

---

# BAYESIAN ASSIGNMENT

---

## Bayesian Linear Regression of University Admittance

### **Author**

Hidde van de Beek  
Utrecht University  
21-06-2023

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Dataset</b>	<b>3</b>
<b>3</b>	<b>Analysis</b>	<b>3</b>
3.1	Bayesian Model Averaging . . . . .	3
3.2	Parsimonious model . . . . .	4
3.2.1	Model Specification . . . . .	4
3.2.2	Convergence . . . . .	5
3.3	Results . . . . .	6
3.3.1	Bayesian Model Averaging . . . . .	6
3.3.2	Parsimonious model . . . . .	6
<b>4</b>	<b>Proof</b>	<b>7</b>
4.1	Synthetic data . . . . .	7
4.2	Bayesian Model Averaging . . . . .	7
4.3	MCMC Sampler . . . . .	8

# 1 Introduction

The aim of this paper is to identify the key factors that play a role in the university admittance process. In predicting the chance of admittance into university, it is crucial to understand the relative importance of each measure. Although different measures are designed to assess a student's abilities, it remains unclear to what extent each measure independently contributes to the overall decision of admittance. Traditional modelling approaches often focus on identifying one single best-fitting model, assuming that the single set of predictors can fully capture the complex relationship between the predetermined variables and outcome.

Bayesian Model Averaging (BMA), however, offers a more flexible framework and acknowledges that there may be multiple plausible models, each with its own set of predictors and estimates. Rather than selecting a single model, BMA considers a weighted average across a range of models, thus incorporating uncertainties of the underlying model structure into the analysis. This approach allows for a more nuanced understanding of the factors that influence the chances of admittance, using an existing dataset.

In addition to exploring the comprehensive model selection provided by BMA, this paper aims to further investigate a more parsimonious model using the best three predictors identified through BMA. Through the application of a subset of predictors, an optimized model can be acquired which captures the essence of the underlying relationships of the metrics. The selection of the three predictors are based on the posterior inclusion probabilities (PIP). These probabilities indicate the likelihood of each predictor being included in the model. The construction of a parsimonious model offers a simplified representation of the admittance process without sacrificing predictive accuracy.

## 2 Dataset

The data used in this article is a public data set called "Data for admission in the university" by Akshay Dattatray Khare on Kaggle [1]. It consists of 400 responses to several variables, including Graduate Record Examination (GRE) scores (out of 340), Test of English as a Foreign Language (TOEFL) scores (out of 120), University Rating (out of 5), statement of purpose strength (out of 5), letter of recommendation strength, undergraduate GPA (out of 10), research experience (either 0 or 1), and the chance of admittance (ranging from 0 to 1). There is no missing data and all continuous variables are normally distributed. To help the interpretation of comparing the importance of factors, all predictor will be standardized.

## 3 Analysis

### 3.1 Bayesian Model Averaging

For the BMA we will adopt a comprehensive approach by considering all possible subsets of predictor variables in our model space. This implies there is no pre-specification of models used during the analysis. First the Bayesian correlation between variables is calculated to look into the existing relationships of the data. To maintain a non-informative stance that might influence the relative importance of the predictors of university admittance, we

will choose an uninformative prior  $(\beta(1, 1))$  for model weights. This prior assigns equal prior weight to all possible models. We will then use Markov chain Monte Carlo (MCMC) methods to estimate the posterior model probabilities and produce weighted predictions using R [2] packages `BAS` and `BayesFactor`. The function will run 10,000,000 iterations to ensure convergence of the MCMC chains. We discard a burn-in period of 100,000 iterations to discard early transient behaviour, and then use the remaining samples to estimate the posterior probabilities of model inclusion and other relevant quantities of interest. Once we obtain the posterior probabilities of model inclusion, the weighted average predictions can be computed by combining the predictions from each model, using their respective probabilities as weights. The importance of variables is determined by the PIPs.

