Stats 202C: Final Project
Bayesian Mortality Prediction
with Sensitivity Analysis

Ryan O'Dell

June 11, 2023

1 Introduction

Patients admitted to the hospital would ideally be returned to good health and discharged in an appropriately timely manner. To make such a decision, it would be reasonable to desire a decision rule on when to discharge a patient based on their predicted likelihood to die.

In this case, we reviewed a dataset from a 2015 study from the Institute of Cardiology and Allied Hospital in Pakistan where 299 patients were admitted with heart failure and roughly 1/3 of the patients had died post discharge. Certain medical history measurements were provided for each patient in the form of age, sex, smoking status, diabetes status, and anaemia status. Along with this, measurements of the patients' ejection fraction, serum creatinine, blood pressure, serum sodium, platelets, and creatinine phosphokinase were also included. For some background, serum creatinine was used as a measurement of kidney performance and ejection fraction was used to measure heart performance. In this paper we will focus on age, smoking status, diabetes status, anaemia status, ejection fraction, and serum creatinine as those were identified as significant in the reviewed literature.

We reviewed two papers, "Survival analysis of heart failure patients: A case study", Ahmad et al. (2017) and "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone", Chicco et al. (2020) who had also taken a look at this data set. Ahmad et al. had applied cox regression proportional hazard model and used out of sample AUC in order to measure the model. They found that age, serum creatinine, blood pressure, ejection fraction, and anaemia were significant factors and their model had an out of sample AUC of 0.81. Chicco et al. had applied a large variety of machine learning methods including Random Forest, SVM, XGBOOST, etc. However, they found that the Random Forest did the best with

an out of Sample AUC of around 0.80 and identified that serum creatinine and ejection fraction were significant predictors of mortality.

Based on the above readings, we had noted the large variation in methodological approaches and resulting difference in model parameters selected. We decided to take a Bayesian approach to the dataset and apply Markov Chain Monte Carlo (MCMC) methods to analyze mortality predictions. In order to compare our models, we decided to use out of sample AUC as the other authors also used this approach.

2 Methodology

We wish to consider two goals in our analysis. One is the data analysis question of predicting the patients' mortality rates given health record data with the covariates described above. To accomplish this, we can implement a Bayesian logistic regression, where the outcome variable y we wish to predict is death ($y = 1$) versus no death ($y = 0$). The other, more nuanced question we have is related to the specifications of the Bayesian regression: How might one's judgments and assessments of model uncertainty differ when the utilized prior distribution for the regression coefficients changes?

We may measure this uncertainty using a sensitivity analysis on the different priors. In particular, we look at three different choices of priors: (1) the uninformative/uniform, (2) a normal prior with $\mu = 0$ and $\sigma = 5$, and (3) a double exponential (Laplace) prior with $\mu = 0$ and $\sigma = 1$. We modeled the prior distribution of each regression coefficient was independent. The intuition behind our choices of the normal and Laplace priors is that they correspond to the Ridge and LASSO regression, respectively, so it seemed appropriate to investigate these contexts. In our comparison, we will look at the posterior distributions of the parameters as well as the predictive posterior distributions for each of the three priors. To measure out-of-sample performance later, we will also need to split the data into a train set and a test set; we choose to keep 75% of the data for training and the remaining 25% for testing.

To carry out the Bayesian logistic regression, we utilize MCMC sampling methods. We believe that this is a good use case for MCMC sampling since it is well known that there does not exist a conjugate prior for Bayesian logistic regression. In particular, we try three different MCMC samplers; for each sampler, we initialize the parameters at their maximum likelihood estimates. First, we implement the block Metropolis-Hastings algorithm but this proves to be problematic. Some of the covariates - namely, those deemed insignificant either by existing papers or by the MLE estimates themselves - fail to update at all, even after approximately 1 million iterations of Metropolis-Hastings. Next we can look at the full Metropolis-Hastings but here we come across numerical issues when drawing the full vector of coefficients. This implementation similarly fails to provide updates, since we end up with very restrictive acceptance criteria for the sampler.

Lastly, we introduce a Hamiltonian Monte Carlo (HMC) which solves our problems with updating. After setting a sufficiently small step size and a large number of leap frog iterations (0.00025 and 250, respectively), the HMC sampler achieved stationarity for all three models and all regression coefficients. We can see for instance in the trace plots for the first 50,000 iterations post 10,000 burn in iterations (see Appendix Figures 2-4) that the HMC sampler displays good convergence behavior, with all of the parameters converging to a stationary distribution very rapidly.

