A Research Project Report On

Bayesian Linear Regression: A Probabilistic Approach of Linear Regression

A Project Report submitted in Partial Fulfilment of

Requirement for the 3^{rd} semester

M.Sc. (Statistics)

With specialization in Industrial Statistics

DEPARTMENT OF STATISTICS

SCHOOL OF MATHEMATICAL SCIENCES

KAVAYITRI BAHINABAI CHAUDHARI NORTH MAHARASHTRA UNIVERSITY, JALGAON

Submitted by:

Mr. Kalpesh Dhanraj Tele (*Seat no.366365)*

Mr. Vishal Dileep Mali (*Seat no. 366345)*

Miss. Sapana Anil Pawar (*Seat no.366361)*

Under the guidance of

Prof. K. K. Kamalja

Dept. of Statistics, KBCNMU

December - 2024

CERTIFICATE

This is to certify that **Mr. Kalpesh Dhanraj Tele**, **Mr. Vishal Dileep Mali** and **Miss. Sapana Anil Pawar** are the students of M.Sc. (Statistics) with specialization in Industrial Statistics, at Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon. They have successfully completed their project work entitled **"Bayesian Linear Regression: A Probabilistic Approach of Linear Regression"** based on research paper reviews as a part of M.Sc. (Statistics) program under my guidance and supervision during the academic year 2024-2025.

Date: 17 December 2024

Prof. K.K. Kamalja
(Project Guide)
Head of Department Statistics,
Kavayitri Bahinabai Chaudhari
North Maharashtra University, Jalgaon

ACKNOWLEDGEMENT

We want to convey our sincere regards to **Prof. K. K. Kamalja**, Head, Department of Statistics, Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon for seeking us the desire permission for this project.

We take this opportunity in expressing our sincere gratitude to our project guide **Prof. K. K. Kamalja** for her valuable guidance, kind suggestion, co-operation and constant encouragement, which enabled us to take every forward step in our project.

It would be unfair to go without acknowledging our beloved friends at the campus of Department of Statistics, Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon for their precious support to us.

Last but not the least, we are thankful to our parents for their moral support and blessings.

Mr. Kalpesh Dhanraj Tele

Mr. Vishal Dileep Mali

Ms. Sapana Anil Pawar

Contents

Chapter 1
Introduction to Bayesian Regression

Introduction

Bayesian Linear Regression is a probabilistic approach to linear regression that uses Bayesian inference to estimate the parameters of the model. In traditional linear regression, the parameters are estimated using Ordinary Least Square or Maximum likelihood estimation, which provides a point estimate of the parameters. In contrast, Bayesian Linear Regression provides a full posterior distribution over the parameters, which can be used to quantify uncertainty and make predictions.

1.1 Key Components of Bayesian Linear Regression

In this section we mention the three basic components of Bayesian Regression, these are follows.

a) Prior Distributions

In Bayesian Linear Regression, we specify prior distributions for the model parameters, can be denoted as $p(\theta)$, which represent our beliefs about the parameters before observing the data. The prior distributions are typically chosen to be conjugate priors, which simplify the computation of the posterior distribution.

b) Likelihood Function

The likelihood function specifies the probability of observing the data given the model parameters, denoted as $p(y|\theta)$. In linear regression, the likelihood function is typically a normal distribution with a mean that is a linear function of the predictors and a variance that is a function of the model parameters.

c) Posterior Distribution

The posterior distribution is the product of the prior distribution and the likelihood function, normalized by the evidence, denoted as $p(\theta|y) \propto \mathrm{p}(\theta).\mathrm{p}(\mathrm{y}|\theta)$. The posterior distribution represents our updated beliefs about the model parameters after observing the data.

1.2 Idea of Bayesian Linear Regression (BLR)

Bayesian Linear Regression (BLR) is a probabilistic approach to linear regression that combines the principles of Bayesian inference with the traditional linear regression model. In BLR, the goal is to model the relationship between a dependent variable y and one or more independent variables X, while accounting for uncertainty in the model parameters. Bayesian Linear Regression is founded on Bayes' theorem, which provides a mathematical framework for updating prior beliefs based on new data.

In a Standard linear regression model of the form,

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where Y is the response variable of $n \times 1$ vector, X is the predictor matrix of size $n \times p$. β is the coefficients vector of size $p \times 1$, and ϵ is an error term (assumed to be normally distributed $\boldsymbol{\epsilon} \sim N(0, \sigma^2)$). The coefficients are found using ordinary least squares to minimize the residual sum of squares between the observations (training data) and the fitted model.

In a Bayesian framework, Linear Regression takes the form

$$\boldsymbol{Y} \sim N(X\boldsymbol{\beta}, \sigma^2)$$

where $\boldsymbol{Y}$ is a random variable with each data point distributed according to a normal distribution. The mean of the normal distribution comes from the data points multiplied by the weights vector $\boldsymbol{\beta}$ with an error variance of σ^2.

1.3 Assumption in Bayesian Linear Regression

Bayesian Linear Regression has specific assumptions that align with the Bayesian approach to statistical modeling. These assumptions govern how the model treats the data, parameters, and Error. Here are the key assumptions:

i) Assumption about distribution of Response

- Linear relationship: The dependent variable y is a linear function of the independent variables X plus noise ε

  $y = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$
- Error distribution: The error term ε is assumed to follow a normal distribution with mean 0 and variance σ^2:

  $\boldsymbol{\varepsilon} \sim N(0, \sigma^2)$

ii) Prior Distributions for parameters

- Parameter priors: Prior beliefs about the regression coefficients β and variance σ^2 are specified.
  - For $\boldsymbol{\beta}$, a common choice is a multivariate normal distribution: $\boldsymbol{\beta} \sim N(\mu_0, \Sigma_0)$
  - For σ^2, a typical choice is an inverse-gamma distribution or a Jeffreys prior.
- Priors should reflect domain knowledge or be chosen to be weakly informative to avoid excessive influence.

iii) Conditional Independence

- Observations are assumed to be conditionally independent given the parameters $\boldsymbol{\beta}$ and σ^2
- This means that knowing the parameters, each observation y_i is independent of the others.

iv) Posterior Inference

- Bayesian Linear Regression uses Bayes' Theorem to update prior beliefs with data evidence:

  $p(\boldsymbol{\beta}, \sigma^2|X, y) \propto p(y|X, \boldsymbol{\beta}, \sigma^2)\, p(\boldsymbol{\beta}, \sigma^2)$
- The posterior is a combination of the likelihood and prior, reflecting both data and prior information.

v) Assumptions about predictors

- The matrix of predictors $\boldsymbol{X}$ is typically assumed to be non-collinear (full rank) to ensure identifiable regression coefficients.
- The relationship between predictors and response is assumed to be reasonably captured by a linear model.

1.4 Limitations of Traditional Linear Regression

The below four points are the Limitation and drawbacks of the Traditional Linear Regression:

- **Assumptions:** Traditional linear regression assumes normality, linearity, and homoscedasticity, which may not always hold in real world data.

- **Lack of Uncertainty Quantification:** Frequentist methods provide point estimates of model parameters, but do not account for uncertainty in these estimates.

- **Overfitting:** Traditional linear regression can suffer from overfit ting, especially when dealing with high-dimensional data.

- **Model Selection:** Frequentist methods often rely on ad-hoc procedures for model selection, which can lead to biased results.

1.5 Advantages of Bayesian Linear Regression

Bayesian Linear Regression offers several advantages over traditional linear regression, particularly in its ability to incorporate uncertainty and prior knowledge. Here below mentioned are the key benefits:

Incorporation of Prior Knowledge

Bayesian methods allow the inclusion of prior beliefs about parameters, which can be helpful when domain knowledge is available. Priors can regularize the model, preventing overfitting in cases with limited data.

- **Probabilistic Interpretation**

Instead of providing point estimates, Bayesian regression gives posterior distributions for the parameters, reflecting uncertainty. Predictions come with credible intervals, making it easier to quantify and communicate the confidence in the results.

- **Automatic Regularization**

Priors (e.g., Gaussian priors on coefficients) naturally shrink estimates toward the prior mean, which is similar to regularization techniques like Ridge Regression or Lasso in frequentist approaches. This is particularly useful in high-dimensional settings with many predictors.

- **Flexibility with Complex Models**

Bayesian frameworks can easily handle hierarchical models, time-series structures, and other extensions, making them highly adaptable to complex real-world problems.

- **Handling Multicollinearity**

Bayesian regression can manage multicollinearity by incorporating priors that stabilize coefficient estimates when predictors are correlated.

- **Robustness to Overfitting**

The explicit modeling of uncertainty helps avoid overfitting, especially with small datasets, since the posterior combines prior information and data evidence.

- **Prediction Uncertainty**

Bayesian regression provides a full predictive distribution for new observations, not just point estimates, allowing for better risk assessment in predictions.

- **Model Comparison**

Bayesian methods facilitate model comparison using metrics like Bayes Factors or marginal likelihood, enabling principled decision-making when comparing different models.

- **Improved Interpretability**

The probabilistic nature allows a deeper understanding of parameter uncertainty, making Bayesian models more interpretable in terms of confidence about parameter values.

1.6 Bayesian Regression Model Accuracy Measures

a) Mean Squared Error (MSE): A measure of the average squared difference between predicted and observed values.

Mean Squared Error (MSE) is a metric used to evaluate the performance of a regression model. It measures the average squared difference between the predicted values ($\hat{y}$) and the observed values (y). A smaller MSE value indicates that the model's predictions are closer to the actual values, while a larger MSE suggests poor performance.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Here:

- n: The total number of observations.
- y_i: The actual observed value for the i-th data point
- $\hat{y}_i$: The predicted value for the i-th data point.
- $(y_i - \hat{y}_i)^2$: The squared difference between the observed and predicted values.

b) Mean Absolute Error (MAE): A Measure the average absolute difference between predicted and observed values.

Mean Absolute Error (MAE) measures the average of the absolute differences between the predicted values ($\hat{y}_i$) and the observed values (y_i). It is a common metric for regression models and provides a straightforward interpretation of model performance. A lower MAE value indicates better accuracy, as it shows that predictions are closer to the actual values on average.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

Here:

- y_i: The actual observed value for the i-th data point.
- $\hat{y}_i$: The predicted value for the i-th data point.
- $|y_i - \hat{y}_i|$: The absolute difference between observed and predicted values.

c) Bayesian R-squared: A measure of the proportion of variance in the dependent variable that is explained by the independent variables, which is calculated using the posterior distribution of the model parameters.

Bayesian R^2 is a measure of the proportion of the variance in the dependent variable that is explained by the independent variables. Unlike traditional R^2 , Bayesian R^2 accounts for the uncertainty in the model parameters using their posterior distribution. It provides a probabilistic interpretation of model fit rather than a single point estimate.

$$R^2 = \frac{var(\hat{y})}{var(\hat{y}) + var(\epsilon)}$$

where:

- $var(\hat{y})$: Variance of the predicted values (fitted values).
- $var(\epsilon)$: Variance of the residuals (unexplained variance).

In Bayesian regression, these variances are computed using the posterior predictive distribution.

- $var(\hat{y})$: Posterior Variance of the predicted values, representing the explained variance.

- $var(\epsilon)$: Posterior Variance of the residuals, representing the unexplained variance.

The range of Bayesian R^2 is $(0 \leq R^2 \leq 1)$ which is same as the classical R^2. The value of the R^2 is around 0 then the model none explained the variability of the response variable and the value of R^2 is around 1 then the model perfectly fitted on the data, it explains the all variance in the response variable.

d) Deviance Information Criterion (DIC): It is a measure of model complexity and fit, which can be used to compare models. Lower DIC values indicate a better fit.