## 3.2 Parsimonious model

### 3.2.1 Model Specification

The three predictors with the highest PIP are chosen for the parsimonious analysis. To calculate the Bayesian Regression Analysis we will use MCMC methods using Gibbs sampling, since the joint posterior distribution can be factorized into conditional distributions. Each parameter is then sampled from its own conditional distribution given the values of the other parameters. The regression parameters  $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ , and the precision  $\tau$  are estimated using non-informative conjugate priors. For  $\beta_0 - \beta_4$  a normal distribution with  $\mu = 0$  and  $\sigma = 0.001$  were used. For the prior variance of the outcome variable  $Y$  a gamma distribution with  $\alpha = 0.001$  and  $\beta = 0.001$  was used. Theorem 3.1 for the conditional distributions can be derived using these priors.

For the MCMC sampler, three chains with different starting values are set-up to probe the parameter space of the conditional distributions. The sampler applies a burn-in period of 10,000 iterations, allowing the chains to converge to the parameter estimates. After the burn-in, an additional 90,000 samples are collected from the posterior distribution for each parameter of interest. This iterative process produces reliable estimates of the regression parameters using the Gibbs sampling framework with conjugate priors.

**Theorem 3.1.**

$Y_i = b_0 + b_1x_{1i} + b_2x_{2i} + b_3x_{3i} + e_i$  with  $e_i$  i.i.d.  $\sim N(0, 1/\tau)$  and priors

$b_0 \dots b_3 \sim N(0, 1/\tau_b)$

$\tau \sim \text{gamma}(\alpha, \beta)$

**which result in the conditional distributions**

$f(b_0|b_1, b_2, b_3, \tau, Y_1, \dots, Y_n) \sim N\left(\frac{\tau}{n\tau + \tau_b} \sum_{i=1}^n (Y_i - (b_1x_{1i} + b_2x_{2i} + b_3x_{3i})), \frac{1}{n\tau + \tau_b}\right)$

$f(b_1|b_0, b_2, b_3, \tau, Y_1, \dots, Y_n) \sim N\left(\frac{\tau \sum_{i=1}^n (Y_i - (b_0 + b_2x_{2i} + b_3x_{3i})x_{1i})}{\tau \sum_{i=1}^n (x_{1i}^2 + \tau)}, \frac{1}{\tau \sum_{i=1}^n x_{1i}^2 + \tau_b}\right)$

...

$f(\tau|b_0, b_1, b_2, b_3, Y_1, \dots, Y_n) \sim \text{gamma}\left(\alpha + n/2, \beta + \frac{1}{2} \sum_{i=1}^n (Y_i - (b_0 + b_1x_{1i} + b_2x_{2i} + b_3x_{3i}))^2\right)$

### 3.2.2 Convergence

To obtain proper estimates it is important for the Gibbs Sampler to converge, since non-convergence will lead to poor and biased parameters. This is checked using several methods: trace plots, autocorrelation plots, Gelman Rubin Statistics and MCMC error. Trace plots and autocorrelation plots are generated to visualize the convergence and mixing properties of the MCMC sampler for each model parameter. The Gelman-Rubin Statistic is computed to measure convergence by comparing between-chain variance to within-chain variance. MCMC error is calculated to check if it is less than 5% of the parameter's standard deviation, ensuring reliable estimates. These diagnostic measures assess the performance of the analysis, and verify the parameter estimates in the Bayesian analysis.

In Figure 1 the sampled values of the parameters visualize convergence to a common region and do not exhibit extreme fluctuations. It also doesn't appear to get stuck in a local spot. A rapid decay of autocorrelation is observed Figure 2, indicating the decrease of correlation between consecutive samples. This indicates that the chains are efficiently exploring the parameter space. We also find that the Gelman-Rubin Statistics are 0.999, 0.999, 1.000, 1.000, 1.00, 0.999 for  $\beta_0 - \tau$ , respectively., indicating convergence. Finally, the MCMC error for each parameter is compute and is well below the threshold of 5% for all parameters, suggesting that the parameter estimates are stable and accurate.