3 Results

3.1 Inference

To make inference about our model parameters in a Bayesian framework we can inspect the marginal posterior distributions of each coefficient. We can make sure the posterior distributions are not multimodal to ensure the region of highest probability / density is not misleading. A visual inspection of the histograms from our HMC sampler shows that indeed all the marginal posteriors are unimodal (Appendix Figures 5-7).

A straightforward Bayesian way to decide which variables are significant is examining the 95% credible intervals. If the 95% credible intervals contain 0 we can conclude that the variable is not significantly impacting mortality. When we examine the 95% credible intervals for each of the models they all identify the same covariates as significant (Table 1). The only variables that our models identify as statistically significant are Age, Ejection Fraction, and Serum Creatinine, suggesting that our results were slightly different from the existing literature. Considering all our models agree on which variables are significant, it appears that the functional form of our prior distributions has minimal impact on identifying which variables are significant predictors of mortality. For a summary of the posterior means of each model please see Table 3 in the appendix.

Posterior Credible Intervals

Coefficients	Normal Prior	Uniform Prior	Laplace Prior
Intercept	[-4.90, -0.72]	[-4.9, -0.72]	[-3.57, 0.13]
Age	[0.02, 0.08]	[0.020, 0.082]	[0.01, 0.07]
Anemia	[-0.23, 1.08]	[-0.23, 1.08]	[-0.32, 0.86]
Diabetes	[-0.42, 0.88]	[-0.42, 0.88]	[-0.42, 0.70]
Ejection Fraction	[-0.11, -0.04]	[-0.11, -0.04]	[-0.11, -0.05]
Serum Creatinine	[0.37, 1.25]	[0.37, 1.25]	[0.26, 1.12]
Smoking	[-0.77, 0.68]	[-0.77, 0.68]	[-0.70, 0.48]

Table 1: Posterior Credible Intervals.

The most striking result is that both smoking and diabetes were not significant predictors of

mortality and that this is common to all three models we used. The existing literature hypothesizes that this is due to the patients already experiencing some form of heart failure when they entered the hospital. Other possible explanations for this result could be the observational nature of the data itself and issues surrounding sampling. Caution should be taken when trying to generalize these results to a population of individuals outside patients who have already suffered some degree of heart failure.

3.2 Prediction

As part of validation of any statistical model, especially regression models, it is often sensible to ensure that a model makes sensible predictions. To that end, we examine the model performance on the testing data set for each model. For predictive inference in Bayesian regression we can use the posterior predictive distribution to summarize the uncertainty in our predictions. The posterior predictive distribution can be obtained by evaluating the following integral:

$$P(Y_{new} = 1|X_{new}, X, y) = \int P(Y_{new} = 1|X_{new}, \theta)P(\theta|X, y)d\theta$$

The term on the left of the integral is simply the likelihood and the term on the right is the posterior distribution. We can simulate this distribution by doing the following procedure.

1. Once we have the samples of our posterior distribution we can generate the probability of death by plugging the samples into the link function.

2. Then we can simulate a Bernoulli random variable with this probability and repeat for each sample in the posterior distribution.

The end result is a distribution of possible outcomes for each individual. We apply this technique to all the individuals in the testing data set. For model performance, we take the means of the posterior predictive distributions and use AUC as our classification metric to have direct comparisons with the existing literature (Table 2). The performance of the models with normal and unif priors are essentially identical, while the model with Laplace priors has the highest out of sample AUC of 0.81. This is noteworthy because the highest AUC reported in both papers was also 0.81.

Below (Figure 1) is the three predictive distributions for an individual in the test set who was lived upon release. The predictive distribution provides a concrete assessment of how likely the patient is to pass away upon release. In particular, in the the posterior predictive distribution is interpreted as an actual probability as opposed to a confidence interval. For the individual below, we see that all the densities are concentrated between $0.1 - 0.2$, indicating that the individual is unlikely to die upon release. This could assist physicians and nurses when deciding if a patient is

Predictive Performance

	Normal	Uniform	Laplace
AUC	0.801	0.795	0.810

Table 2: Out of Sample AUC.

safe to discharge from the hospital. Additionally, we see that the models with uniform and normal priors have identical predictive distributions. Indicating that our priors have minimal impact on our assessments of predictive uncertainty.

