is such that

$$\beta, \tau|y \sim NG(\mu^*, (X'PX)^{-1}, n\hat{\sigma}^2/2, (n - p)/2). \qquad (15)$$

The marginal posterior distribution for the vector β of unknown parameters is then (13) determined by the

multivariate t −distribution

$$\beta|y \sim t\left(\mu^*, \frac{n}{n-p}\hat{\sigma}^2(X'PX)^{-1}, n-p\right). \quad (16)$$

Therefore, the Bayes estimator $\hat{\beta}_B$ of β is obtained by $E(\beta|y)$. Let $\hat{\beta}_B = (X'PX)^{-1}X'Py$. The variance-covariance matrix $D(\beta|y) = \frac{n}{n-p-2}\hat{\sigma}^2(X'PX)^{-1}$.

Then, the marginal distribution for the weight parameter τ obtained from the posterior distribution (15) is the gamma distribution due to (14)

$$\tau|y \sim G(n\hat{\sigma}^2/2, (n-p)/2). \quad (17)$$

The inverse-gamma distribution is therefore the posterior distribution with variance σ^2

$$\sigma^2|y \sim IG(n\hat{\sigma}^2/2, (n-p)/2). \quad (18)$$

We therefore obtain the Bayes estimator of σ^2 defined by $\hat{\sigma}_B^2 = E(\sigma^2|y) = \frac{n}{n-p-2}\hat{\sigma}^2 = \frac{(y-X\hat{\beta}_B)'P(y-X\hat{\beta}_B)}{n-p-2}$. The variance of σ^2 is $V(\sigma^2|y) = \frac{2n^2(\hat{\sigma}^2)^2}{(n-p-2)^2(n-p-4)}$.

Properties 2.1. *The Bayesian estimators $\hat{\beta}_B$ and $\hat{\sigma}_B^2$ have the following properties:*

- $(n-p-2)\hat{\sigma}_B^2/\sigma^2$ *follows a Chi-square with* $n-p$ *degrees of freedom* $\left(\chi^2_{n-p}\right)$,
- $\frac{(n-p)(\beta-\hat{\beta}_B)'X'PX(\beta-\hat{\beta}_B)}{p(n-p-2)\,\hat{\sigma}_B^2}$ *follows a Fisher distribution with* $(p, n-p)$ *degrees of freedom* $(F(p, n-p))$.

2.3. Informative prior

The prior distribution can be chosen to represent the researcher's beliefs prior to observing the results of an experiment, resulting in an approximate subjective Bayesian analysis. After this, the researcher specifies prior beliefs about the model parameters and expresses them in the form of prior probability distribution.

Let's consider the variance factor σ^2, which is now an unknown random variable. To obtain a conjugate prior for the unknown parameters β and σ^2, we introduce in place of σ^2 the parameter of unknown weight τ with $\tau = 1/\sigma^2$. Since $(\det(\tau^{-1}P^{-1}))^{-\frac{1}{2}} = (\det(P))^{\frac{1}{2}}\tau^{\frac{n}{2}}$, the likelihood function is written as follows:

$$f(y|\beta,\tau) = (2\pi)^{-n/2}(\det(P))^{1/2}\tau^{n/2}exp\left[-\frac{\tau}{2}(y-X\beta)'P(y-X\beta)\right] \quad (19)$$

As prior for β and τ, density function (12) of the normal-gamma distribution

$$\beta,\tau \sim NG(\mu, V, a, b) \quad (20)$$

Is chosen (Koch, 2007; Ando, 2010). The joint posterior distribution is then obtained by combining the likelihood function with the joint prior distribution:

$$\begin{aligned}\pi(\beta,\tau|y) &\propto \tau^{p/2+b-1}exp\left\{-\frac{\tau}{2}[2a+(\beta-\mu)'V^{-1}(\beta-\mu)]\right\}\tau^{n/2}exp\left[-\frac{\tau}{2}(y-X\beta)'P(y-X\beta)\right]\\ &= \tau^{n/2+b+p/2-1}exp\left\{-\frac{\tau}{2}[2a+(\beta-\mu)'V^{-1}(\beta-\mu)+(y-X\beta)'P(y-X\beta)]\right\} \quad (21)\end{aligned}$$