The DIC is a statistical measure used to evaluate and compare Bayesian models. It balances model fit and complexity, providing a metric to choose between competing models. A lower DIC value suggests a better-fitting model with an appropriate balance between complexity and goodness of fit.

$$DIC = \overline{D} + p_D$$

where:

- $\overline{D}$: The average deviance over the posterior distribution of the parameters, computed as:

  $$\overline{D} = E_{\theta \sim p(\theta|data)}[D(\theta)]$$

Here, $D(\theta)$ is the deviance for a parameter set θ, defined as:

$$D(\theta) = -2 \log p(data \mid \theta) + C$$

- $p(data \mid \theta)$: The likelihood of the data given the parameters.
- C: A constant that cancels out when comparing models.
- p_D: The effective number of parameters (a measure of model complexity), calculated as:

  $$p_D = \overline{D} - D(\hat{\theta})$$

- $\hat{\theta}$: The posterior mean of the parameters.
- $D(\hat{\theta})$: The deviance evaluated at $\hat{\theta}$.

e) Watanabe-Akaike Information Criterion (WAIC): It is a measure of model complexity and fit, which is similar to DIC but is more robust to model misspecification. The WAIC is a model selection criterion that is used to evaluate the trade-off between model fit and complexity. It is particularly used in Bayesian models and is considered a more robust alternative to other criteria like the DIC. WAIC accounts for model uncertainty by considering the posterior distribution of the model parameters, which makes it a better choice in cases where the model may be mis specified.

WAIC is often preferred over Akaike Information Criterion (AIC) or DIC because it uses the entire posterior distribution, rather than relying on a single point estimate of the parameters. WAIC is calculated using the log-likelihood of the model evaluated at the posterior samples of the model parameters. The formula is as follows:

$$WAIC = -2\left(\sum_{i=1}^{n} log\left(\frac{1}{S}\sum_{s=1}^{S} p(y_i|\theta_s)\right) - \frac{1}{S}\sum_{s=1}^{S} \log\left(p(y_i|\theta_s)\right)\right)$$

Here:

- y_i: The observed data point for the i_{th} observation.
- $p(y_i|\theta_s)$: The likelihood of the data point y_i given the parameters θ_s for the s_{th} posterior sample.
- S: The total number of posterior samples.
- n: The number of observations in the dataset.

f) Cross-Validation (CV): It is a technique for evaluating the predictive performance of a model by splitting the data into training and testing sets.

Cross-validation (CV) is a technique used to assess the performance of a machine learning model by splitting the dataset into multiple subsets or folds. The model is trained on some of these folds (training set) and tested on the remaining fold(s) (testing set). This process is repeated multiple times with different splits, providing a more reliable estimate of the model's generalization performance**.** The general formula for k-fold cross-validation is:

$$CV = \frac{1}{k}\sum_{k=1}^{k} f_k$$

Here:

- f_k is the evaluation metric (such as MSE, accuracy, etc.) obtained from testing the model on fold k.
- The sum of the performance measures from each fold is averaged to provide an overall estimate of model performance.

f) Leave-One-Out (LOO) Cross-Validation: It is a special case of cross validation where each observation is used as a test set, and the remaining observations are used as a training set. Leave-One-Out Cross-Validation (LOO-CV) is a special case of cross-validation where each individual observation in the dataset is used as a test set exactly once, while the remaining observations serve as the training set. This process is repeated for each data point, and the model's performance is evaluated by averaging the results of all these iterations.

LOO-CV is particularly useful for small datasets, as it maximizes the use of available data for training. However, it can be computationally expensive for large datasets since the model needs to be trained as many times as there are data points.The general formula for calculating the model's error using Leave-One-Out Cross-Validation is:

$$LOO_{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Chapter 2
Fitting Bayesian Linear Regression Model

In this Section we present the Bayesian linear regression models. Also, we introduce the Bayesian perspective, prior distributions and estimation methods.

2.1 Basics concepts of Bayesian inference

Under the Bayesian perspective, we aim, both through data and subjective information, to draw conclusions about a certain unknown quantity of interest by using probabilistic models. Under the classical point of view, the inference is also made utilizing probabilistic models, but only the information from the data is considered and any other extra information is not incorporated into the decision process. The inclusion of the subjective information in the inference process is the main point in Bayesian analysis. That is made by inserting a prior distribution that describes all the available knowledge that one can have about the quantity of interest, and its choice may influence the final results depending on how much data is available. That is, the more data, the less the impact of the prior information on the final conclusions.

There are many types of prior distributions, such as non-informative, conjugate, improper etc. The non informative ones are based on the sampling distribution and their idea is to have a default prior distribution when there is no information about the problem at hand. For instance, the Jeffreys prior is non-informative and it is proportional to the Fisher Information, which is the expected value of the second derivative of the log-likelihood function with respect to the parameter of interest. Although the Jeffreys prior is called non-informative. Thus, evidencing that the data helps to estimate the quantity of interest. The conjugate priors are a class of distributions that present the same parametric form of the likelihood function and their choice is frequently related to mathematical and computational convenience. As a consequence of conjugacy, the posterior distribution may be obtained analytically and posterior samples are generated straightforwardly. On the other hand, improper priors are distributions that, in their parametric space, do not integrate to 1. For instance, in some cases Jeffreys priors are improper, but the posterior distribution is proper.

Consider $\theta = \left(\theta_1, \dots, \theta_p\right)^T$ an unknown vector of parameters that we are interested in estimating and $= (y_1, y_2, \dots, y_n)^T$ an n×1 column vector assumed to be a realization of the random variable whose distribution is $p(y_i|\theta)$. The likelihood function of the y_i is given by

$$\mathcal{L}(\theta|y) = \prod_{i=1}^{n} p(y_i|\theta)$$

All the information from the observations y_i about θ is included. The difficulty in estimating θ becomes an optimization problem of maximizing the likelihood function (or its logarithm). In contrasts, under the Bayesian methodology the estimate of θ is given by the joint posterior distribution, which is defined by the Bayes' theorem,

$$p(\theta|y) = \frac{\mathcal{L}(\theta|y)p(\theta)}{\int \mathcal{L}(\theta|y)p(\theta)d\theta},$$

The above equation can be written as

$$p(\theta|y) \propto \mathcal{L}(\theta|y)\, p(\theta),$$

Since $\int \mathcal{L}(\theta|y)p(\theta)d\theta$ is the marginal distribution of y and does not depend on θ. The posterior distribution $p(\theta|y)$ provides all the information that one can have about θ. For instance, it is possible to evaluate $p(\theta|y)$ and instance, it is possible to evaluate $p(\theta|y)$ and y its mean, median, variance and some other quantities such as quantiles in order to have point and interval estimates. Besides, the posterior distribution frequently has no closed form, thus depending on computational methods to be obtained.

When the posterior distribution is available, one can be interested about the predictive posterior distribution, which is utilized to predict unobserved values of the response outcome, $\tilde{y}$, and the marginal distribution of y.

$$p(\tilde{y}|y) = \int p(\tilde{y}|\theta, y)\, p(\theta|y)\, d\theta, \quad \tilde{y} \sim p(\tilde{y}|\theta, y),$$
$$p(y) = \int \mathcal{L}(\theta|y)\, p(\theta) d\theta.$$

Here, we discuss the main common priors, namely Conjugate, Non-Conjugate, and Non-Informative priors, for fitting the BLR model parameters. The procedure for drawing the posterior distribution of the response variable is also outlined. In the Conjugate prior for BLRM is Normal-Inverse Gamma, Normal-Scaled Inverse χ^2, and Normal-Cauchy priors for the error variance. Additionally, we mention one non-informative prior, the flat Jeffreys prior, for the BLR model parameters. For the Non-Conjugate prior for β, when the error variance is known is the Laplace, Student's t, and flat priors are also mention.

2.2 Conjugate priors for Bayesian Linear Regression Model

Here we consider a Bayesian linear model is the form

$$\boldsymbol{y} = X\boldsymbol{\beta} + \epsilon, \ where \ \epsilon \sim N_n(0, I_n)$$

where $\sigma^2 > 0$, I_n an identity matrix, $\beta = (\beta_0, \beta_1, \dots\dots, \beta_{n-1})^T$ a $p \times 1$ vector, X an $n \times p$ design matrix and we assume that $\epsilon_i's$ are independent. The likelihood function is

$$f(y|X, \beta, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}(y - X\beta)^T(y - X\beta)\right\}$$

2.2.1 Normal – Inverse Gamma as a Conjugate Prior

This derivation of Normal – Inverse Gamma prior is mention as outlined by the authors of Baruna Wundervald (2019).

Under the Bayesian point of view, the inference process involves data and prior information. Thus, we assume a $N_p(\mu, \Sigma)$, which is a conjugate prior distribution for $\beta|\sigma^2$ as follow.

$$f(\beta|\sigma^2, \mu, \Sigma) = (2\pi\sigma^2)^{-\frac{p}{2}} |\Sigma|^{-\frac{p}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\beta - \mu)^T \Sigma^{-1} (\beta - \mu)\right\}$$

For σ^2, also set a conjugate prior distribution given by an Inverse Gamma denoted by $IG(a, b)$ in the form of

$$f(\sigma^2|a, b) = \frac{b^a}{\Gamma(a)}(\sigma^2)^{a-1} \exp\left\{-\frac{b}{\sigma^2}\right\},$$

Where $a > 0$ and $b > 0$. Since we have the likelihood function and the proper priors, we can then find the posterior distribution to make inference on the parameters β and σ^2. Using the Bayes' theorem, we have

$$f(\beta, \sigma^2|y, X) = \frac{f(y|X, \beta, \sigma^2)\, f(\beta|\sigma^2, \mu, \Sigma)\, f(\sigma^2|a, b)}{f(y)},$$

$$f(\beta, \sigma^2|y, X) \propto f(y|X, \beta, \sigma^2)\, f(\beta|\sigma^2, \mu, \Sigma)\, f(\sigma^2|a, b),$$

$$f(\beta, \sigma^2|y, X) \propto (\sigma^2)^{\frac{n}{2} - \frac{p}{2} + a - 1} \exp\left\{ -\frac{A}{2\sigma^2} \right\},$$

Where,

$$A = (y - X\beta)^T(y - X\beta) + (\beta - \mu)^T\Sigma^{-1}(\beta - \mu) + 2b$$

$$= y^Ty - yX\beta - \beta^TX^Ty + \beta^TX^TX\beta + \beta^T\Sigma^{-1}\beta - \beta^T\Sigma^{-1}\mu - \mu^T\Sigma^{-1}\beta + \mu^T\Sigma^{-1}\mu + 2b$$

$$A = \beta^T(X^TX + \Sigma^{-1})\beta - \beta^T(X^Ty + \Sigma^{-1}\mu) + (\mu^T\Sigma^{-1}\mu + 2b + y^Ty) - (y^TX + \mu^T\Sigma^{-1})\beta$$

For convenience, let $\Lambda = (X^TX + \Sigma^{-1})^{-1}$ a $p\ x\ p$ matrix and $m = (X^TX + \Sigma^T)^{-1}(X^Ty + \Sigma^{-1}\mu)$ a $p\ x\ 1$ vector. Hence,

$$f(\beta, \sigma^2|y, X) \propto f(y|X, \beta, \sigma^2)\, f(\beta|\sigma^2, \mu, \Sigma)\, f(\sigma^2|a, b),$$

$$f(\beta, \sigma^2|y, X) \propto (\sigma^2)^{-\frac{p}{2}} \exp\left\{ -\frac{(\beta - m)^T\, \Lambda^{-1}\, (\beta - m)}{2\sigma^2} \right\}$$

$$x\ (\sigma^2)^{-\frac{n}{2} + a - 1} \exp\left\{ -\frac{\mu^T\Sigma^{-1}\mu - m^T\Lambda^{-1}m + 2b + y^Ty}{2\sigma^2} \right\}$$

Therefore, the above equation shows that the posterior distribution $f(\beta, \sigma^2|y, X)$ is proportional to the multiplication of kernels of the $N_n(m, \sigma^2\Lambda)$ and $IG\left(a^* = -\frac{n}{2} + a, b^* = b + \frac{\mu^T\Sigma^{-1}\mu - m^T\Lambda^{-1}m + y^Ty}{2}\right)$.