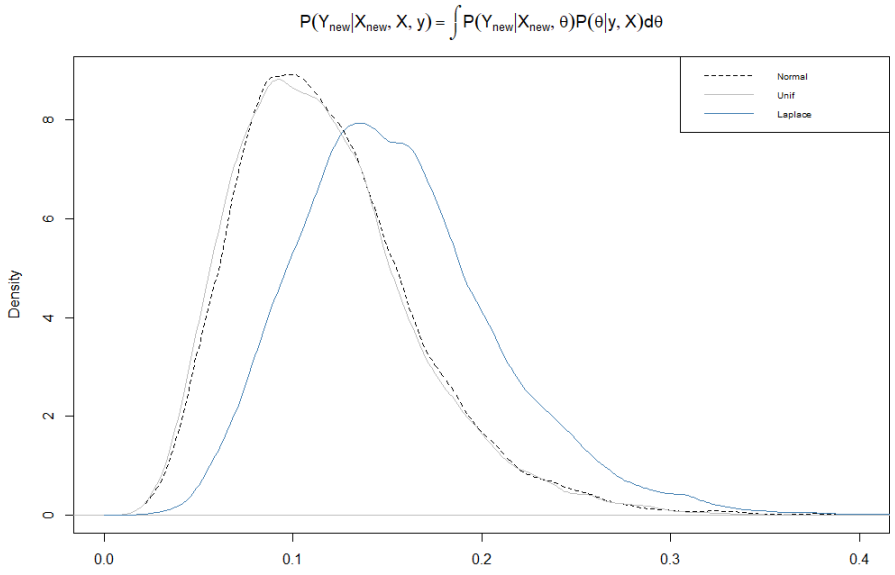

Figure 1: Posterior predictive distributions for an individual who lived post discharge.

4 Conclusion

Our goal was to take a bayesian approach to predict mortality outcomes for discharged heart failure patients using MCMC methods. We were able to achieve a model with an AUC of about 0.81 with only three predictors, which is a comparable result to those presented by published papers. To ensure our models were robust, we performed a sensitivity analysis and saw that the choice of prior had minimal impact on the inference and overall predictive performance.

Although sensitivity analysis is useful for ensuring model robustness, it does not provide a direct method for model selection. In the future, we may want to explore model choice uncertainty via a Bayesian Model Averaging as our method had chosen different predictors than that of the other

two papers in order to appropriately weight models by their posterior probabilities.

5 References

1. Ahmad T, Munir A, Bhatti SH, Aftab M, Raza MA (2017) Survival analysis of heart failure patients: A case study. PLoS ONE 12(7): e0181001. https://doi.org/10.1371/journal.pone.0181001

2. Chicco, D., Jurman, G. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. BMC Med Inform Decis Mak 20, 16 (2020). https://doi.org/10.1186/s12911-020-1023-5

6 Appendix And Figures

Summary of Posterior Means

Coefficients	Normal Prior	Uniform Prior	Laplace Prior
Intercept	-2.78 (1.05)	-2.95 (1.06)	-1.42 (0.95)
Age	0.05 (0.01)	0.05 (0.01)	0.04 (0.01)
Anemia	0.41 (0.33)	0.43 (0.34)	0.24 (0.30)
Diabetes	0.24 (0.34)	0.26 (0.35)	0.12 (0.28)
Ejection Fraction	-0.07 (0.02)	-0.071(0.02)	-0.08(0.02)
Serum Creatinine	0.79 (0.22)	0.80 (0.23)	0.67(0.22)
Smoking	-0.048 (0.36)	-0.01(0.36)	-0.10 (0.29)

Table 3: Standard Deviations in Parenthesis.

