Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 11, No. 3, 2024

Application of Bayesian Method in Linear Regression

Yiteng Zhang *

Jurong Country Garden School, Zhenjiang, Jiangsu, 212499, China
*** Corresponding author Email:** z2406254047@foxmail.com

Abstract: This paper investigates the application of Bayesian methods in linear regression. Firstly, the basic principles of linear regression and Bayesian methods were introduced. Then, the construction and inference methods of Bayesian linear regression models were discussed in detail. Furthermore, the application of Bayesian methods in other regression problems was explored, and their limitations and improvement directions in practice were analyzed. Finally, the main research findings were summarized and suggestions for future research directions were proposed.

Keywords: Bayesian Method; Linear Regression; Bayesian Linear Regression Model.

1. Introduction

Linear regression is an important forecasting model in statistics, which is widely used in economic, financial, medical and other fields. The traditional linear regression method is mainly based on least square method, but this method may have bias and instability when dealing with some complex data problems. In recent years, Bayes method as a statistical inference method has gradually attracted attention. Bayesian methods provide more accurate parameter estimates and model uncertainties, and therefore have unique advantages when dealing with linear regression problems.

In the real world, data often has problems such as noise, outliers and missing values, which can lead to the failure of traditional linear regression methods. Bayesian method can deal with these problems better, and can provide more comprehensive information, including the prior distribution of parameters, the uncertainty of the model and the evaluation of the model effect. Therefore, the purpose of this study is to explore the application of Bayesian method in linear regression, in order to provide new ideas and methods for practical application.

By combing and analyzing the relevant literature, it can be found that the existing researches mainly focus on the theoretical research, algorithm optimization and numerical calculation of Bayesian linear regression model. However, there are still relatively few researches on the application of Bayesian method in linear regression, and the existing researches have some problems such as sample selection bias. Therefore, this study aims to explore the application of Bayesian method in linear regression in order to provide a more comprehensive, accurate and reliable method for practical application.

Linear regression is a common forecasting model in statistics, which is widely used in various practical problems. However, the traditional linear regression method may have some limitations when dealing with some complex data. As a statistical inference method, Bayesian method can provide a more robust and flexible model, so it has great research significance in dealing with linear regression problems.

Bayesian method can better deal with uncertainty in data. In traditional linear regression, models are usually built based on limited observational data, while Bayesian methods allow for both data uncertainty and prior information to be used to build models. This gives Bayesian methods greater flexibility and accuracy when dealing with complex data. Second, Bayesian methods can provide more robust models. Traditional linear regression methods usually build models based on linear assumptions, but in practical applications, the distribution of data is often non-linear. Bayesian methods are better able to capture the nonlinear features of the data, thus providing a more robust and accurate model. Finally, Bayesian methods are more interpretive when dealing with complex data. Traditional linear regression methods often build models based on statistical assumptions that are often difficult to explain. Bayesian methods, on the other hand, pay more attention to the interpretability of the model, allowing the selection of appropriate model parameters according to the characteristics of the data, so as to better interpret the model's predictions.

The application of Bayesian method in linear regression has great research significance. It can better handle uncertainty in data, provide more robust and accurate models, and have higher interpretability when dealing with complex data. These advantages make the Bayesian method has great application value in dealing with practical problems.

2. Overview of Linear Regression and Bayesian Methods

2.1. Basic Principles of Linear Regression

Linear regression is a basic forecasting model that predicts future data by predicting the linear relationship between the dependent variable (y) and the independent variable (x). The basic principle of linear regression is to find the best fit line by minimizing the residuals sum of squares (RSS). In linear regression, ordinary least squares (OLS) are usually used to estimate the parameters.

Multiple linear regression analysis is one of the most widely used statistical methods. If it is described by matrix form, its mathematical postmortem model is as follows [1]:

$$\underset{(n\times 1)}{Y} = \underset{(n\times K)}{X}\underset{(K\times 1)}{\beta} + \underset{(n\times 1)}{\cup} \tag{1}$$

Each component in the formula u_i , i= 1, 2, ... , k independent identically distributed, and u_1 follow the normal distribution that mean is 0, variance is σ^2, be $\cup \sim \mathrm{N}(0, \sigma^2 I_k)$ It is obvious that the sample likelihood function is:

$$L(\beta,\sigma|Y,X)=(2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\left[vS^2+\left(vS^2+(\beta-\beta)'\right)X'X'(\beta-\beta)\right]\right\} \quad (2)$$

In the equation, $v=n-K$,while S^2 and β are respectively $S^2=(Y-X\beta)'(Y-X\beta)/v$ and $\beta=(X'X)^{-1}X'Y$.

