Bayesian Learning and Monte Carlo Simulation Project

Awad Joy,
Basso Paolo,
Cavalli Dario,
Cimpeanu Vlad

POLITECNICO
MILANO 1863

Contents

1 The problem and the data

We have a dataset regarding the "Vinho Verde" wine. In this dataset some physical measurements taken from bottles of wine are combined with a sensory judgement about the quality (a vote 0 to 10) of the wine itself.

The objective is to determine which factors influence the quality wine and to use the physical measurements to correctly classify the wines in their quality category. We have the following physical measurements:

- fixed acidity
- volatile acidity
- citric acid
- residual sugar
- chlorides
- free sulfur dioxide
- total sulfur dioxide
- density
- pH
- sulphates
- alcohol

The problem we are facing is an ordinal regression problem, in fact, even though the goal is to assign each wine to a specific category these categories have an order. Common regression problem differs to ordinal regression for the domain of the target, indeed, in this problem we have a restricted discrete domain of values for the target quality, instead of having a continuous domain.

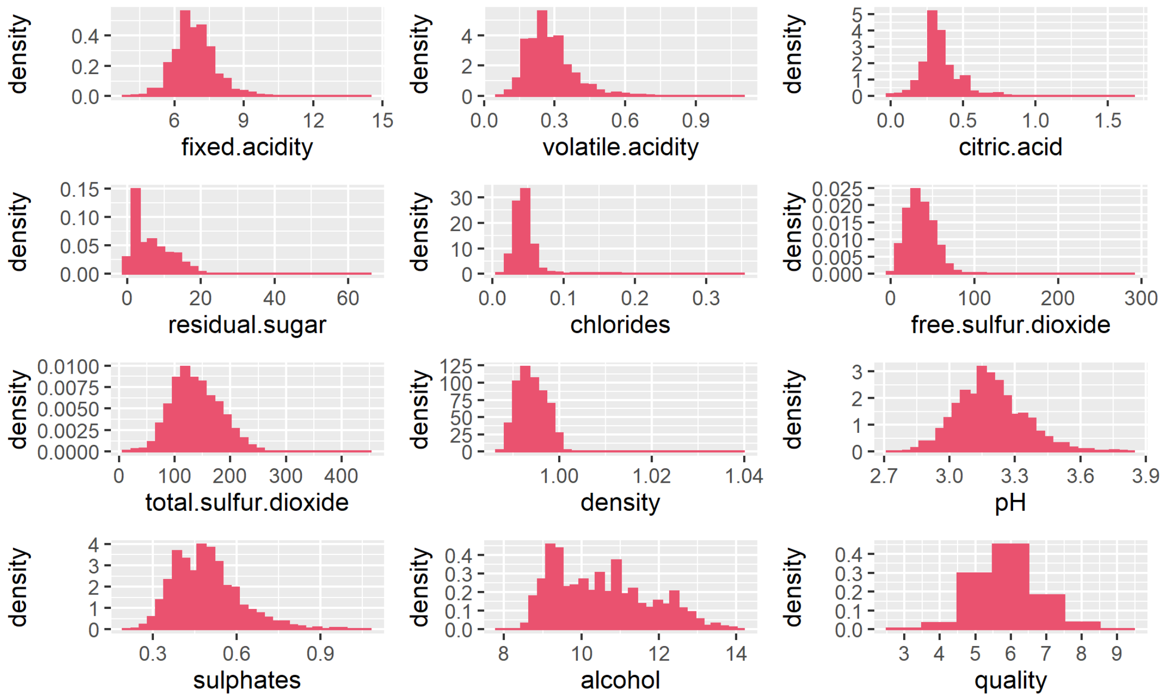

Figure 1: data distribution

2 Analysis of the dataset

For this problem we consider the feature quality as a response variable, whereas the remaining ones as explanatory variables.
In this section we try to make an exploratory data analysis to better understand the nature of our data. By plotting the histograms of the data (Figure 1) we notice all features show positive values, which make sense since, a part from quality, they all represent a physical volume measure. Given the explanatory variables nature we can claim all of them are continuous variables, whereas the quality is clearly a discrete variable. Because explanatory variables are continuous we decided to standardize them to achieve better results with our models.
Quality has most values concentrated in the categories 5, 6 and 7. Only a small proportion is in the categories [3, 4] and [8, 9] and none in the categories [0, 1, 2] and 10, so the dataset is unbalanced and may give some problems when predicting uncommon categories.

2.1 Outliers analysis

In order to determine either a feature is an outlier or not, we follow John Turkey's indications, who states a point is an outlier if it is outside the following range:

$$[Q_1 - k(Q_3 - Q_1), Q_3 + k(Q_3 - Q_1)]$$

where Q_1 is the first quartile, Q_3 is the third quartile and k is set to 1.5.
From the box plots (Figure 2) is clear that all the features but alcohol have outliers. We are not able to identify which is the nature of these outliers, we cannot say if these outliers are the fruit of errors in the measurement process or if they are just rare but possible values so we can not remove them. One reason for the existence of outliers might be the unbalanced dataset, thus, the outliers may be related to rare characteristics of the wine.

One may argue that the outliers are somehow correlated to high quality or low quality wines, since these categories occupy only a small portion of the dataset. For this reason we want to investigate if