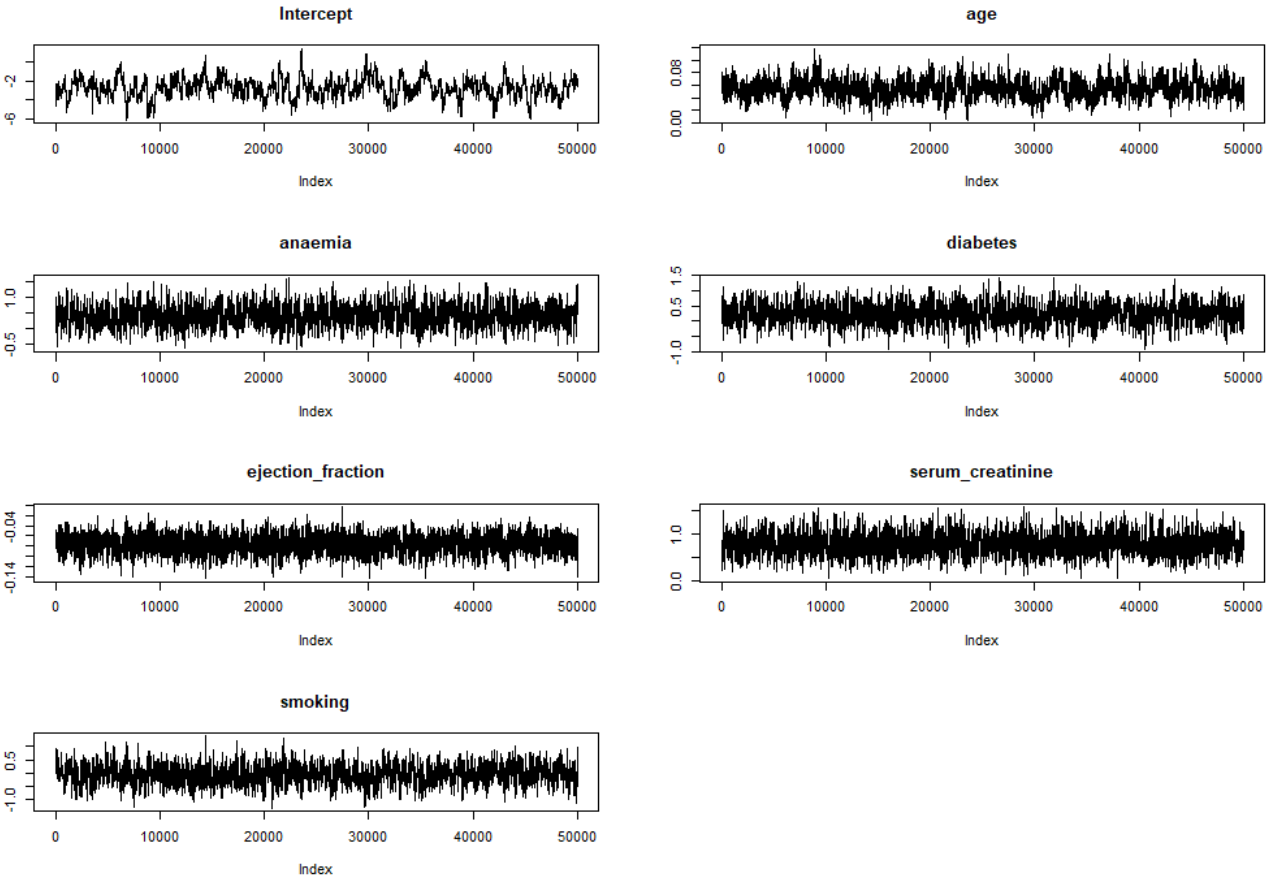

Figure 2: Trace Plots of HMC sampler for model with Normal Priors.