The bracketed expression for the exponent can be written as:

$$\begin{aligned}&2a+y'Py+\mu'V^{-1}\mu-2\beta'(X'Py+V^{-1}\mu)+\beta'(X'PX+V^{-1})\beta\\ &= 2a+y'Py+\mu'V^{-1}\mu-(\mu^*)'(X'PX+V^{-1})\mu^*+(\beta-\mu^*)'(X'PX+V^{-1})(\beta-\mu^*)\\ &= 2a+y'Py+\mu'V^{-1}\mu-2(\mu^*)'(X'PX+V^{-1}\mu)+(\mu^*)'(X'PX+V^{-1})(\mu^*)'+(\beta-\mu^*)'(X'PX+V^{-1})(\beta-\mu^*)\\ &= 2a+(\mu-\mu^*)'V^{-1}(\mu-\mu^*)+(y-X\mu^*)'P(y-X\mu^*)+(\beta-\mu^*)'(X'PX+V^{-1})(\beta-\mu^*) \quad (22)\end{aligned}$$

with $\mu^* = (X'PX+V^{-1})^{-1}(X'Py+V^{-1}\mu)$. Substituting (22) into (21) and comparing with (12), we obtain:

$$\beta,\tau|y \sim NG(\mu^*, V^*, a^*, b^*) \quad (23)$$

where

- $V^* = (X'PX+V^{-1})^{-1}$;
- $a^* = [2a+(\mu-\mu^*)'V^{-1}(\mu-\mu^*)+(y-X\mu^*)'P(y-X\mu^*)]/2$;
- $b^* = n/2+b$.

The marginal posterior distribution of β is according to (13) and (24) the multivariate t −distribution

$$\beta|y \sim t(\mu^*, a^*V^*/b^*, 2b^*) \quad (24)$$

Consequently, the Bayes estimator $\hat{\beta}_B$ of β is given by $E(\beta|y) = \mu^*$. Thus, $\hat{\beta}_B = (X'PX+V^{-1})^{-1}(X'Py+V^{-1}\mu)$. Moreover, the marginal posterior distribution for the weight parameter τ is, according to (14) the gamma

distribution. Consequently, the variance $\tau = 1/\sigma^2$ has an inverse-gamma distribution

$$\sigma^2|y \sim IG(a^*, b^*) \quad (25)$$

Therefore, the Bayes estimator $\hat{\sigma}_B^2$ of σ^2 is obtained from $E(\sigma^2|y) = \frac{a^*}{b^*-1}$ and we have $\hat{\sigma}_B^2 = \frac{2a+(\mu-\hat{\beta}_B)'V^{-1}(\mu-\hat{\beta}_B)+(y-X\hat{\beta}_B)'P(y-X\hat{\beta}_B)}{n+2b-2}$, and the variance $V(\sigma^2|y)$ of σ^2 is $V(\sigma^2|y) = \frac{(\hat{\sigma}_B^2)^2}{b^*-2}$. The variance-covariance matrix $D(\beta|y)$ of β is $D(\sigma^2|y) = \hat{\sigma}_B^2(X'PX + V^{-1})^{-1}$.

Properties 2.2. *Bayesian estimators for informative prior distribution have the following properties:*

1. $2(b^* - 1)\frac{\hat{\sigma}_B^2}{\sigma^2}$ *follows a Chi-square distribution with* $2b^*$ *degrees of freedom ;*
2. $\frac{b^*(\beta-\hat{\beta}_B)'(X'PX+V^{-1})(\beta-\hat{\beta}_B)}{p(b^*-1)\hat{\sigma}_B^2}$ *follows a Fisher distribution with* $(p, 2b^*)$ *degrees of freedom.*

III. MCMC METHODS

In order to estimate the unknown quantities in the Bayesian models, Markov Chain Monte Carlo (MCMC) methods are useful tools to sample from those posterior distributions that have no closed form. In this section we briefly introduce two MCMC algorithms commonly used: Gibbs Sampling.

Bayes' theorem (Ando, 2010), we have

$$\pi(\beta, \sigma^2|y, X) \propto f(y|X, \beta, \sigma^2)\pi(\beta|\sigma^2, \mu, V)\pi(\sigma^2|a, b)$$
$$\propto (\sigma^2)^{-\frac{n}{2}-\frac{p}{2}-b-1} exp\left\{-\frac{B}{2\sigma^2}\right\} \quad (26)$$

where B is quantity defined in expression (22). Finally, the joint posterior distribution for σ^2 and β is given by

$$\pi(\beta, \sigma^2|y, X) \propto \left(\frac{1}{\sigma^2}\right)^{b^*+1} exp\left\{-\frac{a^*}{\sigma^2}\right\} (\sigma^2)^{-\frac{p}{2}} exp\left\{-\frac{(\beta-\mu^*)'(V^*)^{-1}(\beta-\mu^*)}{2\sigma^2}\right\}. \quad (27)$$