2.2. Bayesian Method Basics

Bayesian method is a statistical method based on probability, which describes the relationship between data and model parameters through probability distributions. An important feature of the Bayesian approach is to treat the parameters as random variables and infer their posterior distributions based on the data. This posterior distribution describes parameter uncertainty and provides more comprehensive information.

The main advantage of the Bayesian approach is its ability to deal with uncertainty and noisy data and provide more precise predictions and interpretations. In addition, Bayesian methods can also be used for model selection and diagnosis to determine the complexity and effect of the model.

2.3. Bayesian Linear Regression Model

The Bayesian linear regression model is a model that combines the Bayesian approach and linear regression[2]. In such models, the independent and dependent variables are treated as random variables and Bayes theorem is used to infer their joint distribution. In this way, Bayesian methods can be utilized to deal with the uncertainty and noise in the data and the posterior distribution can be used to evaluate the performance of the model and to interpret the results[3].

In Bayesian linear regression, algorithms such as Gaussian process Regression (GPR) or Bayesian Linear Regression (BLR) are commonly used to fit the model. These algorithms are able to automatically deal with nonlinear relationships and outliers in the data and provide more flexible and accurate predictions[4].

The application of Bayesian methods in linear regression provides a more comprehensive and precise way to deal with the uncertainty and noise in the data, and provides more flexible and accurate prediction results[4].

Bayesian analysis of multiple linear regression model:

Assume that there are k independent variables in the model:$x_1,x_2,\cdots,x_k$, m dependent variable:$y_1,y_2\cdots,y_m(m<k)$,There is the following linear functional relationship between them [4]

$$\begin{cases} y_1=\beta_{01}+\beta_{11}x_1+\beta_{21}x_2+\cdots+\beta_{k_1}x_k+\varepsilon_1 \\ y_2=\beta_{02}+\beta_{12}x_1+\beta_{22}x_2+\cdots+\beta_{k_2}x_k+\varepsilon_2 \\ \cdots \\ y_m=\beta_{0m}+\beta_{1m}x_1+\beta_{2m}x_2+\cdots+\beta_m x_k+\varepsilon_m \end{cases} \quad (3)$$

In the equation, β_{ij}(i=0,1,…,k;j=1,2,…,m) are unknown parameter,,$\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_m$ Is the random error term,they are not necessarily independent of each other, assuming that they follow a multivariate normal distribution, i.e.

$$(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_m)^T\sim N_m(0,\Sigma),\Sigma>0 \quad (4)$$

Since it is convenient to study the problem of multiple linear regression model by matrix method, the above model is transformed into matrix form

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}=\begin{pmatrix} \beta_{01} & \beta_{11} & \cdots & \beta_{k1} \\ \beta_{02} & \beta_{12} & \cdots & \beta_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{0m} & \beta_{1m} & \cdots & \beta_{km} \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}+\begin{pmatrix} \varepsilon_1 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_m \end{pmatrix} \quad (5)$$

Assuming that n sets of values of k independent variables are given and the corresponding m dependent variables are observed, the following n sets of observed values are obtained:

$$(x_{11},x_{12},\cdots,x_{1k};y_{11},y_{12},\cdots,y_{1m})$$
$$(x_{21},x_{22},\cdots,x_{2k};y_{21},y_{22},\cdots,y_{2m})$$
$$\cdots\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots$$
$$x_{n1},x_{n2},\cdots,x_{nk};y_{n1},y_{n2},\cdots,y_{nm}$$

By substituting it for Model (5), there is

$$\begin{pmatrix} y_{11} & y_{22} & \cdots & y_{1m} \\ y_{21} & y_{22} & \cdots & y_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nm} \end{pmatrix}=\begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix}\begin{pmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0m} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{k1} & \beta_{k2} & \cdots & \beta_{km} \end{pmatrix}+\begin{pmatrix} \varepsilon_{11} & \varepsilon_{12} & \cdots & \varepsilon_{1m} \\ \varepsilon_{21} & \varepsilon_{21} & \cdots & \varepsilon_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{n1} & \varepsilon_{n1} & \cdots & \varepsilon_{nm} \end{pmatrix} \quad (6)$$

If each matrix in the above equation is denoted as Y, X, β and ε from left to right, it can be further reduced to

$$Y_{n\times m}=X_{n\times(k+1)}\beta_{(k+1)\times m}+\varepsilon_{n\times m} \quad (7)$$