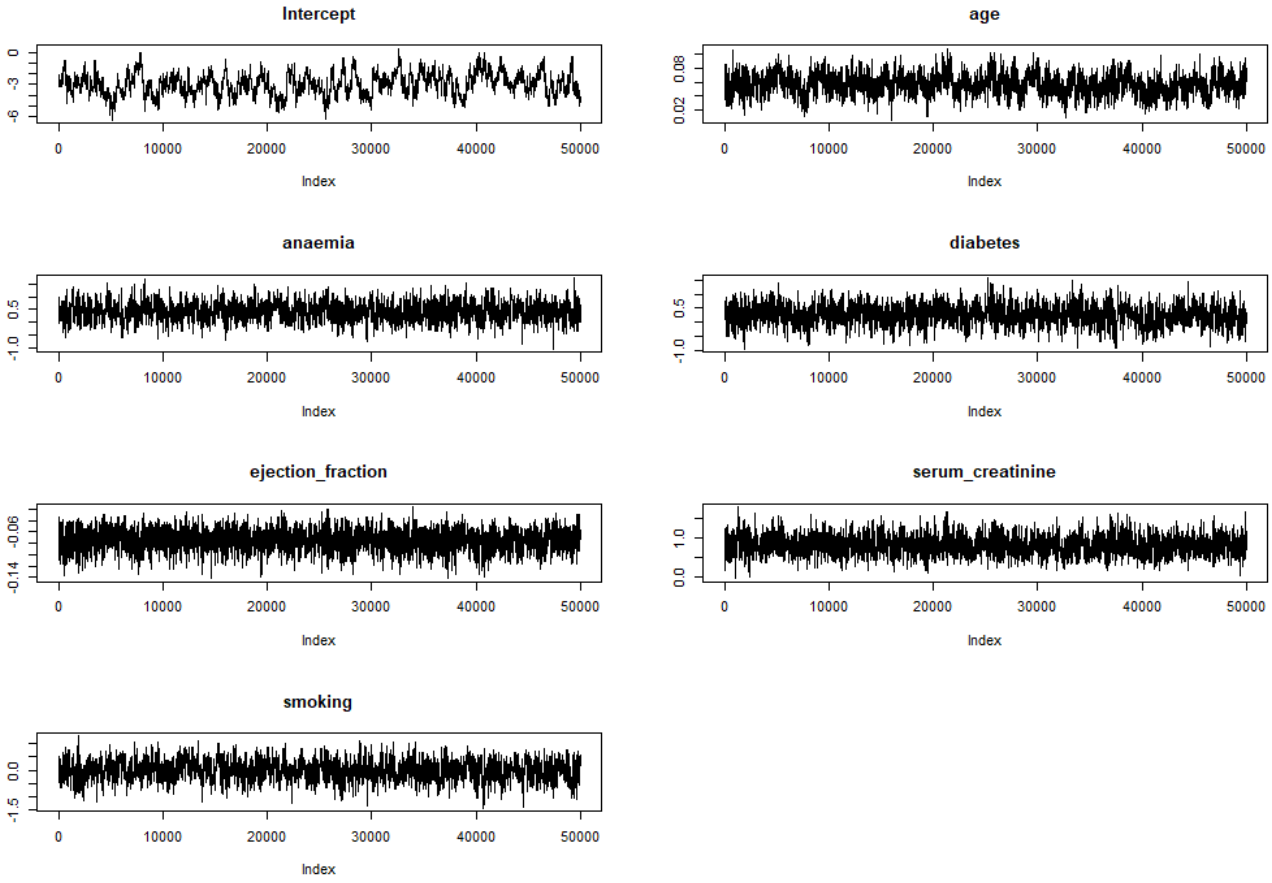

Figure 3: Trace Plots of HMC sampler for model with Uniform Priors.