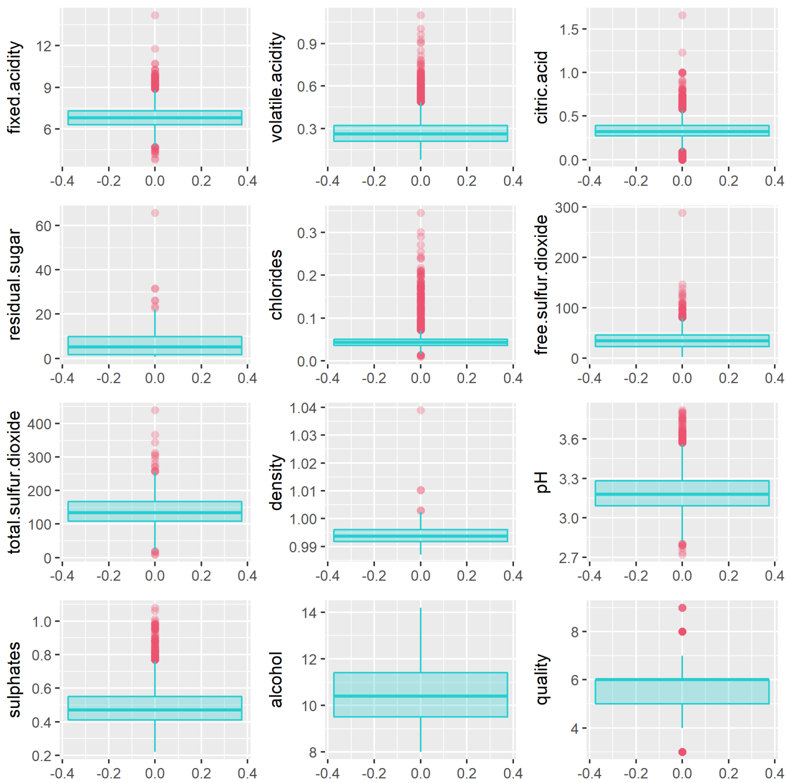

Figure 2: Outliers

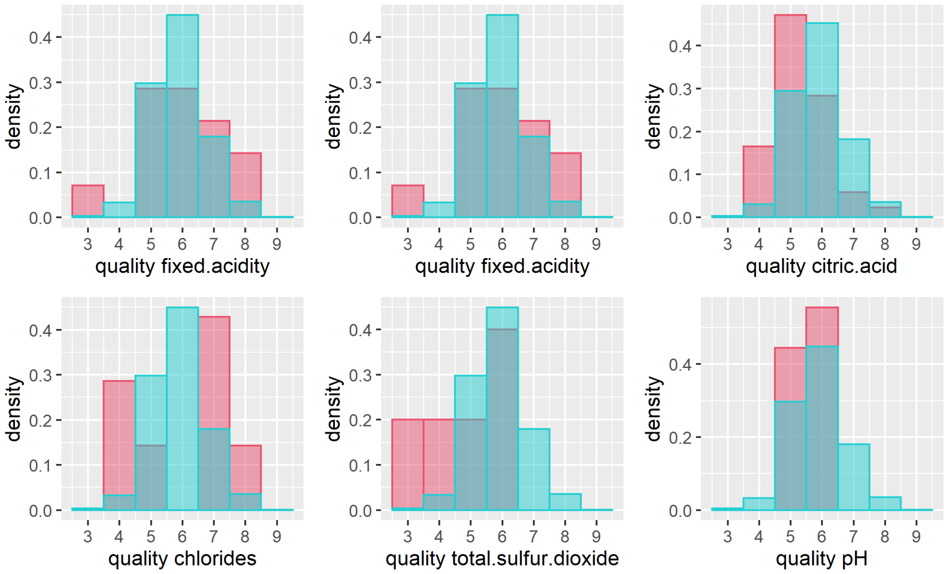

Figure 3: The red histograms represent the quality distributions of outliers, the blue histograms represent the quality distribution of the remaining dataset.

we can apply an outlier detection algorithm to identify good/bad wines. We plotted and compared the distribution of the quality without outliers with the distribution of the quality containing only outliers. First, for each explanatory variable we consider the quality of outlier wines below the lower bound and compare it with the quality of the wines which are above the lower bound (Figure 3). From these plots we notice very low values of total sulfur dioxide may negatively influence the quality of the wine, nevertheless, it does not seem to be a sufficient condition to claim a wine has poor quality, in fact, more than 50% of data with very low values of total sulfur dioxide are still decent wines (belongs to categories 5 and 6).
Finally, for each explanatory variable we consider the quality of outlier wines above the upper bound and compare it with the quality of the wines which are below the upper bound (Figure 4). Even in this case we can not say anything about the outliers, as all the distributions are almost comparable, thus, we can not exploit outliers to detect good/bad wines.

2.2 Correlation analysis

Now we want to analyze the correlation among the covariates.
We used the Pearson correlation coefficient to measure the linear correlation between variables.
From the correlation matrix we observe there are some explanatory variables with high correlation between them. Between the explanatory variables the ones with the highest correlation with the quality are alcohol and density, which are highly correlated between them. From (Figure 5) we might expect the alcohol's regressor to be positive, whereas the density's regressor to be negative.

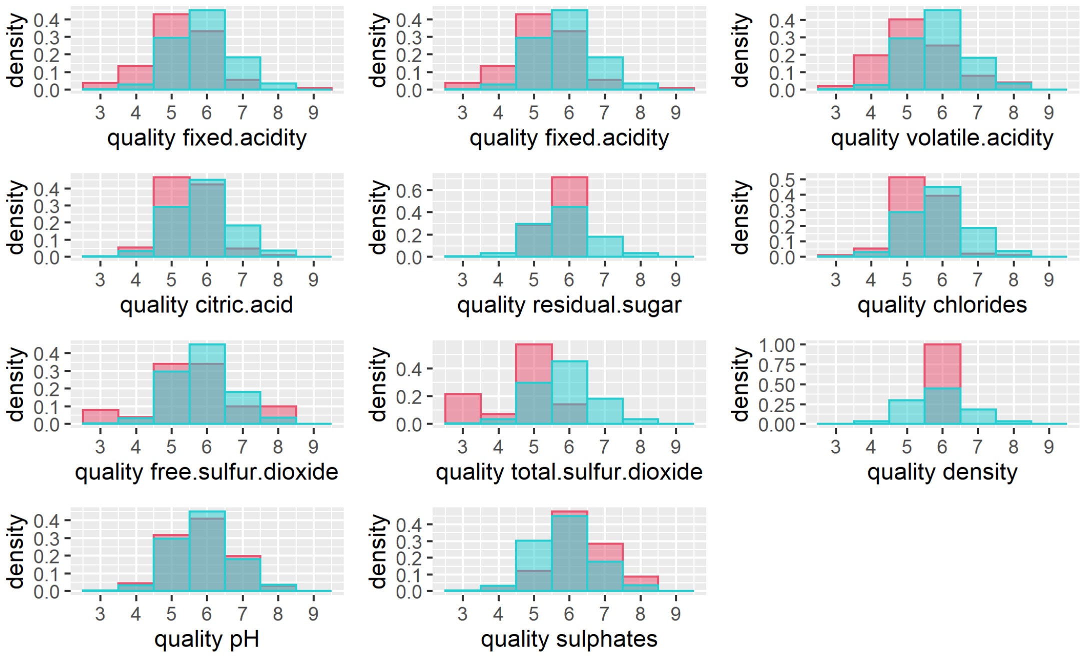

Figure 4: The red histograms represent the quality distributions of outliers, the blue histograms represent the quality distribution of the remaining dataset.

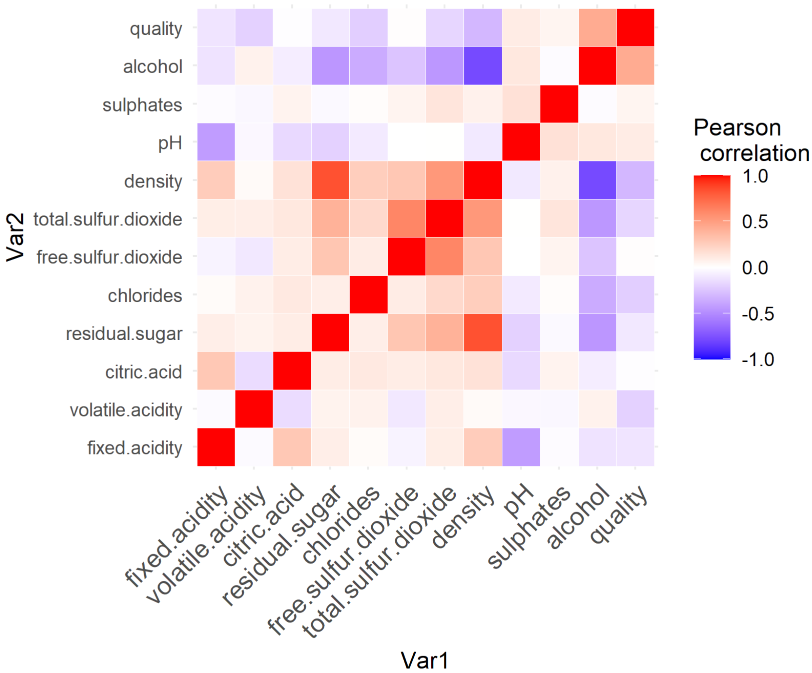

Figure 5: Pearson correlation coefficient of the variables in the dataset.