Or its equivalent

$$\begin{pmatrix} Y_{(1)}^T \\ Y_{(2)}^T \\ \vdots \\ Y_{(n)}^T \end{pmatrix}=\begin{pmatrix} (X\beta)_{(1)}^T \\ (X\beta)_{(2)}^T \\ \vdots \\ (X\beta)_{(n)}^T \end{pmatrix}+\begin{pmatrix} \varepsilon_{(1)}^T \\ \varepsilon_{(2)}^T \\ \vdots \\ \varepsilon_{(n)}^T \end{pmatrix},\varepsilon_{(j)}\sim N_m(0,\Sigma),j=1,2,\cdots,n \quad (8)$$

Here $Y_{(j)}$,$(X\beta)_{(j)}$,$\varepsilon_{(j)}$, j=1, 2, ... ,n denote the transpose of the matrix Y, Xβ and row j of ε respectively, which are m-dimensional column vectors. Obviously, $Y_{(j)}=(Y\beta)_{(j)}+\varepsilon_{(j)},j=1,2,\cdots,n$ since $\varepsilon_{(1)},\varepsilon_{(2)},\cdots,\varepsilon_{(n)}\sim i.\text{i.d.}N_m(0,\Sigma)$ so $Y_{(1)},Y_{(2)},\cdots,Y_{(n)}$ is also independent of each other and $Y_{(j)}$ follows the multivariate normal distribution with $(X\beta)_{(j)}$ mean vector and covariance matrix is Σ, i.e

$$Y_{(j)}\sim N_m\left((X\beta)_{(j)},\Sigma\right),j=1,2,\cdots,n \quad (9)$$

In this case, the sample likelihood function is

$$L(\beta,\Sigma^{-1})=\prod_{j=1}^{n}(2\pi)^{-m/2}|\Sigma^{-1}|^{1/2}\exp\left\{-\frac{1}{2}\left[Y_{(j)}-(X\beta)_{(j)}\right]^T\Sigma^{-1}\left[Y_{(j)}-(X\beta)_{(j)}\right]\right\}$$
$$=(2\pi)^{-mn/2}|\Sigma^{-1}|^{n/2}\exp\left\{-\frac{1}{2}\sum_{j=1}^{n}(Y-X\beta)_{(j)}^T\Sigma^{-1}(Y-X\beta)_{(j)}\right\}=(2\pi)^{-mn/2}|\Sigma^{-1}|^{n/2}\exp\left\{-\frac{1}{2}\text{tr}(Y-X\beta)^T(Y-X\beta)\Sigma^{-1}\right\} \quad (10)$$

Before further analysis of the model, matrix normal distribution is introduced, which is a generalization of multivariate normal distribution and is defined as follows: If $\tilde{Y}$ is a random matrix of households $k_1 \times k_2$ The distribution density function of $\text{vec}(\tilde{Y})$ is

$$f(\tilde{y}|\theta, P, Q) = \frac{1}{(2\pi)^{k_1 k_2/2}|P|^{k_2/2}|Q|^{k_1/2}} exp\left\{-\frac{1}{2}tr[Q^{-1}(\tilde{y}-\theta)^T P^{-1}(\tilde{y}-\theta)]\right\} \quad (11)$$

The random matrix $\tilde{Y}$ is said to follow the matrix normal distribution,And let's write it as $\tilde{Y} \sim MN_{k_1 \times k_2}(\theta, Q \otimes P)$; Here the $P_{k_1 \times k_2}$ and $Q_{k_1 \times k_2}$ is a positive definite matrix $\tilde{y} \in R^{k_1 \times k_2}, \theta \in R^{k_1 \times k_2}$,While the symbol " ⊗ " denotes the Kronecker product of the matrix.

If write $\acute{\beta} = (X^T X)^{-1} X^T Y$,S= $(Y - X\acute{\beta})^T X^T X(Y - X\acute{\beta})$ It is not difficult to verify that the following equation holds

$$(Y - X\acute{\beta})^T X(\beta - \acute{\beta}) = [Y - X(X^T X)^{-1} X^T Y]^T X(\beta - \acute{\beta}) = 0 \quad (12)$$

Using this condition, the quadratic term $(Y - X\beta)^T (Y - X\beta)$ in Equation (10) is decomposed as follows

$$(Y - X\beta)^T (Y - X\beta) = S + (\beta - \acute{\beta})^T X^T X(\beta - \acute{\beta}) \quad (13)$$

Accordingly, the sample likelihood function in Equation (10) can be decomposed as follows:

$$L(\beta, \Sigma^{-1}) = \frac{|\Sigma^{-1}|^{(k+1)/2}}{(2\pi)^{mn/2}} \exp\left\{-\frac{1}{2}\text{tr}(\beta - \acute{\beta})^T X^T X(\beta - \acute{\beta})\Sigma^{-1}\right\} \cdot |\Sigma^{-1}|^{(n-k-1)/2} \exp\left(-\frac{1}{2}\text{tr}S\Sigma^{-1}\right) \quad (14)$$