Therefore, the equation (27) shows that the posterior distribution $\pi(\beta, \sigma^2|y, X)$ is proportional to the multiplication of kernels of the $IG(a^*, b^*)$ and $\mathcal{N}(\mu^*, \sigma^2V^*)$. It turns out that, under this prior distribution, $\pi(\sigma^2|y, X)$ is inverse-gamma distribution and $\pi(\beta|\sigma^2, y, X)$ is an multivariate normal distribution, which means that we can directly sample (σ^2, β) from their posterior distribution by first sampling from $\pi(\sigma^2|y, X)$ and then from $\pi(\beta|\sigma^2, y, X)$. Since we can sample from both of these distributions, samples from the joint posterior distribution $\pi(\sigma^2, \beta|y, X)$ can be made with Gibbs sampling approximation (Monte Carlo Chain Markov Method) (Gelfand & Smith, 1990; Robert & Casella, 1999).

A sample value of (σ^2, β) from $\pi(\sigma^2, \beta|y, X)$ can be made as follow:

- sample $\sigma^2 \sim Inverse - Gamma(a^*, b^*)$;
- sample $\beta \sim Multivariate\ normal(\mu^*, \sigma^2V^*)$.

The pseudo-code for this method is detailed in the algorithm 1 below:

Algorithm 1 Estimation of regression model parameters

Require: Database

Ensure: VP: parameter values

1. $VP \leftarrow \emptyset$
2. Consider the Bayesian linear regression of y_i on $(x_{i,1}, x_{i,2}, \dots, x_{i,p-1})$, for $i = 1, \dots, n$, i.e. $y = X\beta + \varepsilon$
3. **Prior information**

- The conjugate prior distribution for σ^2 is: $\sigma^2 \sim IG(a, b)$
- The conjugate prior distribution for $\beta|\sigma^2$ is: $\beta|\sigma^2 \sim \mathcal{N}(\mu, \sigma^2V)$

4. **Posterior distribution of β and σ^2**

Apply the Gibbs sampling method, assuming conditional densities for each parameter:

- $\pi(\sigma^2|y, X) = IG(a^*, b^*)$
- $\pi(\beta|\sigma^2, y, X) = \mathcal{N}(\mu^*, \sigma^2V^*)$

Initialization

$\theta^{(0)} = \left(\sigma^{2(0)}, \beta^{(0)}\right)'$, the given initial value of $\theta = (\sigma^2, \beta)'$

Iteration

For $i = 1,2, \ldots, N$ do

Sample:

- $\sigma^{2(i)} \sim \pi(\sigma^2 | y, X)$
- $\beta^{(i)} \sim \pi\left(\beta | \sigma^{2(i)}, y, X\right)$

5. $VP \leftarrow \left\{\hat{\sigma}_B^2, \hat{\beta}_B\right\}$: optimal values obtained using a usual cost function
6. Return VP parameters values

IV. APPLICATION

In this section, we apply the Bayesian approach using R codes. To do this, we will use a database called "eucalyptus", which is a study of data provided by height and circumference of 1,429 eucalyptus trees, where the variable to be explained is their height ($euca$) and the explanatory variable is their circumference ($circ$) and its square root ($\sqrt{circ}$). Thus, this model is written as

$$euca = \beta_0 + \beta_1 * circ + \beta_2 * \sqrt{circ} + \varepsilon \qquad (28)$$

In this model, we will estimate the parameters of this regression using two different methods, namely:

1. Parameter estimation using the maximum likelihood method

First, we will use the maximum likelihood method, a parameter estimation method employed in classical statistics, to estimate the unknown parameters of model (28). Here is the result obtained:

Call:

lm(formula = euca ~ circ + I(sqrt(circ)), data = eucalypte)

Coefficients:

(Intercept)	circ	I(sqrt(circ))	sigma^2
-24.3520	-0.4829	9.9869	1.290783

2. Bayesian estimation

Now, we will apply the Bayesian approach. To do this, we will use the algorithm 1. Priors were already assigned as: $\sigma^2 \sim IG(a, b)$ and $\beta | \sigma^2 \sim \mathcal{N}(\mu, \sigma^2 V)$, with:

- $\mu = (0,0,0)'$;
- $V = \begin{pmatrix} 10 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 10 \end{pmatrix}$;
- $a = 1$;
- $b = 0.5$.

We will apply the Gibbs sampling method with the following conditional densities for each parameter:

- $\pi(\sigma^2 | y, X) = IG(a^*, b^*)$,
- $\pi(\beta | \sigma^2, y, X) = \mathcal{N}(\mu^*, \sigma^2 V^*)$; with:
  - $\mu^* = (-15.1712308; -0.2809055; 7.2510905)'$;
  - $V^* = \begin{pmatrix} 3.3584375 & 0.073803701 & -1.0000513 \\ 0.0738037 & 0.001662853 & -0.0222621 \\ -1.0000513 & -0.022262096 & 0.2997903 \end{pmatrix}$;
  - $a^* = 943.4281$;
  - $b^* = 715$.