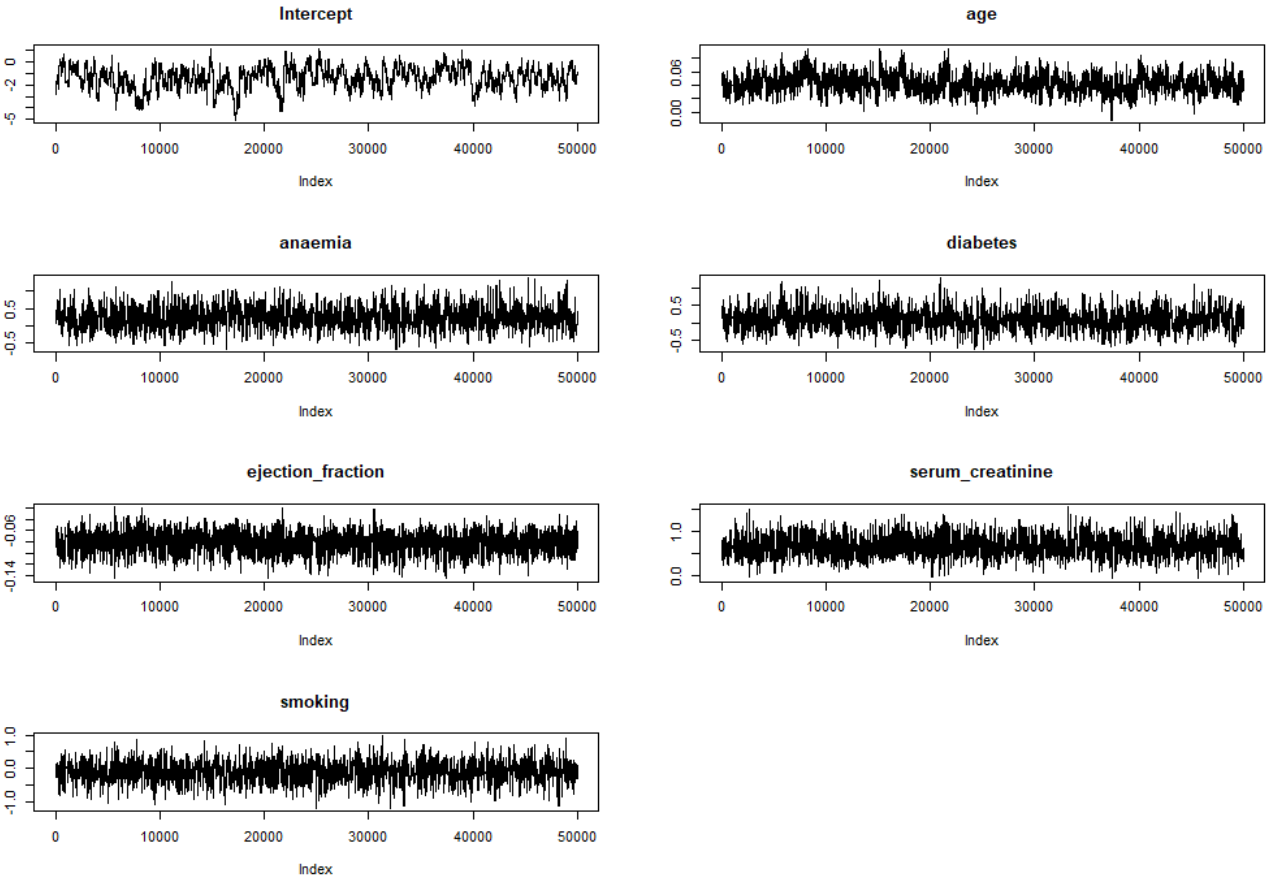

Figure 4: Trace Plots of HMC sampler for model with Laplace Priors.