It is easy to see that the first term on the right side of the above equation is when the precision matrix Σ^{-1} is given, the mean value is $\acute{\beta}$,the covariance matrix is$(X^T X)^{-1} \otimes \Sigma$ the kernel of the density function of the matrix normal distribution $N_{m \times (k+1)}(\acute{\beta}, (X^T X)^{-1} \otimes \Sigma)$,The second term is the kernel of the density function $W_m(n + m - k, S^{-1})$ of the Wishart distribution .Because the family of conjugate distributions of matrix normal distributions is matrix normal when the covariance matrix is known,The conjugate family of Wishart distributions is still Wishart .Therefore, the matrix normal - Wishart distribution is the joint conjugate prior distribution of the coefficient matrix β and the precision matrix Σ^{-1} in the multi-equation linear model system. It can be seen that there isa theoretical basis for choosing the following matrix normal-Wishart distribution as the conjugate prior distribution of the parameters

$$\pi(\beta, \Sigma^{-1}) = \pi(\beta|\Sigma^{-1})\pi(\Sigma^{-1}) \quad (15)$$

Of which

$$\pi(\beta|\Sigma^{-1}) = c_1(A, k, m)|\Sigma^{-1}|^{(k+1)/2} \exp\left\{-\frac{1}{2}\text{tr}(\beta - \mu_0)^T A(\beta - \mu_0)\Sigma^{-1}\right\}, \beta \in R^{(k+1) \times m}, \Sigma^{-1} > 0$$

$$\pi(\Sigma^{-1}) = c_2(D, n, m)|\Sigma^{-1}|^{(v-m-1)/2} \exp\left(-\frac{1}{2}\text{tr}D\Sigma^{-1}\right), \Sigma^{-1} > 0 \quad (16)$$

Where $c_1(A, k, m), c_2(D, n, m)$ are both regularization constant factors, A is a (k+1)×(k+1) positive definite matrix,μ_0 Is an $(k + 1)$ ×m matrix, D is an m ×m positive definite matrix, and ,v is a positive integer greater than m+1.

At this point, the problem of constructing the conjugate prior distribution of the parameters of the model system has been solved, in the following, the joint posterior distribution of the coefficient matrix β and the precision matrix Σ^{-1} is inferred according to Bayes theorem. In order to simplify the derivation process of relevant conclusions, the following two equations are proved first.

(1) note $(A + X^T X)^{-1}(A\mu_0 + X^T X\acute{\beta})$

$$C = (c_{ij})_{m \times m} \triangleq D + \mu_0^T A\mu_0 + \hat{\beta}^T \hat{\beta} + S - (A\mu_0 - X^T X\acute{\beta})^T \hat{\beta}$$

Than $D(\beta - \mu_0)^T A(\beta - \mu_0) + (Y - X\beta)^T (Y - X\beta) = (\beta - \acute{\beta})^T (A + X^T X)(\beta - \acute{\beta}) + C$ (17)

(2) $$\int_{\Sigma^{-1}>0} |\Sigma^{-1}|^{(n+v+k-m)/2} exp\left\{\frac{1}{2}\text{tr}\left[C + (\beta - \acute{\beta})^T (A + X^T X)(\beta - \acute{\beta})\right]\Sigma^{-1}\right\} d\Sigma^{-1} \propto \left|C + (\beta - \acute{\beta})^T (A + X^T X)(\beta - \acute{\beta})\right|^{-(n+v+k-m)/2} \quad (18)$$

According to the Bayes theorem, the density function of the posterior distribution of the parameters is proportional to the product of the sample likelihood function and the density function of the prior distribution of the parameters, based on the above conclusions in Equation (18), the density function of the joint posterior distribution of the coefficient matrix β and the precision matrix Σ^{-1} under the matrix normal - Wishart conjugate prior distribution in the form of Equation (15) is

$$\pi(\beta, \Sigma^{-1}|Y, X) \propto \pi(\beta, \Sigma^{-1}|Y, X)$$
$$\propto |\Sigma^{-1}|^{(n+v+k-m)/2} \exp\left\{-\frac{1}{2}\text{tr}[D + (\beta - \mu_0)^T A(\beta - \mu_0) + (Y - X\beta)^T (Y - X\beta)]\Sigma^{-1}\right\}$$
$$\propto |\Sigma^{-1}|^{(n+v+k-m)/2} \exp\left\{-\frac{1}{2}\text{tr}\left[C + (\beta - \acute{\beta})^T (A + X^T X)(\beta - \acute{\beta})\right]\Sigma^{-1}\right\} \quad (19)$$

Obviously, the joint posterior distribution of the coefficient matrix β and the precision matrix Σ^{-1} is still the matrix normal-Wishart distribution, but its distribution parameters are different from the prior distribution.