3 Model specification and posterior analysis

3.1 Categorical likelihood

We can imagine each quality represents a category k, thus, we can use a categorical distribution for the likelihood:

$$y_i \mid \boldsymbol{p_i} \overset{\text{ind}}{\sim} \mathcal{C}at(p_{0,i}, ...p_{10,i})$$
$$p_{j,i} = \frac{e^{\beta_j X_i}}{\sum_k e^{\beta_k X_i}}$$

where y_i and X_i are respectively the quality and the design matrix of wine i, $p_{j,i}$ is the probability of wine i of belonging to category j and $\boldsymbol{\beta_k}$ is the regressor vector for category k.
In this case we choose a Gaussian distribution as prior for the $\beta_{i,j}$:

$$\beta_{j,i} \overset{\text{ind}}{\sim} \mathcal{N}(\mu_0, \sigma^2)$$

and we set μ to 0 and σ to 10 to have a non informative prior, since we have no information to use.

Posterior analysis As we can see from Figure 6, we are not able to reach good results when sampling from the posterior distributions of the parameters (for more insights see Section 5), indeed we find different local maxima on the distributions.
The adoption of a categorical distribution implies the estimation of the posterior probability of $k \cdot n$ regressors, where n is the number of covariates. This because, we have to estimate for each category the regressors associated to each covariate. For this reason computing the posterior distributions results time expensive.
Given that we are not able to reach satisfying results in a reasonable amount of time, we have decided to not investigate more on this model and try a different approach.

3.2 Binomial likelihood

As mentioned in Section 1, we are facing an ordinal regression problem, thus, unlike in a classification problem, we are able to measure the distance between two categories. For this reason we want to use a likelihood that allows us to solve a regression problem, but at the same time we want a likelihood that has the same domain of the quality wine (discrete and in the range $[0, .., 10]$), so, a possible solution could be a binomial distribution:

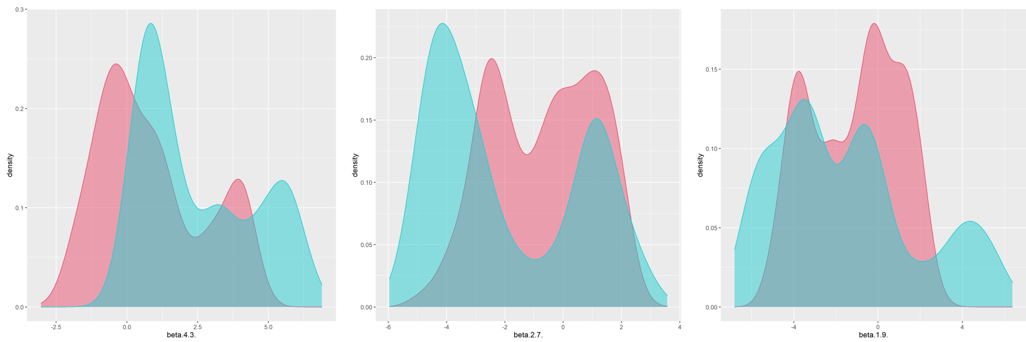

Figure 6: An example of posterior distributions for the regressors of the categorical likelihood. We used two different chains to sample the posterior distributions, thus the red distribution is sampled from the first chain, whereas the blue one is sampled from the second chain.

$$
\begin{aligned}
y_i \mid p_i, n &\overset{\text{ind}}{\sim} \mathcal{B}i(p_i, n)\\
p_i &= \Phi(\beta_0 + X_i\beta)\\
n &= 10
\end{aligned}
$$

where p_i is the probability of success in a Bernoulli trial, n is the number of Bernoulli trials, X_i is the design matrix for the wine i, β is the regressor vector, β_0 is the intercept parameter, finally, Φ is the cumulative distribution function of a Gaussian distribution $\mathcal{N}(0,1)$. The use of a link function as Φ is essential for the regression, since p_i take values in the range $[0,...,1]$ whereas $\beta_0 + X_i\beta$ can take any possible value. In order to fully represent the domain of the response variable we set the number of Bernoulli trials n to 10.

3.2.1 Choice of the prior and posterior analysis

For the binomial regression we adopt and compare three different priors.

Gaussian prior First we try with a non informative Gaussian prior for the same reasons of the Categorical likelihood:

$$
\begin{aligned}
\beta_i \mid \lambda^2 &\overset{\text{ind}}{\sim} \mathcal{N}(0, \frac{1}{\lambda^2}).\\
\lambda^2 &\sim \mathcal{G}(\alpha, \beta)
\end{aligned}
$$

where $\alpha = 1$ and $\beta = 0.01$ to induce a non informative prior over the parameters [1]. From the posterior distributions (Figure 7, 8), we can notice that residual sugar, free sulfur dioxide, sulphates and alcohol have a positive influence on the quality of the wine, whereas volatile acidity and density negatively impact the quality. Notice how alcohol and density satisfy our expectations in Section 2.2.

Laplace distribution Then, we adopt a Laplace distribution[2] as a Lasso prior to make some feature selection:

$$
\begin{aligned}
\beta_j \mid \lambda^2 &\overset{\text{ind}}{\sim} \mathcal{L}aplace(0, \frac{1}{\lambda^2})\\
\lambda^2 &\sim \mathcal{G}(\alpha, \beta).
\end{aligned}
$$

In this case we set different parameters for α and β, in order to make some comparisons between different shrinkage methods. We set the parameters to:

1. $\alpha = 0.01$, $\beta = 0.1$ (λ^2 red posterior distribution in Figure 9)
2. $\alpha = 0.1$, $\beta = 0.1$ (λ^2 green posterior distribution in Figure 9)
3. $\alpha = 0.5$, $\beta = 0.1$ (λ^2 blue posterior distribution in Figure 9)

From the pictures mentioned above, we notice that by decreasing the values of α, the λ^2 posterior distribution moves towards 0, thus, the variance of the Laplace distribution decreases. If the variance of the Laplace distribution increases, the shrinkage effect given by the Lasso prior is lighter, so we should expect that the posterior distribution of the regressors is closer to 0 when α is low. However, in our case changing the hyper-parameter does not affect much the posterior distributions.

The posterior distributions of the betas are shown in Figure 10 (for the statistics go to Section 5). The obtained distributions are very similar, this means that the model has a low sensitivity to the *alpha* hyper-parameter.

[1] In this case $\frac{1}{\lambda^2}$ is the precision of the Gaussian, following the JAGS convention.

[2] Even for the Laplace distribution we consider $\frac{1}{\lambda^2}$ as precision following the JAGS convention