2.2.2 Normal - Scaled Inverse χ^2 distribution as a Conjugate Prior

These derivation of Normal Scaled Inverse χ^2 is derived by us. The draw the posterior distribution, use Normal – Scaled Inverse χ^2 conjugate prior. We assume a $N_p(\mu, \Sigma)$, which is a conjugate prior distribution for $\beta|\sigma^2$ as follows

$$f(\beta|\sigma^2, \mu, \Sigma) = (2\pi\sigma^2)^{-\frac{p}{2}} |\Sigma|^{-\frac{p}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\beta-\mu)^T \Sigma^{-1} (\beta-\mu)\right\}$$

The scaled inverse χ^2 distribution for σ^2 is parameterized by a shape parameter v degree of freedom and a scale parameter τ^2. The prior for σ^2 is given by

$$f(\sigma^2|v, \tau^2) = \frac{\left(\frac{v\tau^2}{2}\right)^{-\left(\frac{v}{2}\right)-1}}{\Gamma\left(\frac{v}{2}\right)} (\sigma^2)^{-(v/2)-1} \exp\left\{-\frac{v\tau^2}{2\sigma^2}\right\},$$

The posterior distribution of $\beta, \sigma^2|y, X$ is

$$f(\beta, \sigma^2|y, X) = \frac{f(y|X, \beta, \sigma^2)\, f(\beta|\sigma^2, \mu, \Sigma)\, f(\sigma^2|v, \tau^2)}{f(y)},$$

$$f(\beta, \sigma^2|y, X) \propto f(y|X, \beta, \sigma^2)\, f(\beta|\sigma^2, \mu, \Sigma)\, f(\sigma^2|v, \tau^2)$$

$$f(\beta, \sigma^2|y, X) \propto (\sigma^2)^{-\frac{n}{2}-\frac{p}{2}-\frac{v}{2}-1} \exp\left\{-\frac{1}{2\sigma^2}((y-X\beta)^T(y-X\beta) + (\beta-\mu)^T \Sigma^{-1} (\beta - \mu) + v\tau^2)\right\},$$

Define,

$$Q = ((y-X\beta)^T(y-X\beta) + (\beta-\mu)^T \Sigma^{-1} (\beta-\mu) + v\tau^2)$$

Substituting Q, then the posterior distribution becomes

$$f(\beta, \sigma^2|y, X) \propto (\sigma^2)^{-\frac{(n+p+v)}{2}} \exp\left\{-\frac{Q}{2\sigma^2}\right\},$$

Therefore, the above equation shows that the posterior distribution $f(\beta, \sigma^2|y, X)$ is proportional to the multiplication of kernels of the Normal – Scaled Inverse χ^2 distribution.

2.2.3 Normal – Cauchy Prior as a Conjugate Prior

This derivation of Normal Cauchy prior is derived by us. The draw the posterior distribution, use Normal – Cauchy conjugate prior. We assume a $N_p(\mu, \Sigma)$, which is a conjugate prior distribution for $\beta|\sigma^2$ as follows

$$f(\beta|\sigma^2, \mu, \Sigma) = (2\pi\sigma^2)^{-\frac{p}{2}} |\Sigma|^{-\frac{p}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\beta-\mu)^T \Sigma^{-1} (\beta-\mu)\right\}$$

The Cauchy distribution for σ^2 is parameterized by a shape parameter v degree of freedom and a scale parameter τ^2. The prior for σ^2 is given by,

The distribution function for a Cauchy prior on σ is

$$f(\sigma^2|\gamma) = \frac{1}{\pi\gamma\left(1+\left(\frac{\sigma}{\gamma}\right)^2\right)},$$

Now, for prior for σ^2

$$f(\sigma^2|\gamma) = \frac{1}{\pi\gamma}\frac{1}{\sqrt{\sigma^2}\left(1+\left(\frac{\sqrt{\sigma^2}}{\gamma}\right)^2\right)},$$

$$f(\sigma^2|\gamma) \propto (\sigma^2)^{-\frac{1}{2}} \frac{1}{1+\frac{\sigma^2}{\gamma^2}}$$

The joint posterior distribution is

$$f(\beta, \sigma^2|y, X) \propto f(y|X, \beta, \sigma^2)\, f(\beta|\sigma^2, \mu, \Sigma)\, f(\sigma^2|\gamma)$$

$$f(\beta, \sigma^2|y, X) \propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}(y-X\beta)^T (y-X\beta)\right\}$$

$$x\ (\sigma^2)^{-p\backslash 2} \exp\left\{-\frac{1}{2\sigma^2}(\beta-\mu)^T \Sigma^{-1} (\beta-\mu)\right\} x\ (\sigma^2)^{-1\backslash 2} \frac{1}{1+\frac{\sigma^2}{\gamma^2}}$$

$$f(\beta, \sigma^2|y, X) \propto (\sigma^2)^{-\left(\frac{n+p+1}{2}\right)} \exp\left\{-\frac{1}{2\sigma^2}((\beta-\mu^*)^T \Lambda^{-1} (\beta-\mu^*) + S)\right\} \cdot \frac{1}{1+\frac{\sigma^2}{\gamma^2}}$$

Here, $\Lambda = (X^T X + \Sigma^{-1})^{-1}$, $\mu^* = \Lambda(X^T y + \Sigma^{-1}\mu)$ and $S = (y - X\mu^*)^T (y - X\mu^*)$

This distribution not form a standard form, so we used numeric methods (such as MCMC, Gibbs sampling, etc.) are generally used to estimate β and σ^2 form the posterior.

2.3 Non - Informative Priors for Bayesian Linear Regression Model

Now discuss the non – information prior for regression parameter in fitting BLR model. Here, we demonstrate the one expression for the posterior distribution with flat prior for regression coefficients and Jefferey's prior for error variance.

Flat prior for $\boldsymbol{\beta}$:

$$f(\beta) \propto 1$$

This is an improper prior, meaning it doesn't integrate to 1 over the entire parameter space Θ. It assign equal probability to all possible values of β, implying no prior knowledge or preference for any particular value.

Jeffreys' prior for $\boldsymbol{\sigma^2}$:

$$p(\sigma^2) \propto \frac{1}{\sigma^2}$$

The Jeffreys prior is a non – informative prior specifically designed to be invariant under reparameterizations. The Jeffreys prior represents scale invariance, meaning the choice of measurement units for the variance doesn't affect the prior.

The joint posterior distribution for β and σ^2 is proportional to the product of the likelihood and the flat priors

$$f(\beta, \sigma^2 | y, X) \propto f(y | X, \beta, \sigma^2).f(\beta).f(\sigma^2)$$

Substituting the likelihood and priors

$$f(\beta, \sigma^2 | y, X) \propto (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}(y - X\beta)^T (y - X\beta)\right\}.\frac{1}{\sigma^2}$$

$$f(\beta, \sigma^2 | y, X) \propto (\sigma^2)^{-\left(\frac{n}{2}+1\right)} \exp\left\{-\frac{1}{2\sigma^2}(y - X\beta)^T (y - X\beta)\right\}$$

This posterior distribution reflects a purely data-driven inference approach, with the non-informative priors allowing the likelihood to dominate the inference process.

Marginal Posterior for β

The marginal posterior for β is a multivariate Student's t- distribution:

$$\beta \mid y \sim t_{n-p}(b, V)$$

- $b = (X^T X)^{-1} X^T y$ (Posterior mean)
- $V = s^2 (X^T X)^{-1}$ (Posterior variance, scaled by residual variance
- $s^2 = \frac{(y - X\beta)^T (y - X\beta)}{n-p}$ (Residual variance).

Posterior for σ^2 is an Inverse – gamma distribution

The marginal posterior for σ^2 is an Inverse Gamma distribution:

$$\sigma^2 | y \sim Inverse - Gamma \left(\frac{n-p}{2}, \frac{(y - Xb)^T (y - Xb)}{2} \right)$$

Where, $p(\sigma^2 | y) \sim IG(\alpha, \beta)$ with shape parameter α, $\alpha = \frac{n-p}{2}$ and scale parameter β,

$$\beta = \frac{(y - Xb)^T (y - Xb)}{2}$$

2.4 Non-Conjugate Priors with known variance (σ^2)

A) Laplace prior for Regression Coefficient β

In Order to illustrate a Bayesian linear model with non – conjugate prior, consider $y \sim N_n(X\beta, \sigma^2 I)$ and assume that σ^2 is known. Also, consider a Laplace distribution as prior for each β_i such as

$$f(\beta_i | \mu_i, v_i) = (2v_i)^{-\frac{p}{2}} \exp \left\{ - \sum_{i=1}^{p} \frac{| \beta_i - \mu_i |}{v_i} \right\},$$

where, $\beta_i \in \mathbb{R}$, $\mu_i \in \mathbb{R}$ and $v_i > 0$. In this case, the posterior distribution is given by

$$f(\beta | y, X, \sigma^2) \propto f(y | X, \beta) \prod_{i=1}^{p} f(\beta_i | \mu_i, v_i),$$

$$f(\beta|y,X,\sigma^2) \propto \exp\left\{ -\frac{1}{2\sigma^2}(y-X\beta)^T\,(y-X\beta) - \sum_{i=1}^{p}\frac{|\,\beta_i - \mu_i\,|}{v_i}\right\}.$$

The equation above has no closed form of a known pdf or pmf and MCMC methods can be utilized an order to obtain samples from the posterior distributions of β_i, Although the Gibbs Sampling is an MCMC method, in this case it cannot be used since it is not possible to sample directly from $f(\beta|y,X,\sigma^2)$. In contrast, Metropolis – Hasting can be applied.

B) Student's t prior for Regression Coefficient $\boldsymbol{\beta}$

Prior for β: Assume a multivariate Student's t-distribution prior for β:

$$f(\beta|v,m,\tau^2) \propto \left(1+\frac{1}{v\tau^2}(\beta-\mu)^T(\beta-\mu)\right)^{\frac{v+d}{2}}$$

where:

- v are the degrees of freedom, controlling the heaviness of the tails.
- μ is the location parameter (mean vector of β).
- τ^2 is the scale parameter, which can be interpreted as the variance of β.

The joint posterior distribution for β is proportional to the product of the likelihood and the student's t prior. Using Bayes' theorem, we have

$$f(\beta|y,X,\sigma^2) \propto f(y|X,\beta,\sigma^2)\,.f(\beta|v,\mu,\tau^2)$$

Substitute the likelihood and the prior

$$f(\beta|y,X,\sigma^2) \propto \exp\left\{-\frac{1}{2\sigma^2}\,(y-X\beta)^T\,(y-X\beta)\right\}.\left(1+\frac{1}{v\tau^2}\,(\beta-\mu)^T(\beta-\mu)\right)^{-\frac{v+d}{2}}$$

The posterior distribution of β with a known σ^2 and Student's t-distribution prior is:

$$f(\beta|y,X,\sigma^2) \propto \exp\left\{-\frac{1}{2\sigma^2}\,(y-X\beta)^T\,(y-X\beta)\right\}.\left(1+\frac{1}{v\tau^2}\,(\beta-\mu)^T(\beta-\mu)\right)^{-\frac{v+d}{2}}$$

This posterior is generally not analytically tractable due to the non-conjugate Student's t prior. Numerical techniques (like MCMC) are needed to approximate the posterior distribution.