3. Expansion and Application

3.1. Application of Bayesian Methods to Other Regression Problems

Bayesian methods have been widely used in many regression problems, not limited to linear regression. For example, in polynomial regression, Bayesian methods can better capture the nonlinear features of the data by modeling higher-order terms. In addition, Bayesian methods have also been applied in nonlinear autoregressive models, such as ARMA models, to capture nonlinear dynamics in time series data. In the classification regression problem, the Bayesian method has also been applied, which can better capture the classification pattern of the data by transforming the classification target into a continuous variable and using the Bayesian method for modeling.

3.2. Limitations and Improvement Directions of Bayesian Methods in Practice

Despite the remarkable success of the Bayesian approach in many regression problems, there are some limitations in practice. Firstly, Bayesian methods typically require large amounts of data and computational resources for modeling, which may limit their application when data is scarce. Secondly, the high sensitivity of the Bayesian approach to model selection may lead to the problem of overfitting or underfitting. In order to solve these problems, the following improvement directions can be considered:

(1) Efficient algorithms: More efficient Bayesian method algorithms are investigated to reduce computational time and resource requirements.

(2) Model selection and tuning: By introducing more sophisticated model selection and tuning techniques, such as hyperparameter optimization and Bayesian inference, to improve the generalization ability of Bayesian methods.

(3) Combining other methods: Bayesian methods are combined with other statistical or machine learning methods, such as random forest, gradient boosting, etc., for more comprehensive data interpretation and prediction capabilities

(4) Consider prior knowledge: In practice, domain knowledge and prior information can be combined to select an appropriate prior distribution to improve the prediction accuracy and interpretation of the model.

Although the Bayesian method has some limitations in practice, through continuous research and improvement, it can be applied to more regression problems and further improve the prediction accuracy and generalization ability of the model.

4. Conclusion and Prospects

4.1. Summary of Key Research Findings

In this study, the application of Bayesian methods to linear regression was explored. By comparing the two methods of frequency linear regression and Bayesian linear regression, it is found that the Bayesian linear regression has the following advantages and characteristics: Bayesian linear regression can provide the probability distribution of output, rather than just a single point estimate, which is particularly important for small sample data sets; Bayesian method can introduce prior knowledge to improve the robustness and generalization ability of the model. Bayesian linear regression can handle noise and uncertainty better by using Bayesian inference methods for parameter estimation.

4.2. Research Limitations and Deficiencies

Although the Bayesian linear regression method has advantages in many aspects, it also has some limitations and shortcomings:

(1) The computational complexity of Bayesian methods is high, especially when dealing with large-scale data sets.

(2) Bayesian linear regression requires the selection of an appropriate prior distribution, which can be challenging for researchers unfamiliar with Bayesian statistics.

(3) Bayesian method requires more subjective judgment and domain knowledge for model selection and parameter adjustment

4.3. Suggestions for Future Research Directions

In view of the limitations and shortcomings of the above research, the following research directions and suggestions are proposed:

(1) Further research on the application of Bayesian methods to other types of problems, such as classification, clustering, etc.

(2) Conduct a comparative study between Bayesian method and other machine learning algorithms to determine the advantages and disadvantages of different algorithms.

(3) The application of dynamic Bayesian methods to regression problems is studied to improve the adaptability and flexibility of the model.

(4) Explore the application of Bayesian methods in nonlinear regression problems to improve the explanatory power and prediction accuracy of the model.

This study provides useful exploration and enlightenment for the application of Bayesian method in linear regression, and future research directions should focus on expanding and applying scope, and improving the flexibility and adaptability of the model.

References

[1] Zhu, H., Han, Y. (2001) Bayesian Statistical Inference of Multiple linear Regression Models. Statistics and Decision Making.

[2] Hu, Z., Yang, X. (2011) Bayesian Analysis of Linear Regression Model under NMAR Mechanism. Journal of Chuxiong Normal University.

[3] Yao, Y. (2010) Nonparametric Bayesian estimation of linear regression models with missing variables. East China University of Science and Technology.

[4] Zhu, H., Han, Y., Wu, Z. (2005) Bayesian prediction Analysis of Multiple linear Regression Model. Operations Research and Management.