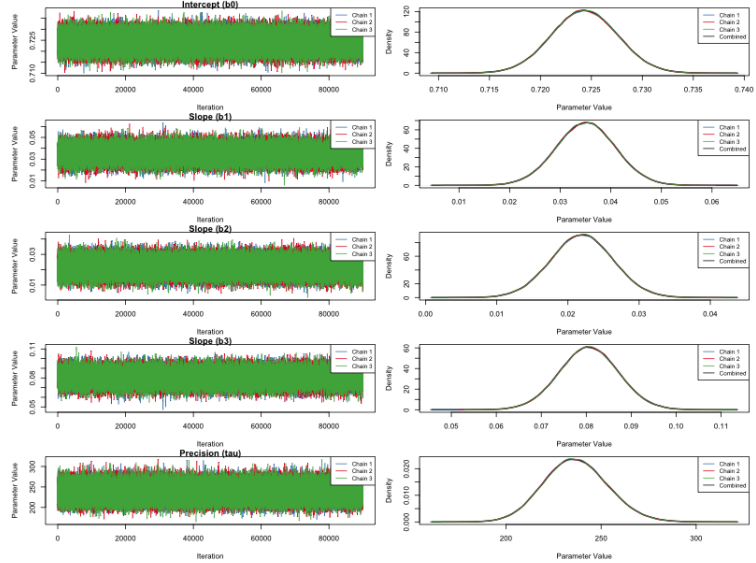


Figure 1: Trace plots and posterior distributions of the standardized 3 predictor model

In Figure 1 the sampled values of the parameters visualize convergence to a common region and do not exhibit extreme fluctuations. It also doesn't appear to get stuck in a local spot. A rapid decay of autocorrelation is observed Figure 2, indicating the decrease of correlation between consecutive samples. This indicates that the chains are efficiently exploring the parameter space. We also find that the Gelman-Rubin Statistics are 0.999, 0.999, 1.000, 1.000, 1.00, 0.999 for  $\beta_0 - \tau$ , respectively., indicating convergence. Finally, the MCMC error for each parameter is compute and is well below the threshold of 5% for all parameters, suggesting that the parameter estimates are stable and accurate.

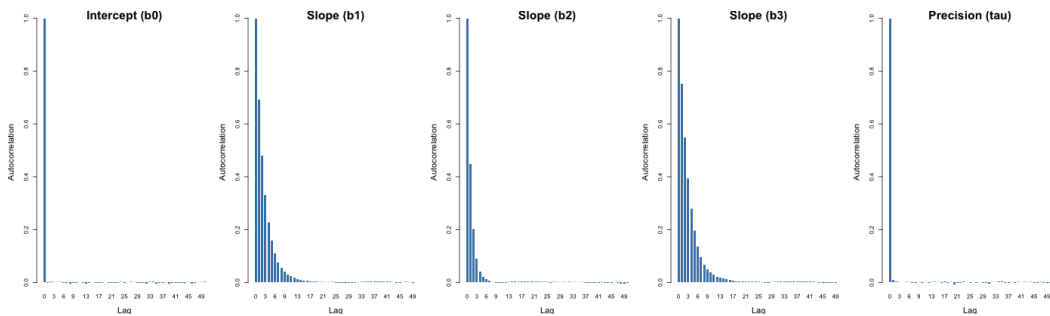


Figure 2: Autocorrelation plots of the standardized 3 predictor model

### 3.3 Results

#### 3.3.1 Bayesian Model Averaging

Bayesian correlation coefficients are estimated between the predictor variables and response variable. The results indicate a positive correlation with GRE score (95% CI = .80 - .86), TOEFL Score ( $r = .79$ , 95% CI = .75 - .83), University Rating ( $r = .71$ , 95% CI = .66 - .75), Statement of Purpose ( $r = .67$ , 95% CI = .62 - .73), Letter of Recommendation ( $r = .67$ , 95% CI = .61 - .72), undergraduate GPA ( $r = .87$ , 95% CI = .85 - .90), and Research Experience ( $r = .55$ , 95% CI = .49 - .62). These results indicate our data set can be used to investigate our research question.

The PIPs suggest that CGPA (PIP = 1.000,  $\beta = 0.123$ ) and Letter of Recommendation (PIP = .999,  $\beta = .023$ ) have the highest probabilities of being included in the predictive model, because they display strong evidence of their association with the response variable. GRE score and Research experience also shows a high PIP (PIP = .919,  $\beta = .002$  and PIP = .910,  $\beta = .022$ , respectively), indicating potential contribution to the model. TOEFL score also has a relative high PIP (PIP = .836,  $\beta = .003$ ) and should be taken into consideration as an important factor. University rating and Statement of Purpose have lower PIPs (PIP = .244,  $\beta = .001$  and PIP = .163,  $\beta = .000$ , respectively), indicating weaker evidence. The flexibility of BMA shows an insight into influential predictors that can guide future research and inform predictive modeling efforts. It is important to note the inherent uncertainty associated with the estimates and the need for validation in independent datasets.