After performing an iteration with $N = 10{,}000$, we obtained the posterior distribution of these parameters. The following curves illustrate the posterior distribution density of each parameter (see Figure 1)

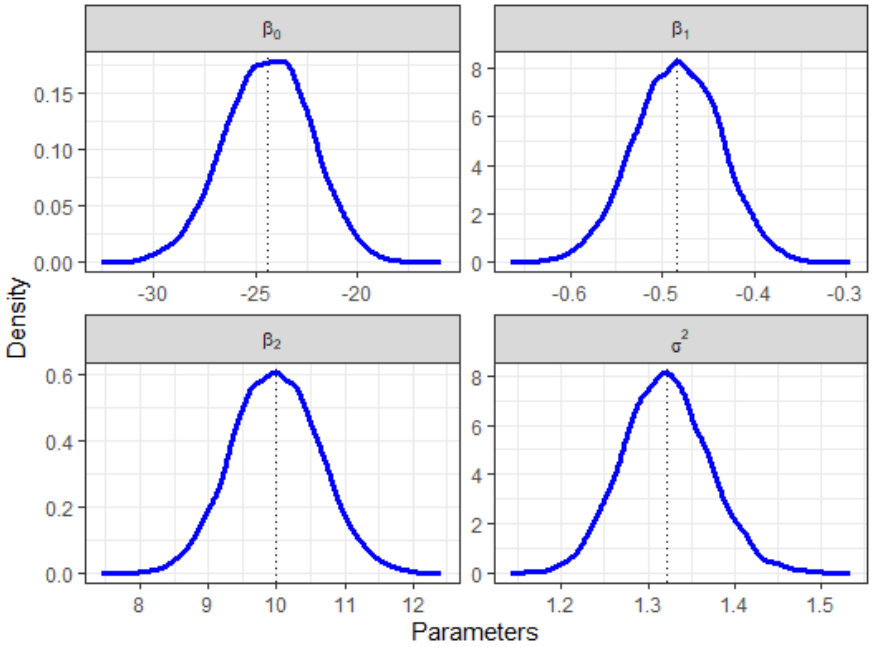

Figure 1: Posterior densities of the parameters

Incidentally, these parameters can be given a value and we obtain:

	Bayes estimator
beta_0	-24.3705404
beta_1	-0.4834148
beta_2	9.9927856
sigma^2	1.3213496

Comments: Thanks to Gibbs sampling (see algorithm 1), we obtain posterior distribution (see Figure 1) and an illustrative value for these parameters. It is very interesting to note that these values are very close to those obtained by the likelihood method and that they are not unique. Given all this, we can say that the estimation method described in algorithm 1 is reliable for estimating the parameters of the Bayesian linear regression model.

V. CONCLUSION

In this study, we examined the mathematical theory of linear regression using a new approach: the Bayesian approach. This approach has several advantages over classical analysis. Through this work, we presented the classical statistics used to calculate estimators of unknown parameters using the maximum likelihood method. In the case of Bayesian linear regression, we studied two prior distributions: the noninformative prior distribution and the informative prior distribution. To study the differences between the classical approach and the Bayesian approach in terms of model formulation and estimation, we used a Monte Carlo experiment and a general linear regression model. This result may be due to the use of vague priors, which reflect the researcher's lack of knowledge about the parameter. In the case, the two methods will give similar (informative), the Bayesian approach is expected to tend to outperform the classical approach.

VI. REFERENCES

1. Ando, T. (2010). Bayesian Model Selection and Statistical Modeling}, Statistics, textbooks, and monographs.

2. Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. Phil Trans Roy Soc London, 5(3):370-418.

3. Gelfand, A., & Smith, A. (1990). Sampling based approaches to calculating marginal densities. Journal of American Statistical Association. 85:398-409 Springer, New York.

4. Jeffreys, H. (1939). Theory of Probability. New York: Oxford University Press.

5. Koch, K. R. (2007). Introduction to Bayesian Statistics, Second Edition, Springer.

6. Koch, K. R. (1999). Parameter Estimation and Hypothesis Testing in Linear Models, 2nd Ed. Springer, Berlin.

7. Laplace, P. S. (1995). A philosophical essay on probabilities, english edn. Dover Publications Inc., New York, translated from the sixth French edition by Frederick William Truscott and Frederick Lincoln Emory, With an introductory note by E. T. Bell.

8. Robert, C. P., & Casella, G. (1999). Monte Carlo statistical methods. Springer.