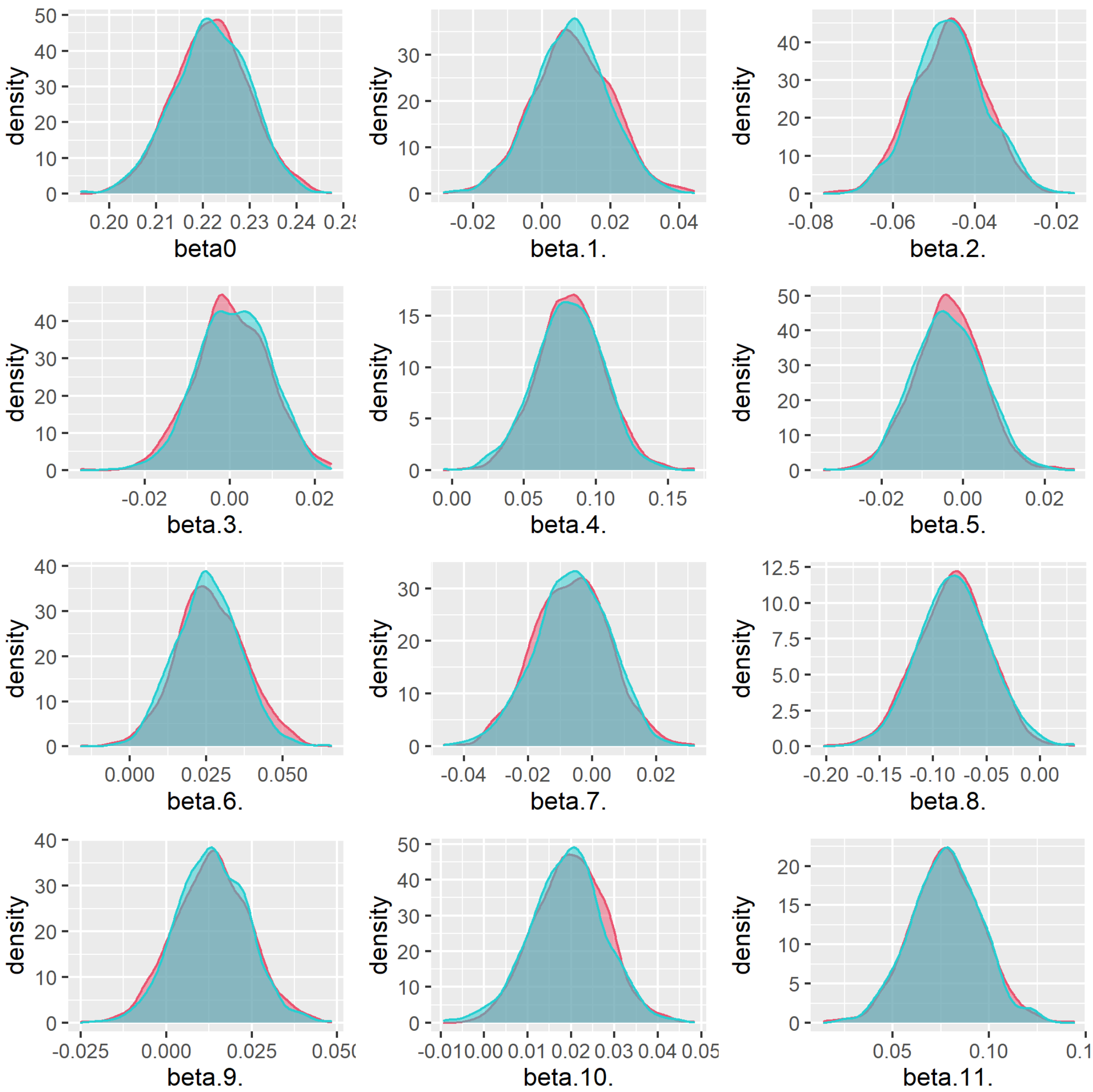

Figure 7: Posterior distributions for a binomial likelihood using a Gaussian prior. The blue and red distributions are taken from different Markov chain.

	statistics.Mean	statistics.SD	statistics.Naive.SE	statistics.Time.series.SE
fixed.acidity	0.00865240670473402	0.0110805354608838	0.000247768305176331	0.000256595459617886
volatile.acidity	-0.0462618432663171	0.0086994474185375	0.000194525557945349	0.000194204575581909
citric.acid	0.000741388809413038	0.00845253991206039	0.000189004538258971	0.000201731213117701
residual.sugar	0.0817348645450253	0.0232183540191936	0.000519178179125724	0.000532394949442075
chlorides	-0.00370512033510826	0.00822624593660328	0.000183944451138764	0.000189699919634919
free.sulfur.dioxide	0.0259902334546648	0.0108683018036918	0.000243022616330386	0.000242125890750749
total.sulfur.dioxide	-0.00642293902151802	0.0117097166159013	0.000261837223504141	0.000261887727879433
density	-0.0809159753147809	0.0329910972448172	0.000737703360917172	0.000813754289250866
pH	0.0133356556464551	0.0105274751179389	0.000235401499951491	0.000241267127311429
sulphates	0.0195379079421313	0.0083011691953635	0.0001856197861356	0.000189378312844201
alcohol	0.0778404312696663	0.0180859569087061	0.000404414290859989	0.000418321857277962
intercept	0.221806727837464	0.00802186519648627	0.00017937435885683	0.00017941808206211

Figure 8: Statistics of posterior distributions got with Gaussian prior, Binomial likelihood.