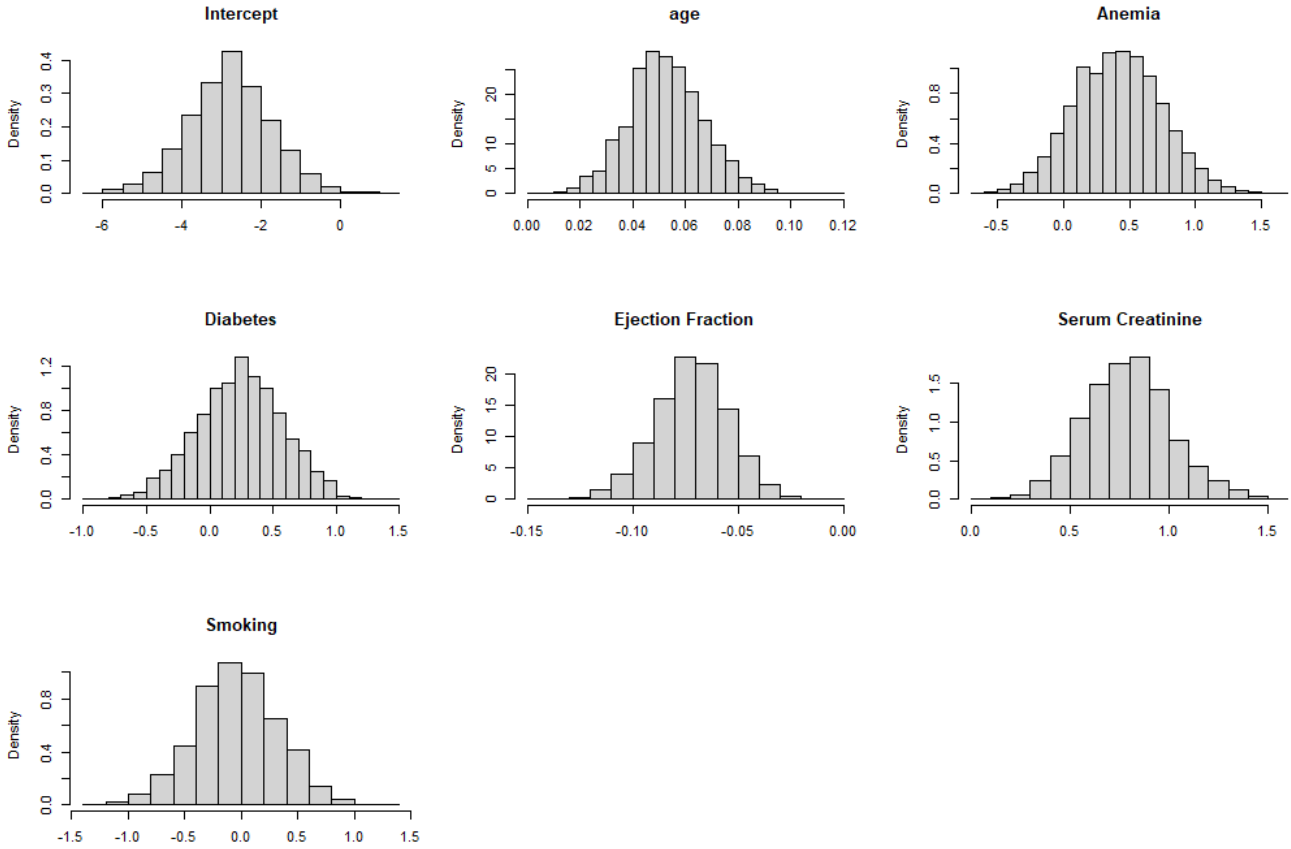

Figure 5: Marginal Posterior Densities of the regression coefficients for model with Normal Priors.

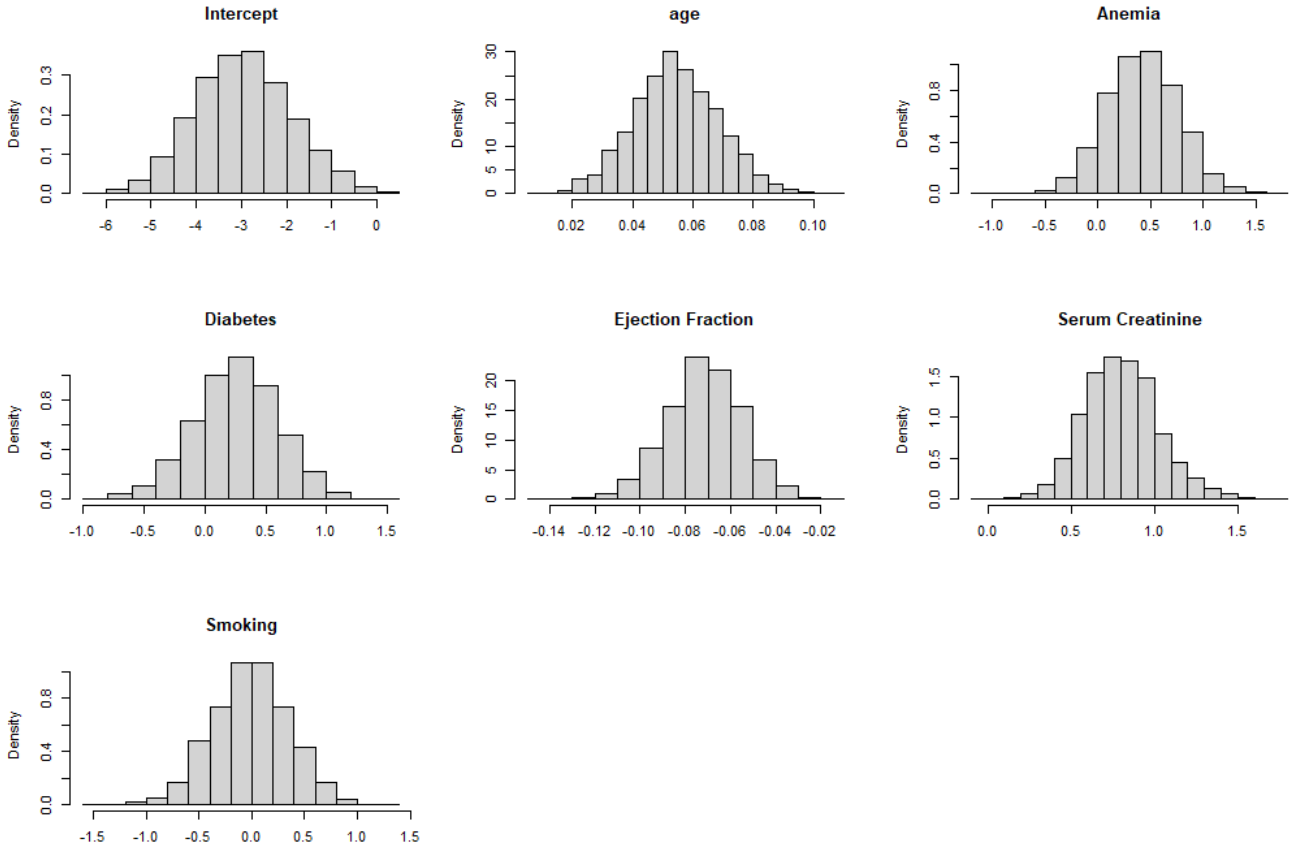

Figure 6: Marginal Posterior Densities of the regression coefficients for model with Uniform Priors.