C) Flat prior for Regression Coefficient ($\boldsymbol{\beta}$), when $\boldsymbol{\sigma^2}$ is known

To represent a lack of prior information you can use a uniform prior over β

$$f(\beta) \propto 1$$

This prior assumes all values of β are equally likely a priori, meaning it does not favor any particular value.

Posterior Distribution

$$f(\beta|y, X, \sigma^2) \propto f(y|X, \beta, \sigma^2).f(\beta)$$

Substitute the likelihood and the uniform prior

$$f(\beta|y, X, \sigma^2) \propto \exp\left\{-\frac{1}{2\sigma^2}\ (y - X\beta)^T\ (y - X\beta)\right\}.1$$

Since the prior $f(\beta)$ is uniform, it doesn't alter the posterior. Therefore, the posterior is directly proportional to the likelihood:

$$f(\beta|y, X, \sigma^2) \propto \exp\left\{-\frac{1}{2\sigma^2}\ (y - X\beta)^T\ (y - X\beta)\right\}$$

This posterior distribution is a **Gaussian distribution** with mean and covariance derived from the likelihood. Specifically, this implies a **Maximum Likelihood Estimation (MLE)** approach for estimating β, where the posterior mean and variance of β will match those obtained by ordinary least squares (OLS) regression. Thus, the posterior distribution can be characterized as:

$$\beta|(y, X) \sim N((X^TX)^{-1}X^Ty, \sigma^2(X^TX)^{-1})$$

This approach allows the data to completely drive the inference for β, as the prior contributes no additional information.

Chapter 3

R-Package for Fitting Bayesian Linear Regression Model

We discuss Bayesian regression models with Stan (`brms`) and Bayesian applied regression modelling via Stan (`rstanarm`) for fitting BLR model in R software. We demonstrate the basic fitting procedure and syntax of these two libraries.

3.1 `brms` Package

The **`brms`** package was developed by Paul-Christian Burkner in 2017, a statistician and software developer. He released the package to simplify Bayesian modeling in R by providing an interface to Stan, a powerful platform for statistical modeling and high-performance computation

Title: Bayesian Regression Models using Stan– R package

Description: Fit Bayesian generalized (non-)linear multivariate models using Stan for full Bayesian inference. A wide range of distributions and link functions are supported, allowing users to fit– among others– linear, robust linear, count data, survival, response times, ordinal, zero-inflated, hurdle, and even self-defined mixture models, all in a multilevel context. Further modelling options include both theory-driven and data-driven non-linear terms, auto-correlation structures, censoring and truncation, meta-analytic standard errors, and quite a few more. In addition, all parameters of the response distribution can be predicted in order to perform distributional regression. Prior specifications are flexible and explicitly encourage users to apply prior distributions that actually reflect their prior knowledge. Models can easily be evaluated and compared using several methods assessing posterior or prior predictions. In addition, model fit can easily be assessed and compared with posterior predictive checks and leave-one-out cross-validation (LOOCV).

Function: `brm`

Syntax:

```
brm(formula, data, family = gaussian(), prior  = NULL,sample= "no", …)
```

Arguments

- **Formula**: An object of class formula, mapping between the response variable and exploratory variables.
- **data:** An object of class data. Frame (or one that can be coerced to that class) containing data for all variables used in the model.

- **family:** A description of the response distribution (likelihood) to be used in the model. This can be a family function, a call to a family function, or a character string naming the family. If not specified, default links are used. By default, a linear Gaussian model is applied.

- **Prior:** One or more brms prior objects created by set prior or related functions and combined using the c method or the + operator.

- **warmup:** A positive integer specifying the number of warmup (aka burn in) iterations. This also specifies the number of iterations used for step size adaptation, so warmup draws should not be used for inference. The number of warmup iterations should not be larger than iter, and the default is iter/2.

- **cores:** The number of cores to use when executing the chains in parallel, which defaults to 1.

- **init :** Initial values for the sampler. If NULL (the default) or "random", Stan will randomly generate initial values for parameters in a reasonable range. If 0, all parameters are initialized to zero on the unconstrained space. This option is sometimes useful for certain families where default random initial values may cause draws to be essentially constant.

- **chains :** Number of Markov chains (defaults to 4). Each chain represents an independent exploration of the posterior distribution, starting from different initial values.

R Code: The sample code for Fitting Bayesian Linear Regression Model

```
# Bayesian Linear Regression Model fitting syntax and arguments
Fit <- brm(formula = y ~ .,data = data, family = gaussian(),
prior = c(prior(normal(5,0.7),class = 'b'), prior(student_t(df
= 3,location = 0,scale = 5),class = 'sigma'))), chain = 4,iter
= 2000,warmup = 1000,cores = 2,seed = 123)
#generate a summary of the results
summary(Fit)
```

Function: bayes R2

Compute a Bayesian version of R-squared for regression models.

Syntax:

```
bayes_R2(object, resp = NULL, summary = TRUE, probs = c(0.025,
0.975), …)
```

Arguments: object is the fitted model of class brmsfit

Function: `pp_check`

Comparing the observed data to data generated from the model's posterior predictive distribution.

Syntax:

```
pp_check(object, ...)
```

Function : `fixef`

Extract the population-level ('fixed') effects from a brmsfit object.

```
fixef(
  object, summary = TRUE,
  probs = c(0.025, 0.975),
  ...)
```

Function: `mcmc_area`

Visualize the posterior distribution of parameters in Bayesian Analysis. It provides an area plot for each parameter's marginal posterior distribution, highlighting the density and credible intervals.

```
mcmc_plot(
  object,
  pars = NA,
  type = "intervals",
  variable = NULL,
  regex = FALSE,
  fixed = FALSE,
  ...)
```

3.2 rstanarm Package

The `rstanarm` package was developed by Jonah Gabry and his team in 2024, a statistician and stan developer. He released the package to simplify Bayesian modeling in R by providing an interface to Stan, a powerful platform for statistical modeling and high-performance computation

Title: Bayesian Applied Regression Modelling via Stan– R package

Description: The rstanarm package allows users to fit Bayesian applied regression models using the Stan backend for full Bayesian inference. It offers a wide range of pre-built model classes that can be used for generalized linear models (GLMs), linear models, and hierarchical models. With rstanarm, users can fit models such as linear regression, logistic regression, survival models, count models, and more. The package offers a straightforward syntax similar to glm and lm functions in base R. It emphasizes ease of use for applied researchers and enables users to add their own priors, which can be informed by domain expertise or data-driven approaches. In addition, the package provides various tools for posterior analysis, including posterior predictive checks, model diagnostics, and performance comparison.

Function: stan_glm

Syntax:

```
Model = stan_glm(formula, data, family = gaussian(),
                 prior = NULL,priorintercept = NULL, ...)
```

Arguments

- **formula:** An object of class formula, specifying the response and predictor variables.

- **data:** A data. Frame containing the data for the model.
- **family**: A description of the error distribution and link function to be used in the model (e.g., gaussian (), binomial (), poisson ()). If not specified, the default is the normal (Gaussian) family.
- **prior:** A prior object that specifies the priors on the coefficients (e.g., normal, student-t). The prior can be customized or left as default weakly informative priors. • prior intercept: Similar to prior, but applies to the model intercept.
- ...: Additional arguments for advanced control of model fitting options.

R Code for example:

```
# Logistic regression using stan_glm
fit = stan_glm(am ~ wt + cyl, data = mtcars,
family = binomial(), prior = normal(0, 2.5),
prior_intercept = normal(0, 10))

# Generate a summary of the results
summary(fit)
```

Function: posterior predict.stanreg

Description: Compute posterior predictive distributions for the fitted model.

Syntax:

```
posterior_predict(object, newdata = NULL,
draws = NULL, ...)
```

Arguments

- **object**: A fitted model of class stanreg (such as models created by stan_glm)
- **newdata:** A data frame of new predictor values for which posterior predictions are to be generated.
- **draws**: The number of posterior draws to sample. If NULL, all draws are used.
- ...: Additional arguments for controlling the posterior sampling process

Chapter 4
Fitting BLRM for Simulated and Real-life data

4.1 Simulation Study

In this chapter, we applied the fitting of BLR model to the simulated and real life data with different priors for regression coefficients and error variance. We explore the effect of conjugate and non – informative prior for model parameters. Evaluate the model performance by using different model accuracy measures.

4.1.1 Introduction

In this chapter, we fit Bayesian linear regression models to simulated data with different priors for the regression parameters of the model. The simulated study aims to explore the effect of conjugate and non-conjugate priors on the regression parameters and their subsequent impact on posterior distributions and predictive performance of the model. The study includes detailed examples, demonstrating the impact of prior choices on parameter estimates and posterior predictive distributions. We conduct a stepwise study to explore and compare various Bayesian linear regression models using R software.

4.1.2 Algorithm for Fitting the Bayesian Linear Regression Model to the Simulated Data

In this section, we discuss the general approach for fitting the Bayesian Regression Model, the present general algorithm for fitting a Bayesian Linear Regression Model (BLRM), as outlined by the authors of *Bayesian Data Analysis* (Gelman et al. 1967). The algorithm can be described stepwise as below:

Step 1: Simulate the Dataset

Generate the predictor matrix $\boldsymbol{X}$ of size $n\ x\ p$ using the random draws from a multivariate normal distribution or uniform distribution. Define the true coefficients $\boldsymbol{\beta_{true}}$ a $p\ x\ 1$ vector and $\boldsymbol{\sigma^2_{true}}$.

Simulate the response $\boldsymbol{y}$ using the linear model

$$\boldsymbol{y = X\beta_{true} + \varepsilon,}$$

where $\varepsilon \sim N(0, \sigma^2_{true})$

Step 2: Define the Bayesian Model

Specify the likelihood:

$$p(y|X, \beta, \sigma^2) \propto \prod_{i=1}^{n} N(y_i | X_i^T \beta, \sigma^2).$$

Choose the prior distribution for parameters, there are mainly two types of priors namely Conjugate priors and Non – Conjugate priors.

- Conjugate prior example: $\beta \sim N(\mu_0, \Sigma_0), \sigma^2 \sim Inverse - Gamma(\alpha_0, \beta_0)$.
- Non Conjugate prior example:$\beta \sim$ Laplace (μ, b) , $\sigma^2 \sim$ Half Cauchy(0, s)

Sometime choose the flat priors when implying no prior knowledge or preference for any particular value of the parameter and Jeffreys priors, it is invariant under reparameterization.

Step 3: Compute the Posterior Distribution

Use Bayes theorem to Compute the posterior distribution

$$p(\beta, \sigma^2 | (y, X) \propto p(y|X, \beta, \sigma^2).p(\beta) \, .p(\sigma^2)$$

It is simply the prior times likelihood yields the posterior distribution.

Step 4: Fit the Model Using Sampling

Use a Bayesian Computation method to estimate the posterior distribution.

- Conjugate priors: Derive the closed – form posterior distribution.
- Non – Conjugate priors: Use Markov Chain Monte Carlo (MCMC) methods such as Gibbs Sampling or Hamiltonian Monte Carlo.

The fitting of BLR model is not the major task, it can be easily fit via brms and rstan are the libraries available in R software.

Step 5: Summarize the Posterior Results

Obtain summaries of the posterior distributions for β and σ^2 (i.e. mean, standard error, credible intervals) and access the posterior predictive distribution and Bayesian R square for the fit.

Step 6: Perform Posterior Predictive Checks

Simulate new data from the posterior predictive distribution then compare the observe data to simulate data to check model adequacy. Plot the posterior distributions, credible intervals and posterior predictive checks.

7) Compare Models with Different Priors

Repeat the above steps for many times by change different prior choices and compare models using criteria such as LOO-CV or posterior predictive accuracy.