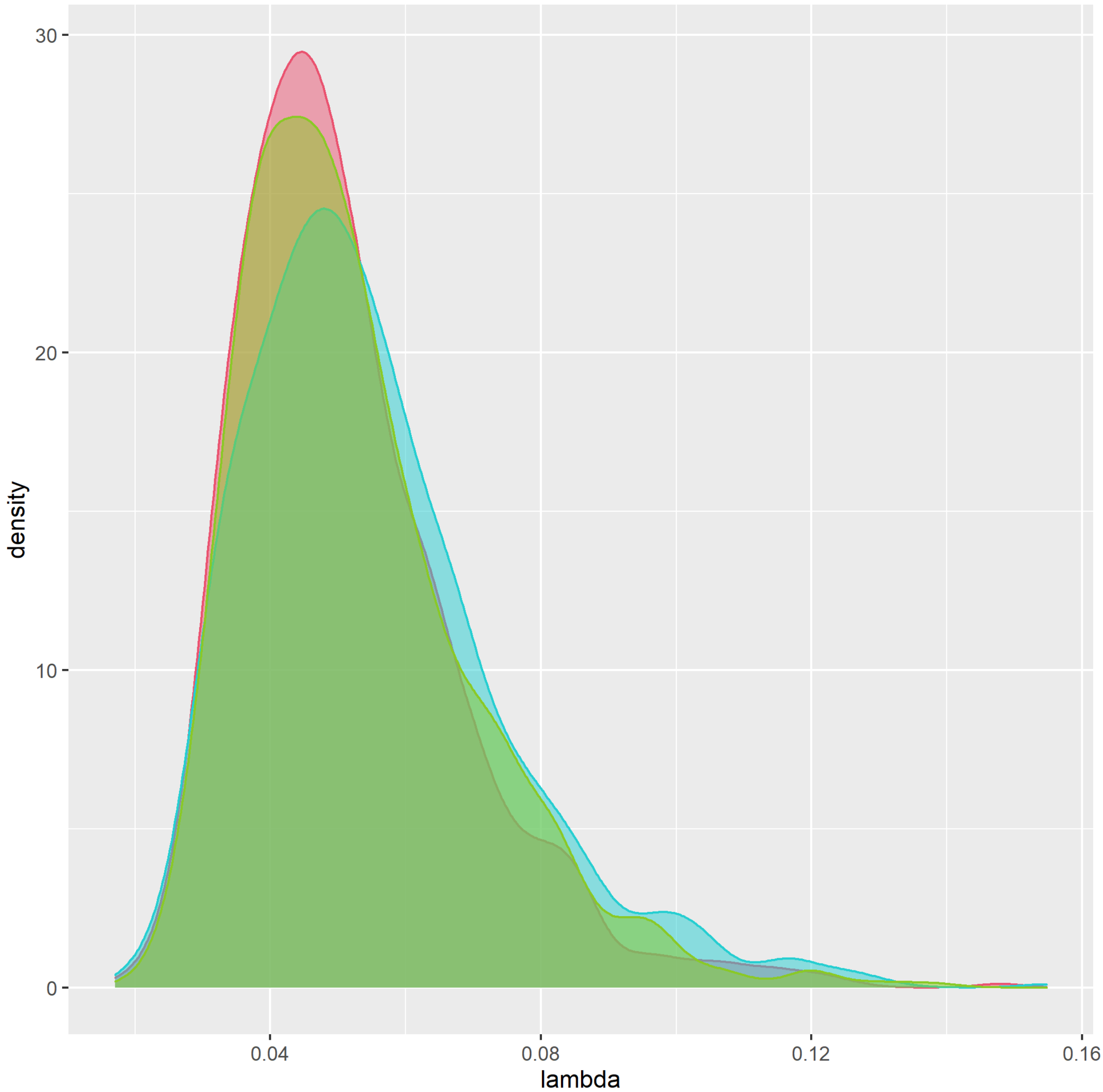

Figure 9: Comparison between the posterior distributions of lambdas for the different Lasso priors. Red distribution corresponds to $\alpha = 0.01$, $\beta = 0.1$. Green distribution corresponds to $\alpha = 0.1$, $\beta = 0.1$. Blue distribution corresponds to $\alpha = 0.5$, $\beta = 0.1$.

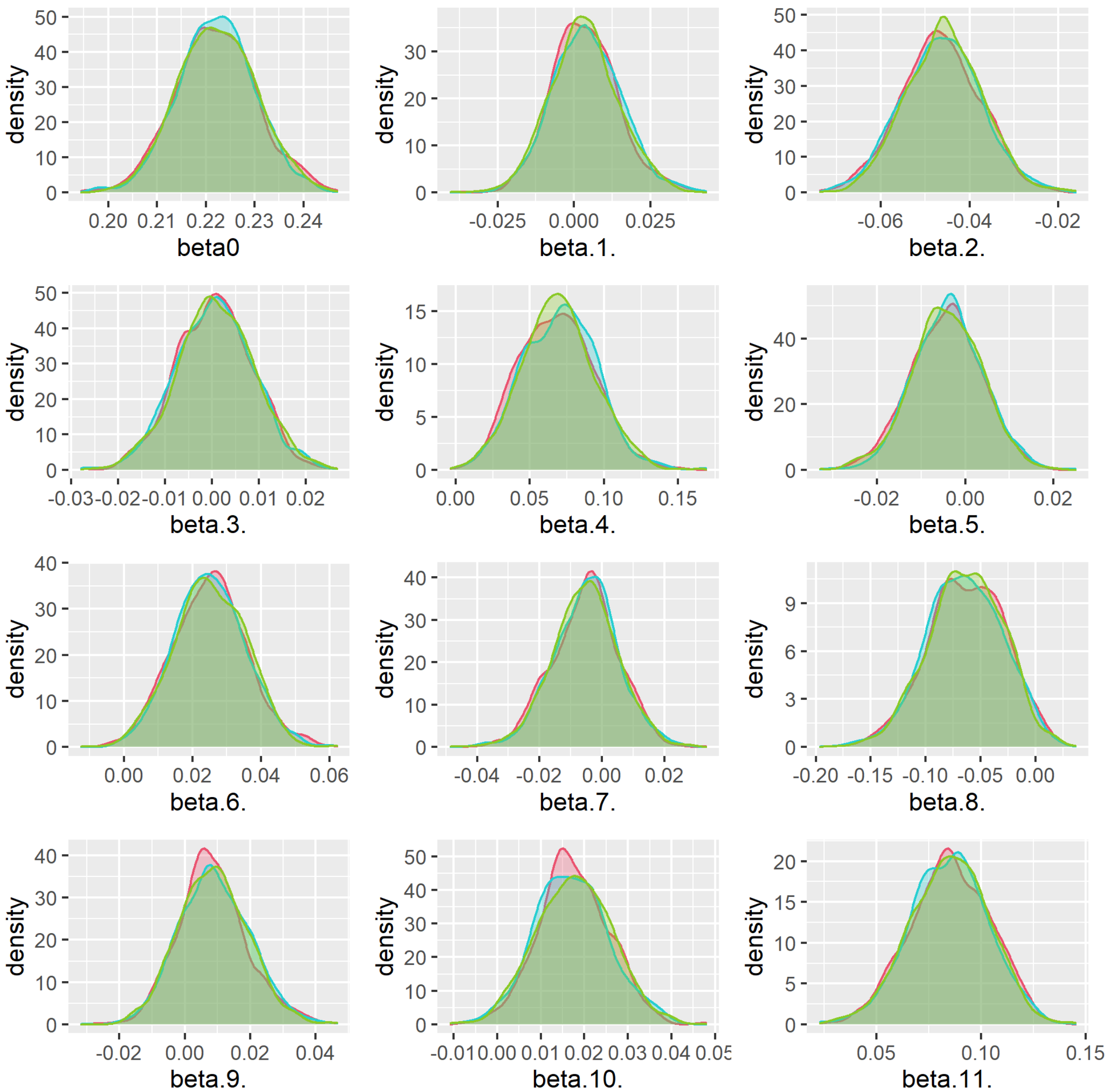

Figure 10: Comparison between the posterior distributions of bets for the different Lasso priors. Red distribution corresponds to $\alpha = 0.01$, $\beta = 0.1$. Green distribution corresponds to $\alpha = 0.1$, $\beta = 0.1$. Blue distribution corresponds to $\alpha = 0.5$, $\beta = 0.1$.

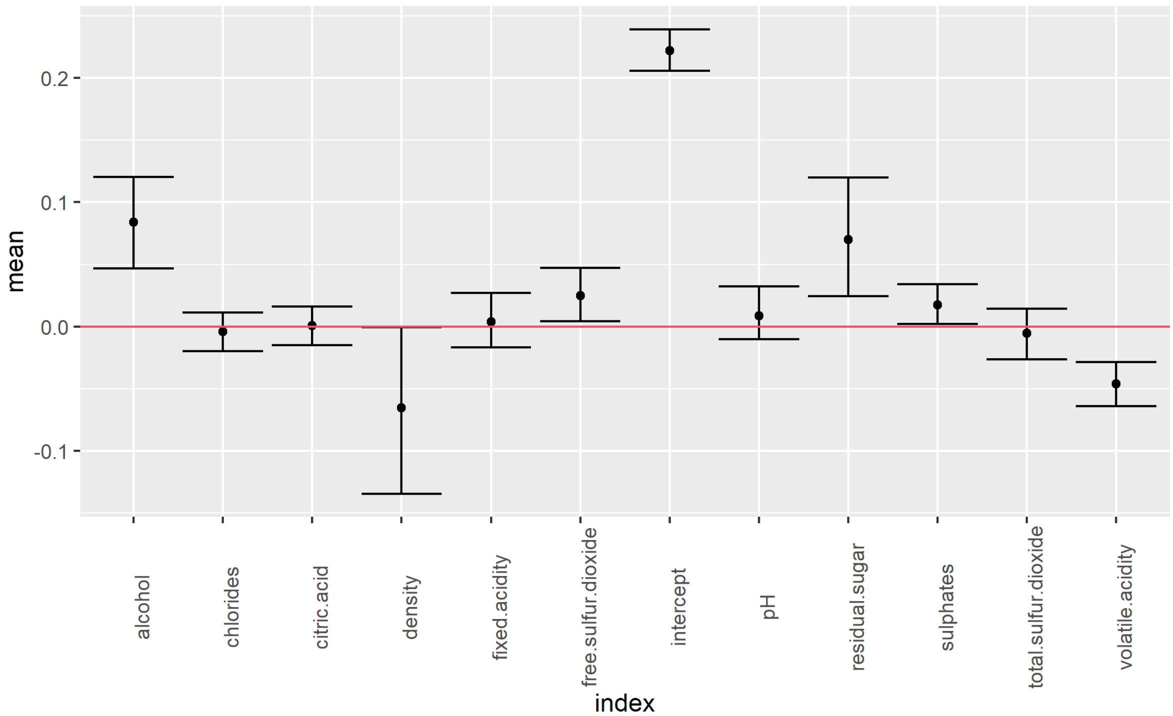

Figure 11: Confidence intervals of covariates after using Lasso prior associated to $\alpha = 0.01$