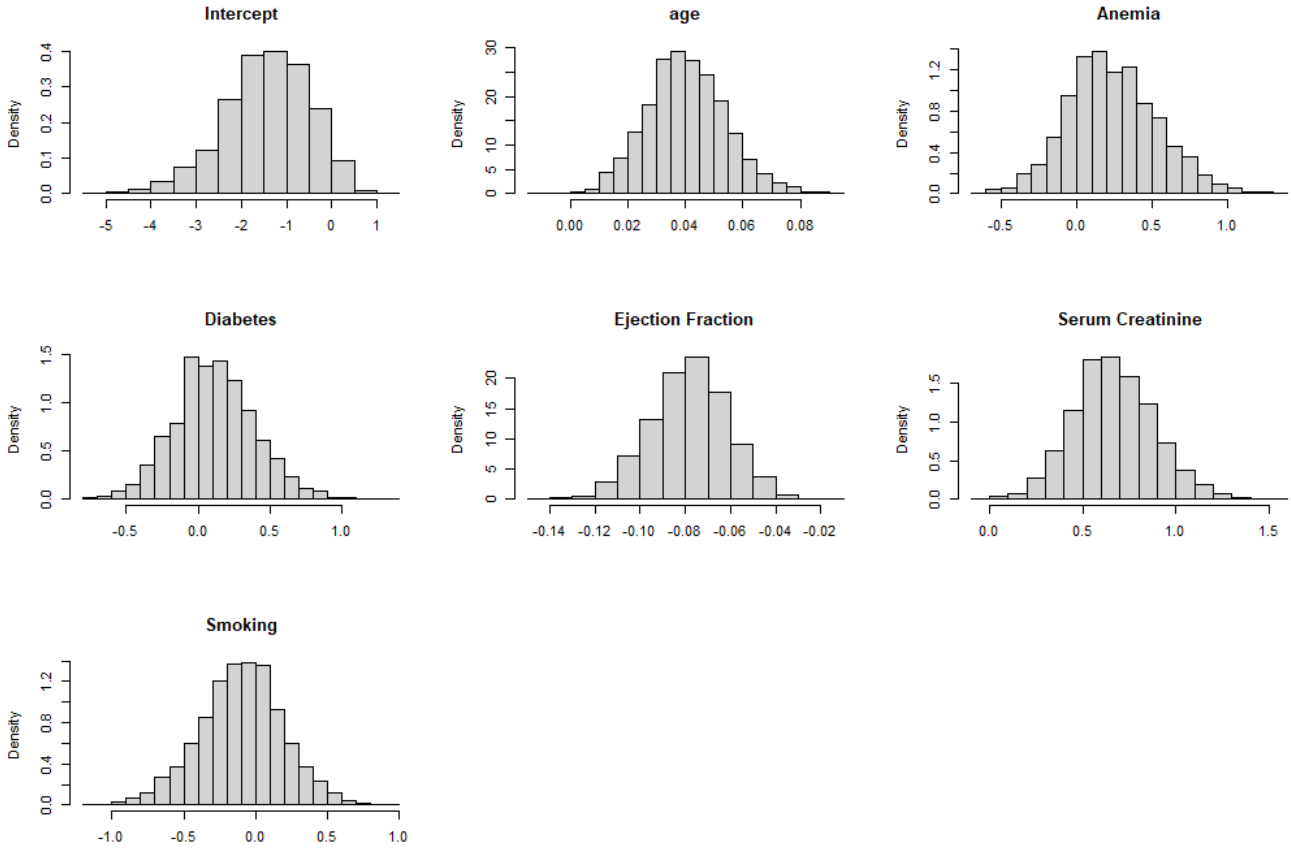

Figure 7: Marginal Posterior Densities of the regression coefficients for model with Laplace Priors.