4.1.3 Simulation based fitting BLRM with Normal - Inverse Gamma Prior

In this study, we generate the 200 observations from normal distribution. Here we define the 5 true regression coefficients and true variance. We choose the conjugate prior, normal with mean 0 and sigma 5 for regression coefficients and variance is Inverse gamma distribution with shape parameter 1.5 and scale parameter 0.1.

i) R code for fitting BLRM with Normal – Inverse Gamma Prior

```
#-------------------------------------------------------------
# Generate the Simulated Dataset
#-------------------------------------------------------------
# -------- Load Libraries and dependencies
library(brms);library(rstanarm);library(bayesplot)
# ---------- Simulated the Data
set.seed(123)  # For reproducibility
n <- 200;      # Number of Observations
p <- 5         # Number of Parameters
# Generate synthetic data
X <- matrix(rnorm(n * p), nrow = n, ncol = p) #Predictor matrix
beta <- c(1, -2, 3, 0, 0)                     # True regression
coefficients
sigma2 <- 2                                   # True variance
```

```
 # Response variable
 y <- X %*% beta + rnorm(n, mean = 0, sd = sqrt(sigma2))
 Response vector
 data <- as.data.frame(cbind(y = y, X))
 colnames(data) <- c("y", paste0("X", 1:p))
#------------------------------------------------------------
# Fitting of BLR model with NIG priors
#------------------------------------------------------------
# ------------  Specify priors
 priors <- c(
    # Normal(0, 5) prior for regression coefficients
   set_prior("normal(0, 5)", class = "b"),
 # Appro IG prior for variance
    set_prior("gamma(2, 0.1)", class = "sigma", lb = 0))
 # ------------- Fit the model
 model <- brm(
   formula = y ~ ., data = data,
   family = gaussian(),
   prior = priors,
   chains = 4,          # Number of MCMC chains
   iter = 2000,        # Number of iterations per chain
   seed = 123)        # For reproducibility
 # -------------- Summarize results
 summary(model)
 # -------------- Bayesian R square Calculation
 bayes_R2(model)
  # ---- Posterior Predictive Distribution Check
  pp_check(model)
```

ii) Output and Interpretation of NIG model

Here, first we fit the BLR model by using NIG prior and estimate the regression coefficients and their standard error. Here we construct the 95% credible interval for each regression parameter, the credible interval interpreted as there is a 95% probability that the true of the respective parameter is lie in that interval.

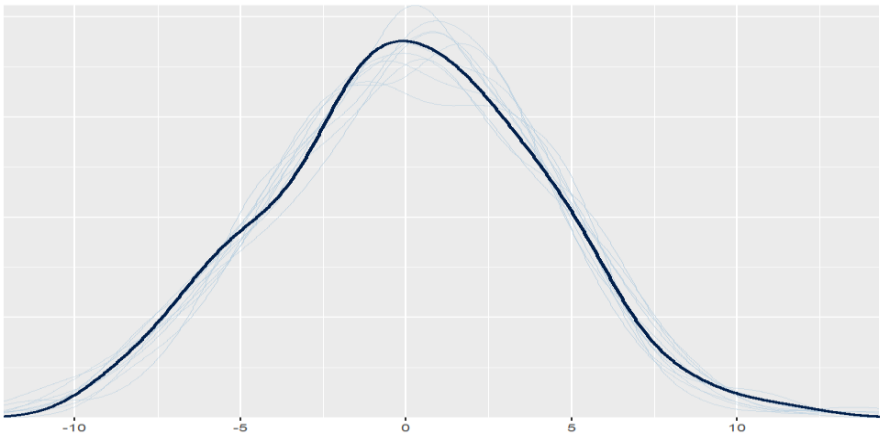
	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	0.08	0.1	-0.12	0.28
X1	1.07	0.11	0.86	1.29
X2	-2.08	0.11	-2.28	-1.87
X3	3.05	0.11	2.84	3.26
X4	0.02	0.1	-0.18	0.22
X5	-0.11	0.1	-0.31	0.09

The Intercept is around 0 indicate that there is negligible effect on the response. The predictor X1, X2 and X3 has significant effect on the response, X1 and X3 has positive effect and X2 has strong negative effect on the prediction of response. The predictor X4 and X5 both have small effect on response and indicate there insignificant for the model.

The posterior variance of the baseline model is estimate as 1.46 with estimated error 0.08 and 95% credible interval for posterior variance is [1.32, 1.63]. The estimate Bayesian R square for the NIG model is 0.87427 with estimated error 0.006377. The (2.5, 97.5)% percentile of the Bayesian R square is (0.85965, 0.88435).

iii) Posterior predictive distribution check

The posterior predictive distribution of NIG model of 10 posteriors drawn and the dark blue line is observed density of response variable.



The graph suggests that, there is bell – shaped symmetric density of observe response and the 10 posterior draws from the NIG model are well capture the relation with the observe response (the dark blue line).

4.4 Simulation Based fitting BLRM with Normal - Scaled Inverse χ^2 Prior

In this study, we generate the 200 observations from normal distribution. Here we define the 5 true regression coefficients and true variance. We choose the conjugate prior, normal

prior with mean 0 and variance for regression coefficients and variance is Normal – Scaled Inverse χ^2 Distribution with parameter 4 degree of freedom and scale is 2

i) R code for Normal – Scaled Inverse χ^2 model.

```
# ------------------------------------------------
 # Fitting of BLR model with Normal - SI χ² priors
 #------------------------------------------------
 # -------------------Define the priors-----------------
 prior2 <- c(
 # Normal(0, 5) prior for regression coefficients
 prior(normal(0, 5), class = "b"),
 # Scaled inverse chi-square prior for variance
 prior(scaled_inv_chi_square(1, 1.5), class = "sigma"))
 # ----------Fit the model ----------------------------
 model2 <- brm(
   y ~ .,data = data,
   prior = prior2,
   chains = 4,           # Number of chains
   cores = 4,            # Number of cores for parallel
 processing
   iter = 2000,          # Number of iterations per chain
   warmup = 1000,        # Warm-up iterations to discard
   control = list(adapt_delta = 0.95))# Control to avoid
 divergent transitions
 # -----------Print the summary of the model -------------
summary(model2)
 # ----------- Bayesian R square Calculation -------------
bayes_R2(model2)
 # ----------- Posterior Predictive Distribution check ----
pp_check(model2)
```

ii) Output and Interpretation of Normal – Scaled Inverse χ^2 model

Here, first we fit the BLR model by using Normal – Scaled Inverse χ^2 prior. Here we define a common normal prior for all regression coefficients and a scaled inverse χ^2 prior for variance and estimate the regression coefficients with their standard error and construct the 95% credible interval for each regression parameter.

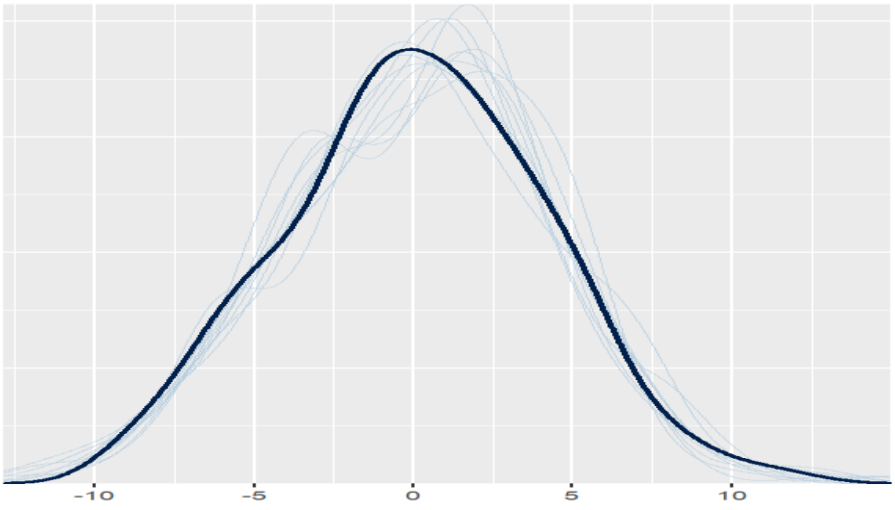
	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	0.08	0.1	-0.13	0.29
X1	1.07	0.11	0.84	1.29
X2	-2.08	0.1	-2.29	-1.87
X3	3.05	0.11	2.84	3.27
X4	0.02	0.1	-0.18	0.22
X5	-0.11	0.1	-0.31	0.08

In this summary output of the Normal-Scaled Inverse χ^2 model, there is a small change in the posterior mean or estimates of the model with their estimated error. There is also a change in the credible interval of the regression coefficients.

The posterior variance and there estimated error both also decreases, that is 1.46 with estimated error 0.3 and 95% credible interval for posterior variance is [1.34, 1.47]. The estimate Bayesian R square for the Normal – Scaled Inverse χ^2 model is 0.8744 with estimated error 0.00616. The (2.5, 97.5)% percentile of the Bayesian R square is (0.86044, 0.8841).

iii) The Posterior Predictive Distribution check

The posterior predictive distribution of Normal - Scaled Inverse χ^2 model of 10 posteriors drawn and the dark blue line is observed density of response variable.



4.1.5 Simulation Study of Fitting Normal - Cauchy Prior

In this study, we generate the 200 observations from normal distribution. Here we define the 5 true regression coefficients and true variance. We choose the conjugate prior that is normal for regression coefficients with mean 0 and sigma 5. The prior for variance it is Cauchy prior with location parameter 2 and scale 0.1. The aligns with the assumption of the Gaussian likelihood in the Normal - Cauchy model.

i) R code for fitting BLRM with Normal Cauchy model

```
#-------------------------------------------------
# Fitting of BLR with Normal – Cauchy Prior
#-------------------------------------------------
# ------------ Specify the priors ----------------
prior3 <- c(prior(normal(0, 20), class = "b"),
  # Cauchy prior for variance (location = 2, scale = 0.1)
  prior(cauchy(2, 0.1), class = "sigma"))
# ---------- Fit the model ----------------------
model3 <- brm(
  y~.,data = data,          # Use the dataset
  prior = prior3,        # Specify the priors
  chains = 4,           # Number of chains
  cores = 4,            # Number of cores for parallel
processing
  iter = 2000,          # Number of iterations per chain
  warmup = 1000,        # Warm-up iterations to discard
  control = list(adapt_delta = 0.95))  # Control to avoid
divergent transitions
 # -------- Print the summary of the model ------------
summary(model3)
 # -------- Bayesian R square ------------------------
bayes_R2(model3)
 # -------- Posterior Prediction of the model ----------
pp_check(model3)
```

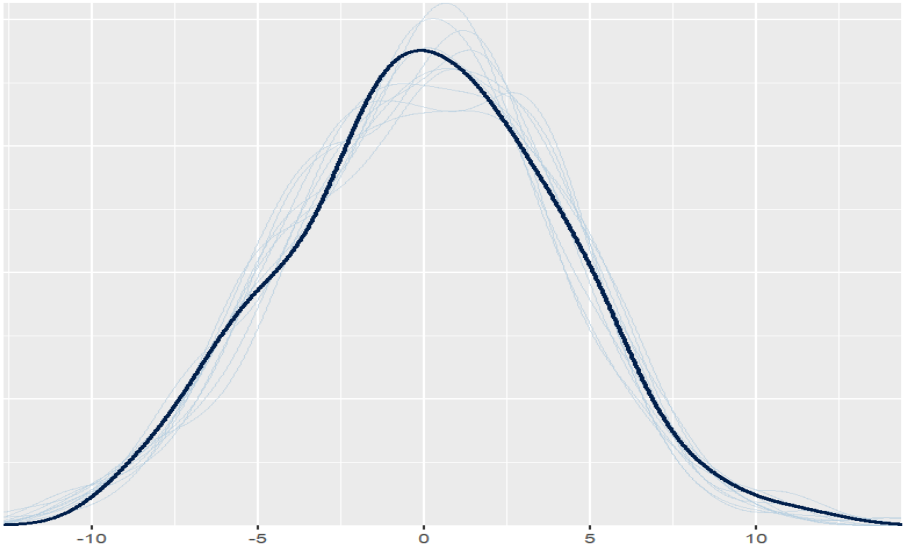
ii) Output and Interpretation of Normal – Cauchy model

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	0.08	0.1	-0.12	0.28
X1	1.07	0.11	0.85	1.3
X2	-2.08	0.11	-2.3	-1.86
X3	3.05	0.11	2.84	3.26
X4	0.02	0.11	-0.18	0.23
X5	-0.11	0.1	-0.31	0.09

In this summary output of Normal – Cauchy model, there is small change is posterior mean or estimates of the model. The posterior variance is 1.46 with estimated error 0.8 and 95% credible interval for posterior va riance is [1.34, 1.65]. The estimate Bayesian R square for the Normal – Cauchy model is 0.87418 with estimated error 0.0065. The (2.5, 97.5)% percentile of the Bayesian R square is (0.85892, 0.88437).

iii) Posterior Predictive Distribution check for Normal Cauchy BLRM

The posterior predictive distribution of Normal - Scaled Inverse χ^2 model of 10 posteriors drawn and the dark blue line is observed density of response variable.