#### 3.3.2 Parsimonious model

The three highest PIP's are used in the parsimonious model: undergraduate GPA, Letter of Recommendation, and GRE score. Parameter estimates and credible intervals using the Gibbs sampling are calculated and are displayed in Table 1. The largest parameter estimate is CGPA (beta = 0.135, CI = 0.118-0.153), then LOR (beta = 0.024, CI = 0.015-0.034), and finally GRE (beta = 0.003, CI = 0.002-0.004). The credible intervals in the tables represent the 95% range of my belief what the regression coefficients, given both on the data and uninformative priors. The regression coefficients are standardized, meaning the magnitude of the coefficient is comparable. This means there is evidence that key factor for University Admittance is undergraduate GPA in the simplified model. Furthermore, we see that the CI for undergraduate GPA does not overlap with the CI for GRE score or Letter of Recommendation, indicating that it is clearly the most important factor for a succesful application. Next to overall better scores, this be an important conclusion for students seeking for better approval rates: The entire undergraduate program and according grades play an important role in eventual application.

Table 1: Summary of the Sampled Parameter Values in the Parsimonious Model.

Parameter	Mean	SD	Naive SE	2.5%	Median	97.5%
intercept	0.7243	0.0033	0.0000	0.7180	0.7243	0.7307
CGPA	0.0806	0.0066	0.0000	0.0677	0.0806	0.0935
LOR	0.0220	0.0044	0.0000	0.0134	0.0220	0.0306
GRE	0.0351	0.0059	0.0000	0.0235	0.0351	0.0466
tau	236.7546	16.8571	0.0324	204.9397	236.3438	270.8541

## 4 Proof

### 4.1 Synthetic data

As a proof of concept for the BMA and Gibbs sampler we created two synthetic data sets of 10,000 observations of 6 variables which can represent different aspects or measurements. In the first data set the outcome variable is perfectly predicted by all predictors. The Directed Acyclic Graph (DAG) in figure 3b illustrates the relationships of these generated variables. We can see that variable F is perfectly predicted by all other variables. In the second data set there are no true relations. This is shown in the DAG in figure 3a, where the absence of arrows signifies the lack of direct relationships. In both data sets the truth about the relationships is known, thus the effectiveness of the research methods can be analysed. The descriptive statistics are shown in table 2 and show similar characteristics.

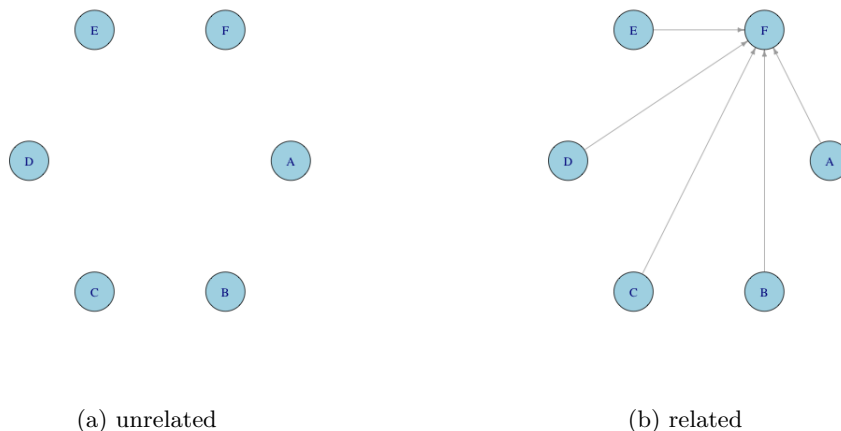


Figure 3: Directed Acyclic Graph for synthetic data

Table 2: Descriptive Statistics of the synthetic data

	Dataset related				Dataset unrelated			
	Mean	SD	Min	Max	Mean	SD	Min	Max
A	6.01	2.46	-3.11	15.20	6.01	2.49	-4.01	14.85
B	30.05	10.05	-6.26	70.00	30.14	10.06	14.81	70.07
C	99.94	14.99	45.98	155.50	99.98	15.04	45.17	155.37
D	60.04	14.94	4.44	117.75	60.03	15.00	3.18	116.33
E	1.00	0.20	0.31	1.71	1.00	0.20	0.23	1.79
F	49.84	22.02	-39.61	133.17	50.04	22.19	-55.95	139.53