With a confidence of 95%, as we can see from Figures 11, 12, 13, even changing the prior we exclude always: chlorides, citric.acidity, fixed.acidity, pH, total.sulfur.dioxide. These covariates are not significant to determine the quality of the wine. On the other hand we keep: alcohol, density, residual.sugar, sulphates, volatile.acidity.

Since the posterior obtained changing the parameters of the Lasso prior are similar, from now on we only consider $\alpha = 0.1$ and $\beta = 0.1$.

We can compare the posterior distributions of the Gaussian and Laplace priors, the result is shown in Figure 14. We can see that the distributions are similar, but the Lasso prior pushes the posterior distributions towards zero, except for beta 11.

Spikes & Slab Spikes and slab is an approach for model selection where each of the possible 2^p subset choices for a model with p covariates are indexed by the vector $\gamma = (\gamma_1, ..., \gamma_p)$. The value β_i of the coefficients' vector β is different from zero if and only if $\gamma_i = 1$, while if $\gamma_i = 0$ then also $\beta_i = 0$

We model the uncertainty underlying variable selection using a mixture prior:

$$\pi(\beta, \gamma) = \pi(\beta \mid \gamma) \pi(\gamma)$$

where:

$$\gamma_j \overset{\text{ind}}{\sim} \mathcal{B}e(\theta_j) \quad j = 1, \ldots, p$$

In this context, θ_j is the probability that β_j is large enough to justify including the corresponding covariate to be included in the model.

To wrap up, the Spike & Slab prior for model selection is given by:

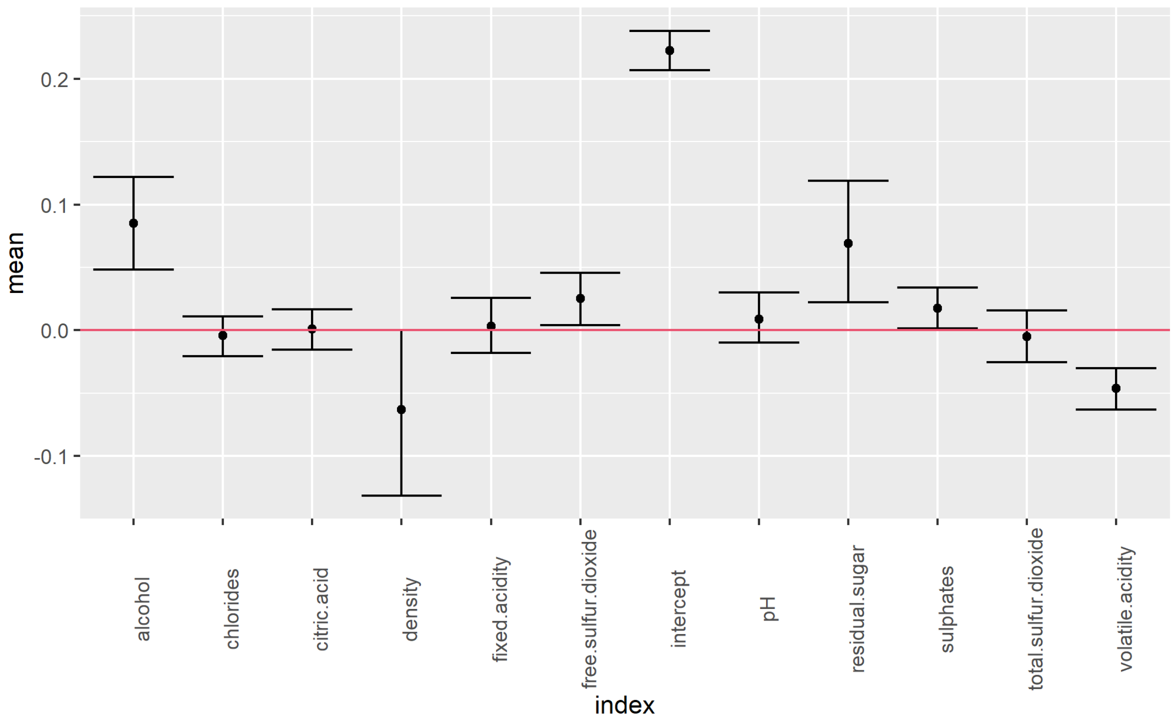

Figure 12: Confidence intervals of covariates after using Lasso prior associated to $\alpha = 0.1$

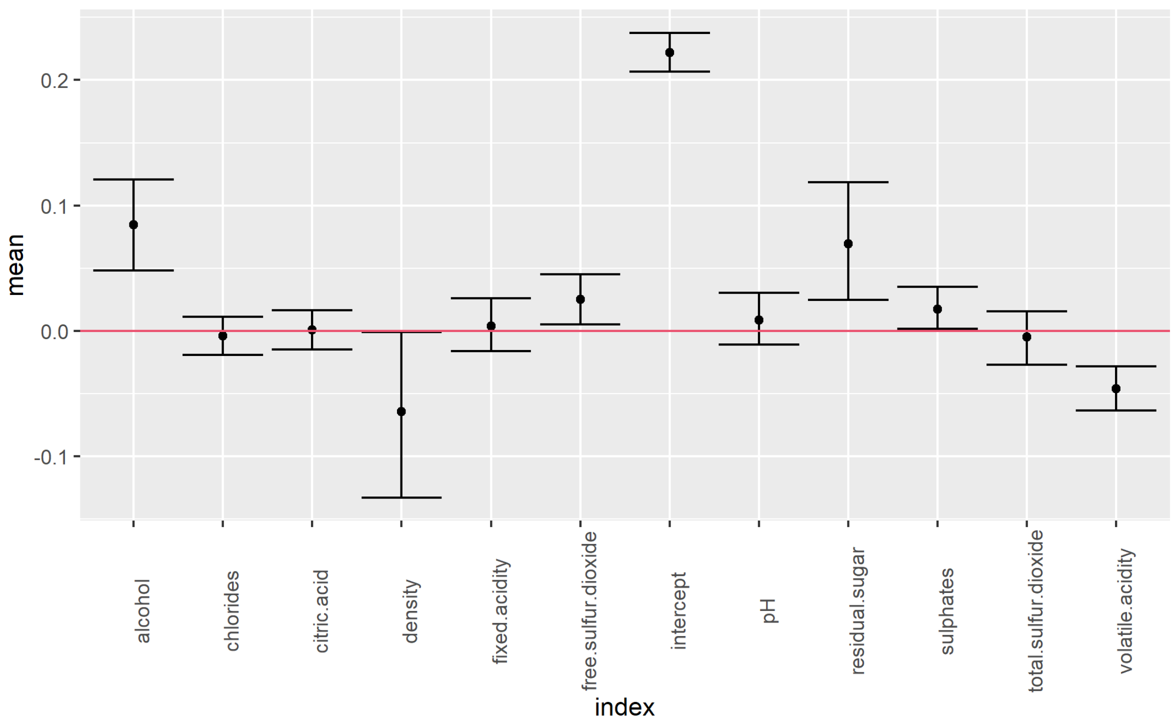

Figure 13: Confidence intervals of covariates after using Lasso prior associated to$\alpha = 0.5$