The graph suggests that, there is bell – shaped symmetric density of observe response and the 10 posterior draws from the Normal Cauchy model are closed to each other. The model well captures the relation with the observe response (the dark blue line) but there is some high variability in prediction at the peak of the distribution where the model is quite fails to capture the variance.

4.1.6 Simulation based fitting BLRM with Flat – Jefferey's' Prior

In this study, we generate the 200 observations from normal distribution. Here we define the 5 true regression coefficients and true variance.

We choose the flat prior for regression coefficients means there is no preference to any value of the parameter and Jefferey's prior for variance. The aligns with the assumption of the Gaussian likelihood in the Flat – Jeffrey's model.

i) R code for Fitting BLRM with Flat – Jefferey's Prior

```
#-----------------------------------------------------------
# Fit the model with flat priors for betas and approximate
Jeffrey's prior for sigma
#-----------------------------------------------------------
# -------- Fit the Model ----------------------------------
 model4 <- stan_glm(y ~ .,data = data,
family = gaussian(),     # Gaussian likelihood
# Flat prior for coefficients (default)
prior = NULL,
# Flat prior for intercept (default)
  prior_intercept = NULL,
# Default prior for sigma (Jeffreys-like)
  prior_aux =  NULL,
adapt_delta = 0.95)   # Increase adapt_delta for stability
# ------ Summarize results ------------------------------
 summary(model4)
# ------ Bayesian R square ------------------------------
bayes_R2(model4)
# ------ Posterior predictive Distribution checks -------
pp_check(model4)
```

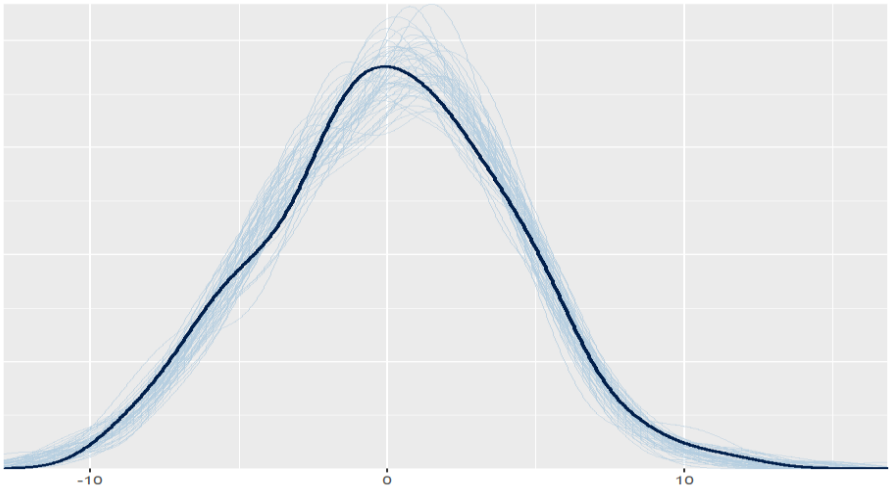
ii) Output and Interpretation of Flat – Jefferey's Model

	mean	sd	10%	50%	90%
(Intercept)	0.1	0.1	-0.1	0.1	0.2
X1	1.1	0.1	0.9	1.1	1.2
X2	-2.1	0.1	-2.2	-2.1	-1.9
X3	3.1	0.1	2.9	3.1	3.2
X4	0	0.1	-0.1	0	0.1
X5	-0.1	0.1	-0.2	-0.1	0
sigma	1.5	0.1	1.4	1.5	1.6

In this summary output Flat – Jefferey's model, there is large changes is posterior mean or estimates of the model. The summary of this model suggests that the predictors X4 and X5 are completely insignificant for the model.

The increase in posterior variance is around 1.5 with estimated error 0.1 and 95% credible interval for posterior variance is [1.36, 1.64]. The estimate Bayesian R square for the Flat – Jefferey model is 0.87297 with estimated error 0.0001. The (2.5, 97.5)% percentile of the Bayesian R square is (0.864839, 0.89606).

iii) Posterior Predictive Distribution of BLRM for Flat – Jefferey Prior



The graph shows that, there is symmetric bell - shaped density of observe response and the 10 posterior draws from the Flat – Jefferey's model is well capture the relation with the observe response (the dark blue line) but the variability in predictions is higher at the peak of the density, where the model seems to underperform in capturing the variance.

4.1.7 Evaluation of the models

To determine which model is best for prediction, we used some measure for evaluation the model. The measures are Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Bayesian R^2. The evaluation metrics

Model	Prior	RMSE	MAE	Bayes R2
M1	Normal Inverse Gamma	4.07606	3.2698	0.78378
M2	Normal Scaled Inverse	4.07568	3.27013	0.79982
M3	Normal Cauchy	4.07813	3.27314	0.79088
M4	Flat Jefferey	4.09687	3.28396	0.36007

The Lower RMSE indicates better model performance as it measures the average of prediction errors. All models have similar RMSE values, but Normal – Scaled Inverse χ^2 model (NSI_fit) has the lowest rmse.

The lower MAE indicates better model performance. All models have Similar MAE, again NSI_fit has the lowest MAE.

The Higher R^2 indicates better model fit to the data. NSI_fit has the highest R^2, indicating it explains the most variance among the models.

Based on all three metrics (RMSE, MAE, R^2), the NSI_fit model performs the best for prediction.

4.2 Fitting Bayesian Linear Regression Model to Real Life Data

- **About the Diabetes Data**

In this example, we examine the Diabetes Data where the response variable is a quantitative measure of disease progression one year after baseline. We first transform the data into standardization, means each variable is centred to 0 and the unit variance of each variable. We aim to compare the performance of a linear model and a Bayesian linear regression model with different prior for regression coefficients and for variance parameter. Given the Standardized data there isn't the bell – shape symmetric distribution of the response variable. This comparison will highlight the advantages of using Bayesian linear regression model with different priors, particularly in capturing the variability that a linear model might overcome.

The Description of Diabetes Dataset

Number of Observation: 442

Number of Variables : 11

The Details of the variables are belows

- age Age in years of patient
- sex Gender of the Patient (Female = 0 and Male = 1)
- bmi Body Mass Index of patient
- bp Average Blood Pressure
- s1 Total serum cholesterol, tc
- s2 Low-density lipoproteins, ldl
- s3 High-density lipoproteins, hdl
- s4 Total cholesterol.
- s5 Possibly log of serum triglycerides level, ltg,
- s6 Blood sugar level
- target A quantitative measure of disease progression one year after baseline.

Reference for Diabetes Dataset: The Diabetes dataset is download from the sklearn package in the python.

4.2.2 Choosing of different priors for error variance

We fit the BLR model to the Diabetes data using a common prior for β and different priors for the error variance. The conjugate priors used for fitting the BLR model are the Inverse Gamma, Scaled Inverse χ^2, and Cauchy priors. A non-informative prior, specifically the choose flat prior, is used for β and Jefferey prior for error variance (σ^2) when fitting the BLR model

4.2.2 Results of BLR Model fitting

i). Summary of OLS model

Here, first we fit the model by using OLS method and estimate the regression coefficients and their standard error.

The OLS model summary

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.00	0.03	0.00	1.00
Age	-0.1298	0.78	-0.17	0.87
Sex	-3.111	0.79	-3.92	0.00
Bmi	6.743	0.86	7.81	0.00
bp	4.208	0.85	4.96	0.00
s1	-10.28	5.41	-1.90	0.06
s2	6.184	4.40	1.41	0.16
s3	1.311	2.76	0.48	0.63
s4	2.297	2.10	1.10	0.27
s5	9.745	2.23	4.37	0.00
s6	0.8772	0.86	1.03	0.31

From the above summary table of OLS model, the intercept is not significant to the response and sex, bmi, bp, s1 and s5 are the significant predictors for the response in the model. The predictors age, s2, s3, s4 and s6 are the non-significant to the OLS model. Our decision for significance and non-significance of the predictors based on p – value of the respective regression coefficient.

Then we fit the baseline model to the response variable with the all predictors by using the Bayesian linear regression model with common prior for regression coefficient i.e. the as prior

available information we set the prior for regression coefficient is normal distribution with mean 0 and variance = 1. The prior for variance is half – normal with mean 0 with variance 1.

ii. Summary of Baseline model

Here, we fit the BLR model to the target variable with the common prior for all regression variable that is a normal prior and half – normal prior for variance.

The baseline model summary

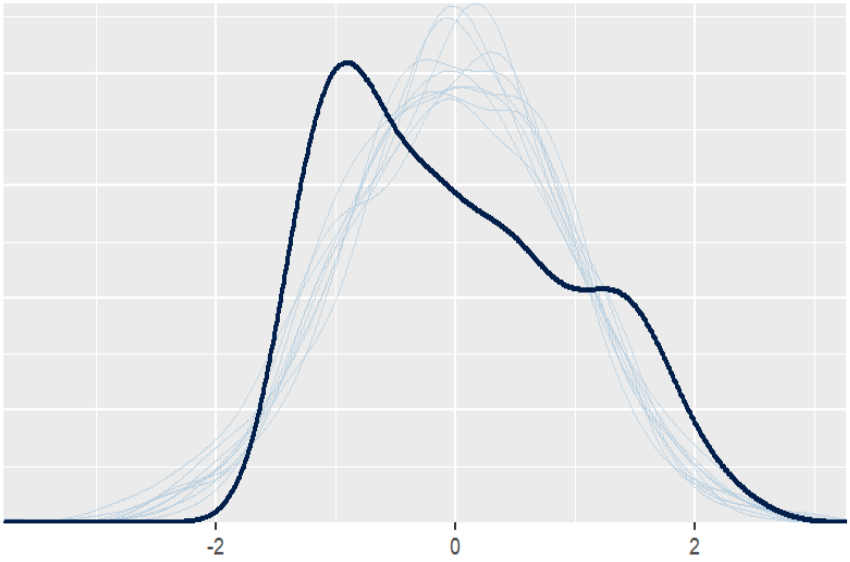
	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	0	0.03	-0.07	0.07
age	0.27	0.61	-0.9	1.45
sex	-1.67	0.63	-2.89	-0.44
bmi	4.92	0.65	3.61	6.18
bp	3.15	0.65	1.85	4.41
s1	-0.16	0.79	-1.67	1.37
s2	-0.75	0.79	-2.32	0.79
s3	-2.27	0.71	-3.63	-0.89
s4	1.57	0.8	0.02	3.12
s5	4.21	0.7	2.84	5.56
s6	1.44	0.64	0.19	2.67

From the above summary table of baseline BLR model, the intercept and s1 are only predictors they have no significant to the model. The predictors sex, bmi, bp and s5 are the not significant predictors for the response in the OLS model. However, with the Bayesian Linear Regression (BLR) model, the appropriate choice of prior affects the regression coefficients and their uncertainty estimates.