### 4.2 Bayesian Model Averaging

To show BMA works properly, it needs to give the same conclusion as the true relationships in the generated data: PIPs need to be 1 in the related model and 0 in the unrelated model. To compute PIPs, the marginal posterior distribution of each parameter is normalized and divided by the sum of the marginal posterior distributions across all parameters. The PIPs

represent the probabilities of each parameter being included in the true model, given the data and the model set. An uninformative prior ( $\beta(1, 1)$ ) is chosen for model weights and assigns equal prior weight to all possible models. The parameter PIPs are then calculated using the R package BAS and show the right PIPs in the unrelated and related dataset (PIPs = 0 and PIPs = 1, respectively). The weighted average is now the value that was used in the data generating mechanism; i.e. the true value.

### 4.3 MCMC Sampler

To show that the Gibbs sampling algorithm is working correctly for the derivation of the conditional posterior distributions, it needs to converge to the true values in the data set. To estimate the parameters we set priors  $\beta_0 - \beta_4$  to a normal distribution with  $\mu = 0$  and  $\sigma = 0.001$ . For the variance of  $Y$  a gamma distribution with  $\alpha = 0.001$  and  $\beta = 0.001$  is used. The trace plots show intermixing of chains, autocorrelation plots are acceptable, Gelman-Rubin Statistics are all close to 1 and MCMC error are also in order, thus giving no proof for non-convergence. Figure 4 shows that the estimates approach the true values and can thus conclude the MCMC sampler is working correctly.

## References

- [1] A. D. Khare, "Data for admission in the university," 2022.
- [2] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2022. [Online]. Available: <https://www.R-project.org/>

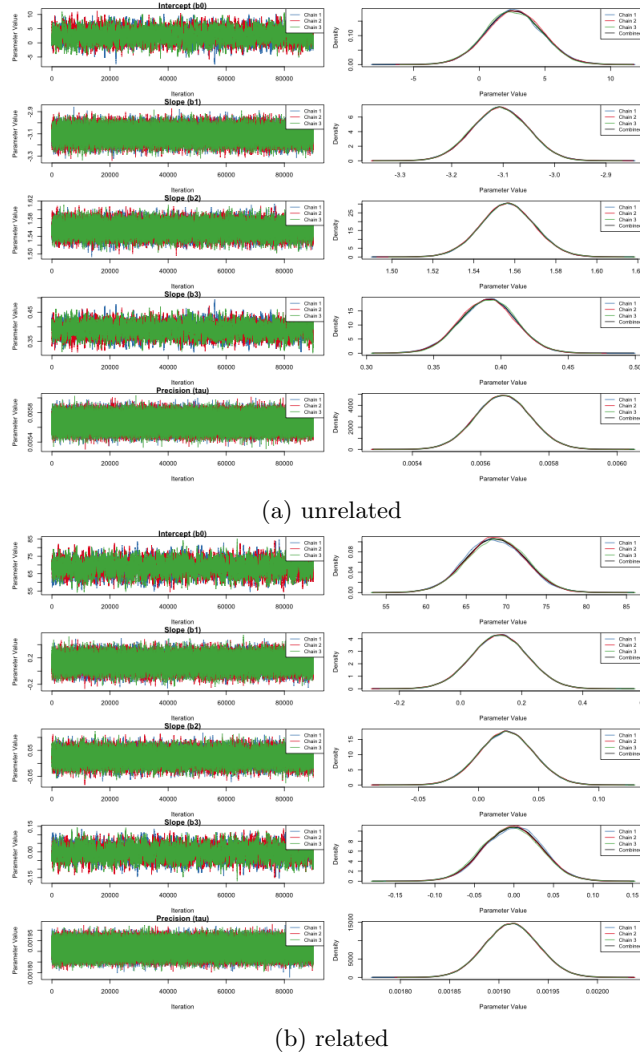


Figure 4: Convergence of synthetic data