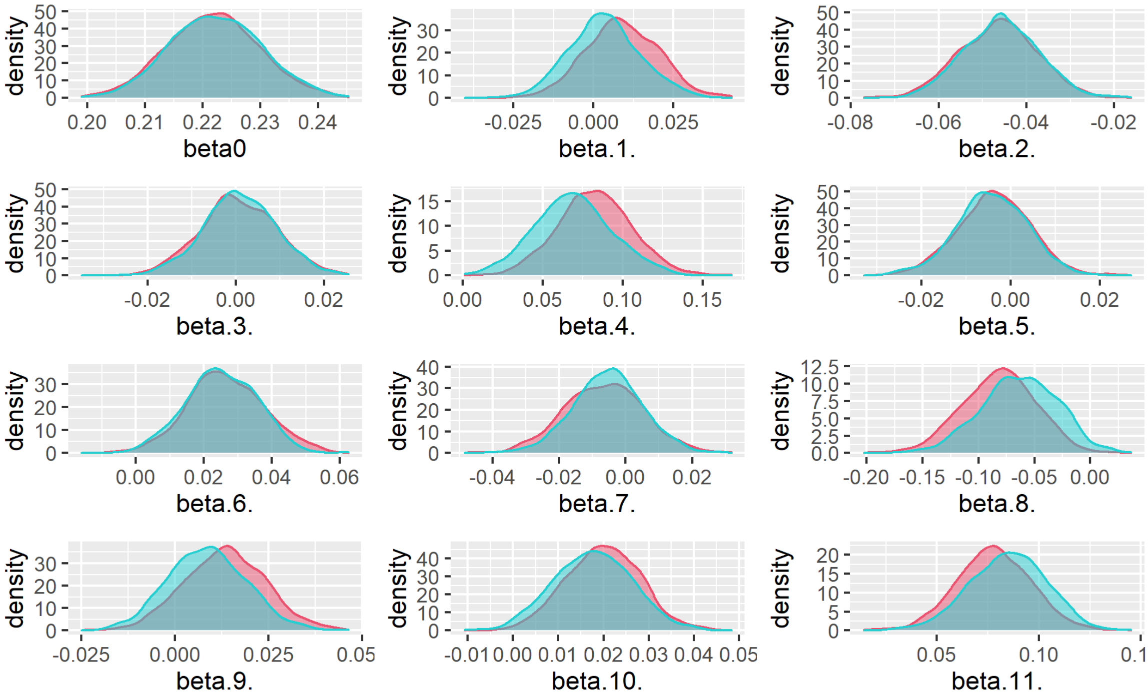

Figure 14: Posterior densities of the beta parameters using the Gaussian prior (red) and Laplace prior (blue).

$$
\begin{aligned}
\beta_j \mid \gamma_j &\overset{\text{ind}}{\sim} (1-\gamma_j)\,\delta_{\{0\}} + \gamma_j\,\mathcal{N}\left(0, \sigma^2_{\beta_j}\right) \\
\gamma_j \mid \theta_j &\overset{\text{ind}}{\sim} \mathcal{B}e\left(\theta_j\right) \\
\theta_j &\overset{\text{iid}}{\sim} \pi\left(\theta_j\right)
\end{aligned}
$$

where $\pi(\theta_j)$ is a Uniform distribution over the interval $[0,1]$. In our dataset we considered the likelihood of the quality target to be distributed as a binomial with $n = 10$ resulting in the following model:

$$
\begin{aligned}
y_j \mid p_j &\overset{\text{ind}}{\sim} \mathcal{B}i(10, p_j) \\
p_j &= \Phi(\beta_0 + x_j^T \beta) \\
\beta_j \mid \gamma_j &\overset{\text{ind}}{\sim} (1-\gamma_j)\,\delta_{\{0\}} + \gamma_j\,\mathcal{N}\left(0, \sigma^2_{\beta_j}\right) \\
\gamma_j \mid \theta_j &\overset{\text{ind}}{\sim} \mathcal{B}e\left(\theta_j\right) \\
\theta_j &\overset{\text{iid}}{\sim} \pi\left(\theta_j\right)
\end{aligned}
$$

The resulting estimated posterior inclusion probabilities (posterior means of the gamma variables) are represented in (Figure 15).

Following the **The Median Probability Model (MPM)** technique where pick variables with estimated posterior inclusion probabilities higher than 0.5, we would select: alcohol, density, pH, residual.sugar and volatile acidity.

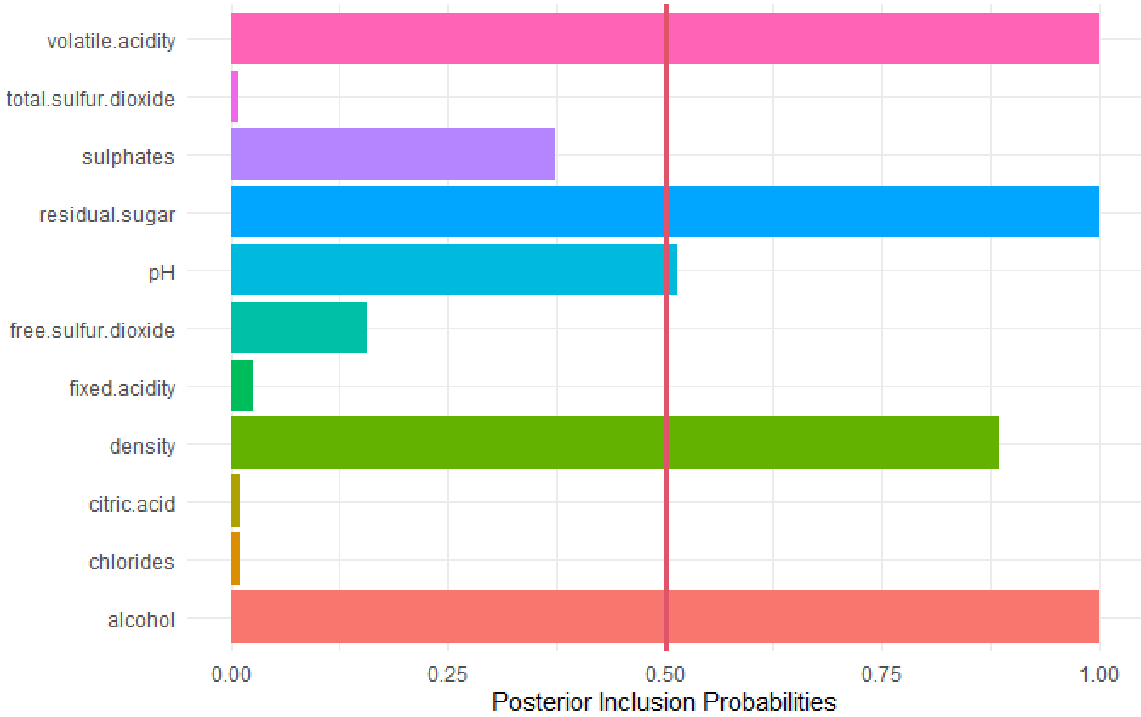

Figure 15: Estimated posterior inclusion probabilities (posterior means of the gamma variables).