The posterior variance of the baseline model is estimate as 0.72 with estimated error 0.03 and 95% credible interval for posterior variance is [0.67, 0.78], means there is 95% probability that true value of posterior variance is lie within this interval.

The Bayesian R^2 estimate for baseline model is 0.3934 with estimate error 0.029. The 95% credible interval for Bayesian R^2 (2.5, 97.5) percentile, is (0.333, 0.4477).

iii) Posterior Predictive Distribution of the Baseline Model



The above graph shows, the posterior predictive distribution of the baseline model. We observe that the observed density of the response variable (the dark blue line) significantly differs from the predictive densities generated by posterior 10 draws.

iv) Summary of refit model

After several trials of fitting and adjusting the priors for both regression coefficients and variance, we finally defined specific prior for each regression coefficient. Here we observed that the density of the response variable was not bell – shaped symmetric. Consequently, we changed the family of the response variable from Gaussian to skew - normal. The resulting refit model gives us the strong improvement in Bayesian R^2 statistics and slight decline in posterior variance.

The summary table of refit model

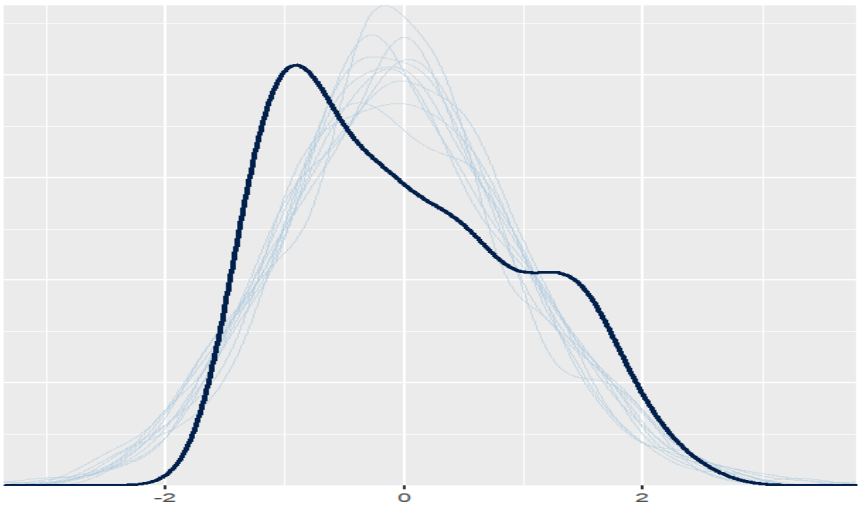
	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	0	0.03	-0.06	0.06
sex	-2.98	0.42	-3.8	-2.18
bmi	6.96	0.62	5.75	8.16
bp	4.08	0.42	3.24	4.91

s1	-4.13	0.75	-5.57	-2.66
s2	1.82	0.74	0.4	3.3
s3	-1.91	0.44	-2.76	-1.05
s4	1	0.59	-0.16	2.14
s5	7.62	0.54	6.57	8.69

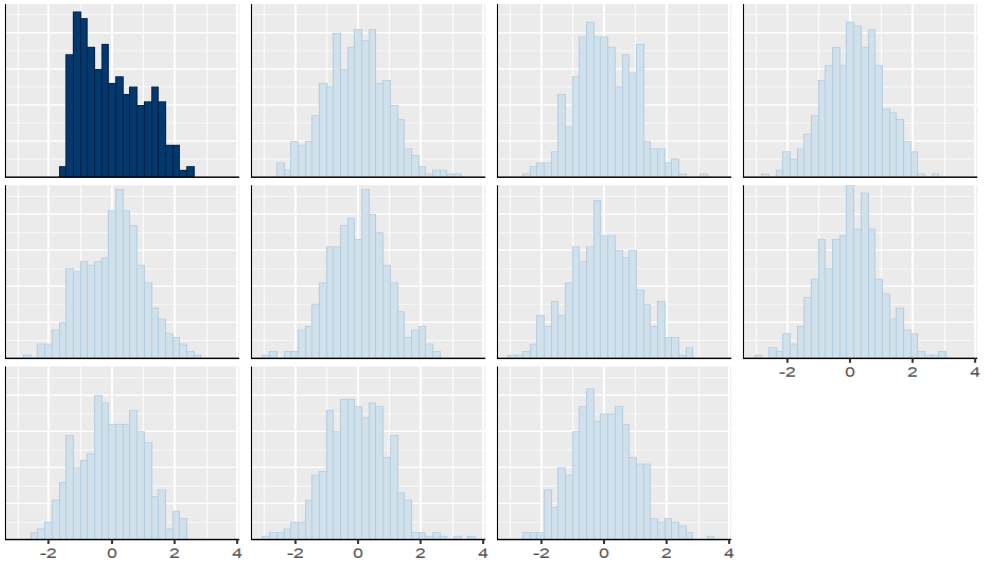
From the above summary table of the refit BLR model, the intercept is the only predictor that is not significant. All other predictors are significant to the model prediction. Specifically, the predictors bmi, bp, s2 s4 and s5 have a positive effect on the model prediction, while the predictors sex, s1 and s3 are the negative effect.

The posterior variance of the refit model is estimated to be 0.69 with an estimated error 0.01 and 95% credible interval for posterior variance is [0.66, 0.72], means there is 95% probability that true value of posterior variance is lies within this interval. The Bayesian R^2 estimate for refit model is 0.50842 with estimate error 0.0194. The (2.5, 97.5) percentile for Bayesian R^2 is (0.4672, 0.543916), indicating a sightly improvement in R^2 statistics and reduction in the posterior variance of the refit model.

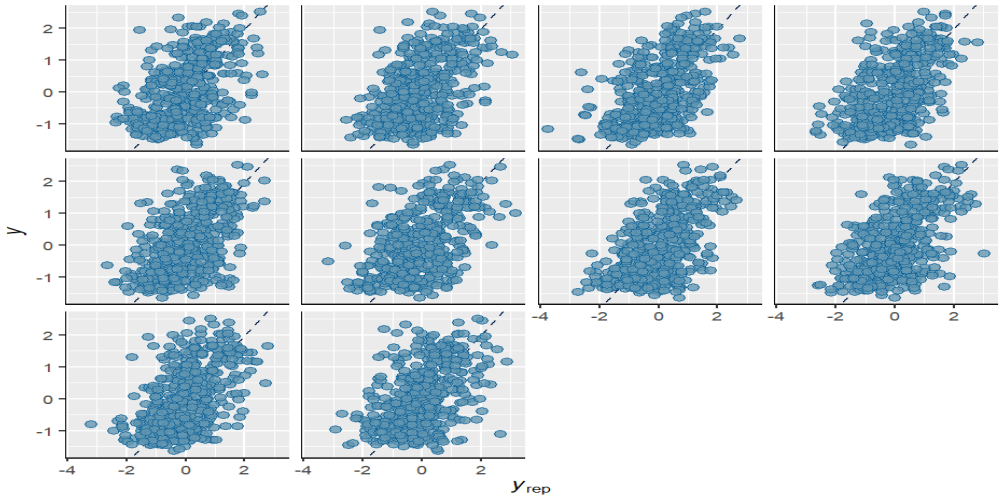
v) The Posterior Predictive Distribution for refit model



The above graph represents the posterior predictive distribution of the refit model. We observe that some peaks of posterior draws are align with the peak of the original response density, indicating slight improvement in the model's predictive performance.



The above 11 histogram shows that the original density of the target variable (the dark blue at left top corner), while the remaining 10 are the posterior draws density generated by the refit model. Although the model does not fully capture the observed distribution of the target variable



The above scatter plot shows the relationship between the observed target values and the posterior predicted samples generated by the refit model. A positive correlation is observed between the observed target and the predicted posterior samples.

4.2.3 Model Comparison

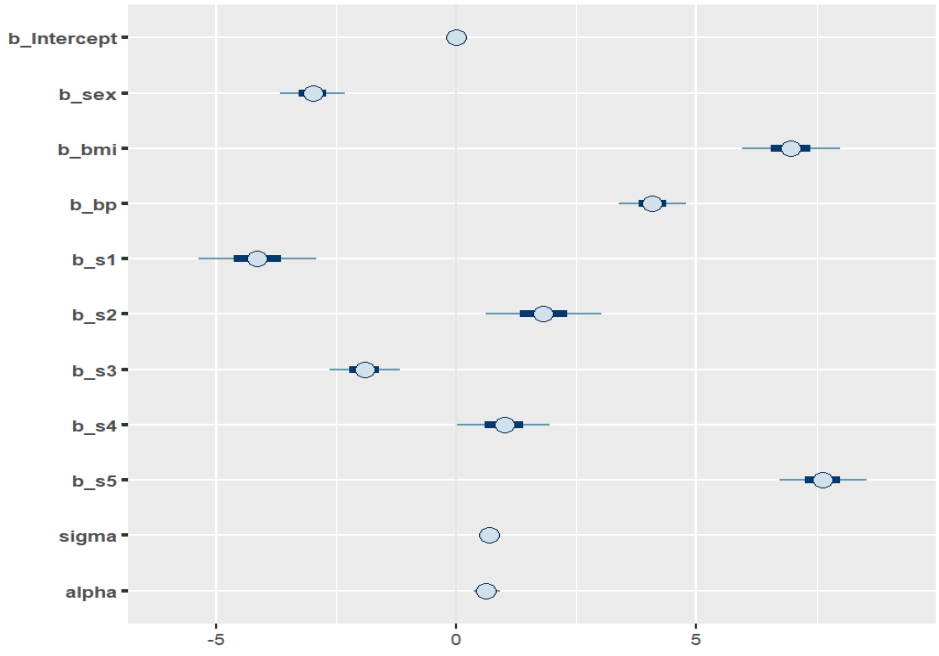
Model Comparison between baseline model and refit model by using Leave – One – Out Cross Validation (LOOCV) based on the Expected Log Predictive Density (ELPD).

	elpd_diff	se_diff
refit	0	0
Baseline	-15.3	5.2

The positive values indicate the refit model performs better in terms of predictive accuracy compared to the baseline fit. Given the higher ELPD, the refit is the preferred model for prediction.

4.2.4 Visualize the posterior distributions and marginal effects of the refit model.

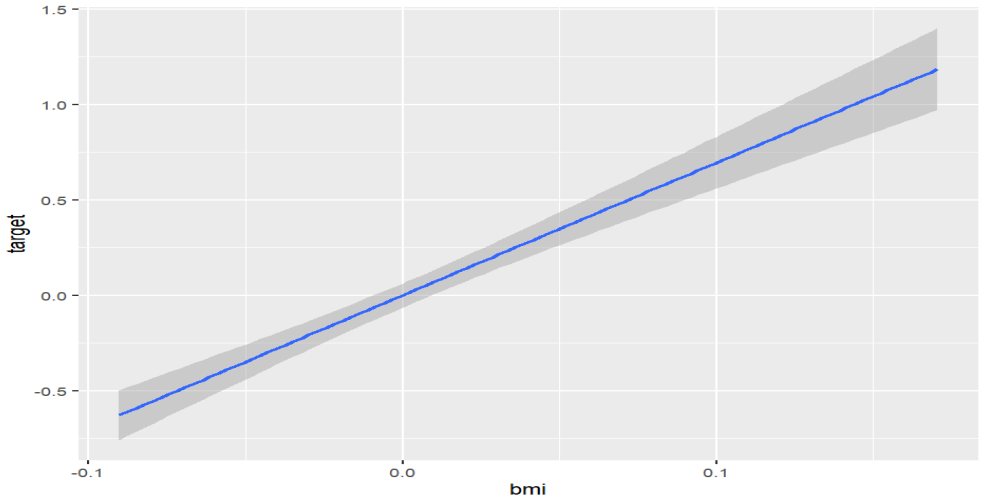
The posterior distributions of the refit model provide insights into the uncertainty and range of plausible values for the model parameters.



The above MCMC plot shows that there has no significant effect of intercept on the model. There is strong positive significant effect of BMI, s5 and BP predictors on the model. The S2 and S4 predictors have a small positive significant effect. There is a negative significant relationship between the sex, S1, and S3 predictors and the model. After standardization, the posterior variance parameters lie around 0.7 with a small estimated error.

The Marginal Effect of BMI predictor on the model

Marginal effects plots visualize the relationship between BMI and the target, showcasing how changes in predictors impact the response while holding other variables constant.



This graph represents the relationship between **BMI (body mass index)** and the **target variable** in the refit model. It suggests that BMI has a significant positive effect on the target variable, with minimal uncertainty (shaded area is narrow).

4.2.5 Testing of Hypothesis for individual coefficients

Here, we test hypothesis related to individual regression coefficient in the BLR model.

Hypothesis	Estimate	Est. Error	CI.Lower	CI.Upper	Evidence Ratio	Posterior Probability
bmi = 0	6.96	0.62	5.75	8.16	NA	NA
bmi > 0	6.96	0.62	5.94	7.99	Inf	1

The results of the hypothesis testing indicate strong evidence for the effect of BMI in the refit model. The Posterior mean of the BMI coefficient is 6.72, with standard error is 0.85. The 95% credible interval for the coefficient is [5.06, 8.37], which does not include zero. Therefore, we reject the hypothesis bmi = 0. The evidence ratio is infinite and the posterior probability 1,

strongly favoring the hypothesis bmi > 0. We conclude that BMI has significant effect on model prediction.

4.2.6 Prediction of new instance

Here, we predict the 5 new instances of Diabetes Data with the fitted BLR model and the observe value and the model value with their estimated error. The (2.5, 97.5) percentile of the predicted response are printed below.

	Observe_value	Estimate	Est.Error	Q2.5	Q97.5
1	-0.0147	0.734346	0.685285	-0.58789	2.112787
2	-1.00053	-1.04886	0.679283	-2.38518	0.334255
3	-0.14442	0.338085	0.694612	-1.01441	1.694874
4	0.698721	0.166073	0.695629	-1.19586	1.547394
5	-0.22224	-0.29223	0.697889	-1.65861	1.164328

The above table shows the observe values and posterior estimates or model prediction with estimated error. For each prediction we compute the (2.5, 97.5) percentile of the prediction.

4.2.7 Posterior Summary of the refit model

The below table is the posterior summary of refit model, estimate for each parameter with their estimated error.

	Estimate	Est.Error	Q2.5	Q97.5
b_Intercept	0.00	0.03	-0.06	0.06
b_sex	-2.98	0.42	-3.80	-2.18
b_bmi	6.96	0.62	5.75	8.16
b_bp	4.08	0.42	3.24	4.91
b_s1	-4.13	0.75	-5.57	-2.66
b_s2	1.82	0.74	0.40	3.30
b_s3	-1.91	0.44	-2.76	-1.05
b_s4	1.00	0.59	-0.16	2.14
b_s5	7.62	0.54	6.57	8.69
sigma	0.69	0.01	0.66	0.72
alpha	0.62	0.17	0.33	0.98
Intercept	0.00	0.03	-0.06	0.06
lprior	-6.01	1.54	-9.73	-3.74
lp__	-476.88	2.36	-482.30	-473.27

The above summary table provides posterior summaries of parameters from a refit model. The posterior uncertainty is quantified by Estimated Error and credible interval [-0.06, 0.06] suggests the intercept is likely close to zero. The predictors BMI, BP, S2 and S5 all have strong positive association with the refit model. The predictors SEX, S1 and S3 have strong negative association with the refit model. The sigma, the residual standard deviation representing the variability of the response variable. The lprior is the log of the prior probability density of the parameters and lp_ is the combining the likelihood and the prior.

4.2.8 Evaluation of model

Here, we evaluate the two fitted BLR models by drawing random samples of size 200 from the mode and calculating the RMSE, MAE and R^2 statistics

Model	RMSE	MAE	R2
Baseline	1.003014	0.865247	0.620917
Refit	0.99891	0.852051	0.649267

The Refit model has a slightly lower RMSE and MAE compared to the Baseline model and higher R^2 value of refit model explains more of the variance in the target variable. The Refit model outperforms the Baseline model on all three metrics (RMSE, MAE, and R^2), indicating it is better for prediction.

4. 2.8 Conclusions

- In the diabetes dataset scenario, the OLS model provides a summary with many insignificant predictors, and there is a negligible effect of the regression coefficients on the response. In the baseline model summary, only one predictor is insignificant, and in the refit model, no predictors are insignificant to the model.

- The refit Bayesian linear regression model shows a significant improvement in predictive accuracy compares to the baseline model, as indicated by the higher Bayesian R^2 (0.51compares o 0.393) and better Leave – One – Out Cross- Validation result. The posterior variance decreased slightly in the refit model (0.69 vs 0.72), indicating better model precision.

- Adjusting priors and define the prior for individual prior for each regression coefficient. The family using a skew normal response distribution improved the model's performance, demonstrating the importance of the prior and likelihood in Bayesian regression.

- The refit model's posterior predictive distributions align better with the observed target density, indicating improved model predictions. Since the observed distribution of response variable is not bell – shaped symmetric, selecting appropriate priors for regression coefficients and variance could further enhance the model's performance.

4.2.9 R code

```
# ------------------------------------------------------------
# Bayesian Linear Model Fitting
# ------------------------------------------------------------
-
# ------- Load necessary libraries
library(brms);library(rstanarm)       # for Bayesian
Regression
library(bayesplot);library(tidyverse) # for data
visualization
# ------- Importing the dataset
diabetes <-
read.csv("C:\\Users\\kalpe\\OneDrive\\Documents\\Datasets\\
diabetes_dataset.csv")
head(diabetes)     # print first 6 observations of the
Diabetes dataset
diabetes$target <- scale(diabetes$target)
# Scatter plot between all numeric variables
scatter_plot_matrix <- pairs(diabetes)
# --------------------------------
#  Fitting the Linear Model
# --------------------------------
lm_model <- lm(target~.,data = diabetes)
summary(lm_model)  # summary of OLS model
confint(lm_model)  # confidence interval for regression
coefficients.
#------------------------------------------------------------
-
#Fit the baseline model using common prior all regression
coeff
#------------------------------------------------------------
-
```

```
regression coefficients
priors <- c(prior(normal(0,1),class = "b"), # Normal(0,1)
            prior(normal(0,1),class = "sigma",lb = 0)) # half
normal prior
br_fit <- brm(target ~.,data =  diabetes,
              family = gaussian(),
              prior = priors,control =
list(adapt_delta=0.95))
summary(br_fit)
# The Bayesian R^2 calculation
bayes_R2(br_fit)
# print Posterior mean, Estimated Error, and 95% cri
fixef(br_fit)
# Posterior Predictive Distribution Check
pp_check(br_fit)

# ----------------------------------------
# Redefine the priors for individual regression coeff
# ----------------------------------------
repriors <- c(          # Weakly informative priors for
regression coefficients
  prior(normal(7,1), class = "b", coef = "bmi"),  # Shrinkage
prior for 'age'
  prior(normal(4,0.5),class = "b",coef = "bp"),
  prior(normal(-3,0.5),class = "b",coef = "sex"),
  prior(normal(-4,1),class = "b",coef = "s1"),
  prior(normal(2,1),class = "b",coef = "s2"),
  prior(normal(-2,0.5),class = "b",coef = "s3"),
  prior(normal(1,0.7), class = "b", coef = "s4"),
  prior(normal(7.5,0.7),class = "b",coef = "s5"),
  prior(normal(0.68, 0.02), class = "sigma", lb = 0),# Prior
for the standard deviation
  prior(normal(0.6,0.2),class = "alpha",lb = 0.3))
# Fit the model with the specified priors and change the
family of the response
refit <- brm(
  formula = target ~ sex + bmi + bp + s1 + s2 + s3 + s4 +
s5,
  data = diabetes,
  family = skew_normal(),    # Assuming a continuous response
variable
  prior = repriors,
  chains = 4,              # Number of MCMC chains
  iter = 2000,             # Total iterations per chain
  warmup = 1000,           # Warmup (burn-in) samples per
chain
  seed = 123)               # For reproducibility
# Summarize the model
summary(refit)
# Posterior predictive checks
pp_check(refit)  + theme(legend.position = "none")
```

```
 # Bayesian R square
 bayes_R2(refit)
 # Check Model convergence
 plot(refit)
 # posterior Predictive Check
 pp_check(refit,type = "hist")  + theme(legend.position =
 "none")  # Histogram of simulated vs. observed data
pp_check(refit, type = "scatter") # scatter plot of
predictors
# Comparison of LOOCV between baseline model and the refit
model
 loo_br_fit <- loo(br_fit)
 loo_refit <- loo(refit)
 loo_compare(loo_br_fit,loo_refit)


 # Check coefficient of Bayesian regression model
 fixef(refit)
 # Hypothesis Testing for individual regression coefficient.
 hypothesis(refit,"bmi = 0")
 hypothesis(refit,"bmi > 0")
 # Visualize the posterior distributions and marginal effects.
 mcmc_plot(refit)
 marginal_effects(refit,effects = "bmi")
 # Predict the new instance by refit model
 new_data <- diabetes[1:5,]
 # Obtain credible intervals for predictions
 refit_pred <- predict(refit, newdata = new_data, probs =
 c(0.025, 0.975))
 data.frame("Observe_value" = new_data$target,refit_pred)
 # Posterior density summary and credible intervals for each
 regression coefficient.
 posterior_summary(refit)  # Summarize posteriors
 # Get 95% credible intervals
posterior_interval(refit, prob = 0.95)
```

References

[1] W. Feller. An Introduction to Probability Theory and Its applications. 1967; John Wiley, Ed.3, vol. 1.

[2] A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, D. Rubin. Bayesian Data Analysis. 2014; CRC, Ed.3, Boca Raton.

[3] Baruna Wundervald. National University of Ireland, Maynooth. *Bayesian Linear Regression, June* 2019.

[4] C. Robert. The Bayesian Choice: from Decision-Theoretic Foundations to Computational Implementa tion. 2007; Springer, Ed. 2, New York

[5] W. Hastings. Monte Carlo sampling using Markov chains and their applications. Biometrika. 1970; 57: 97-109.

[6] Ekiz, U. (2002) 'A Bayesian method to detect outliers in multivariate linear regression', Hacettepe Journal of Mathematics and Statistics, Vol. 31, pp.77–82 [online]

[7] Evans, Sara,"Bayesian regression analysis". (2012). *Electronic Theses and Dissertations series*.

[8] Bishop, CM & Tipping, ME 2003, Bayesian Regression and Classification. in JAK Suykens, I Horvath, S Basu, C Micchelli & JV (eds), Advances in Learning Theory: Methods, Models and Applications. NATO Science Series, III: Computer and Systems Sciences, vol. 190, IOS Press, pp. 267-285

[9] The Diabetes Dataset used in Application Study