If we use the **Highest Posterior Density model (HPD)** technique instead, in which we pick the model with the highest estimated posterior probability we would select the same covariates (alcohol, density, pH, residual.sugar and volatile acidity)

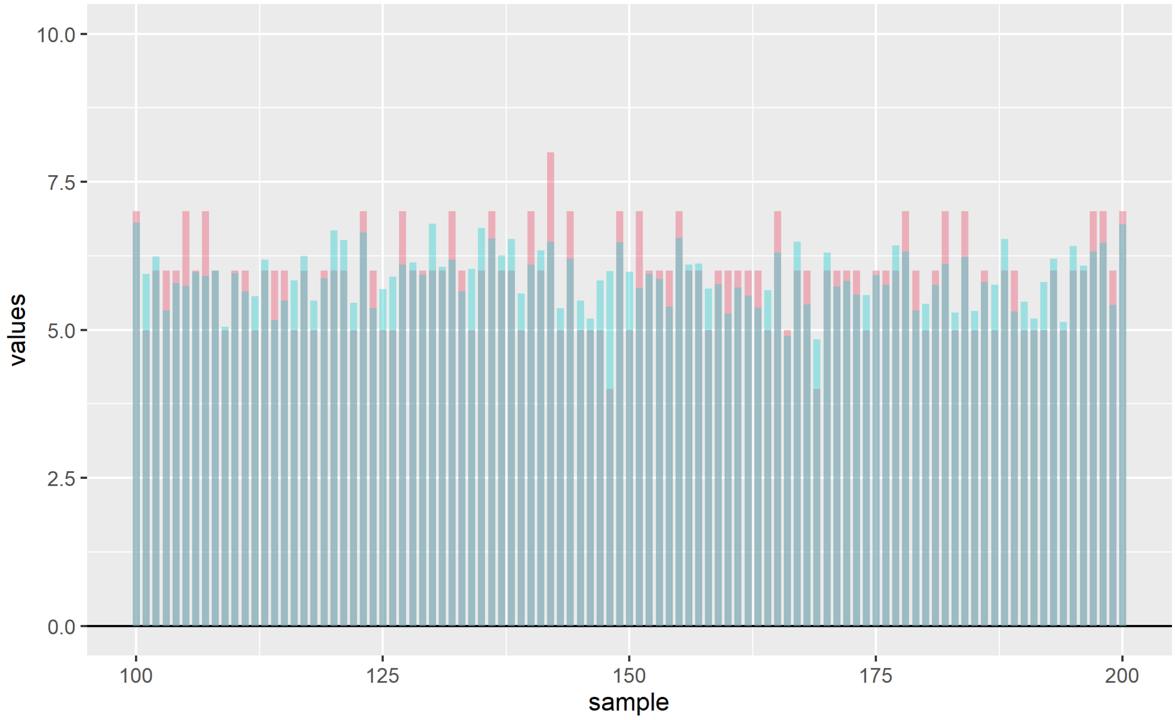

Figure 16: Prediction outcome over a subset of samples belonging to the test set. Blue bars represent the predictions using Gaussian prior with binomial likelihood, whereas the red bars represent the true values of quality.

4 Prediction

Since the categorical likelihood was not able to reach decent posterior results, we try to make prediction only using the Binomial model. We compare the results got using the Gaussian prior with the ones got by the Lasso prior(with hyperparameters of its Gamma prior $\alpha = 0.1$, $\beta = 0.1$). For brevity we make predictions only using the Lasso with the mentioned prior. Using the Gaussian prior we register an MSE=0.5743 and a number of miss-classified wines equal to 1209 out of 2449. We decided to use the MSE to measure the quality of the prediction since we are still able to measure the distance between two categories. For the case of the MSE the distance has been computed as:

$$\epsilon = y_i - \mu_\pi(i)$$

where y_i is the true value, whereas $\mu_\pi(i)$ is the posterior mean of the predictive distribution of the quality of wine i. In order to compute the number of miss-classified wines we have to assign each predictive distribution to a specific category, so we decided to assign wine i to the category k_i:

$$k_i = round(\mu_\pi(i))$$

where $round()$ is a function that returns the closest integer to its input. From this results is clear that our predictions are not precise since almost the 50% of data has been miss-classified, nevertheless, since the MSE is not so high, it means that when a wine is miss-classified, this is assigned to a category close to the true one, as we can see from Figure 16. This is also due to the fact that most wines are in the middle categories and the prediction results always in these few categories.

For the Lasso prior on the other hand we obtain a MSE=0.5755 and a number of miss-classified wines equal to 1200 out of 2449. The result we obtain using the two different priors are very similar.

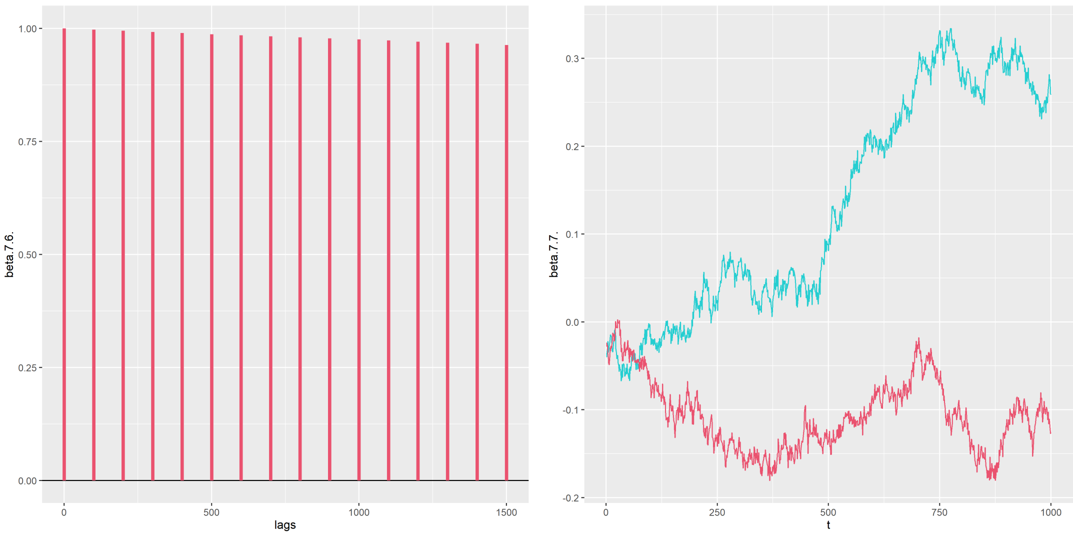

Figure 17: Trace and auto correlation plot for residual.sugar, using the categorical prior.

5 Appendix

In this section we want to add some additional analysis about the posterior distribution we got using a Gibbs sampler.

5.1 Categorical likelihood

As mentioned in the Subsection 3.1, the posterior results present some issues. Since the number of regressors is very high (110) and their correlation plots and trace plots present almost the same issues, for sake of simplicity we show just the results for a single regressor. As we can see from Figure 17, there is a lot of auto correlation even after 15 accepted samples (notice that we used thinning = 100, thus, after 1500 draws from the chain there is still a lot of correlation). High auto correlation implies poor exploration and the presence of strong seasonality in the trace plots (Figure 17). For this reasons we are not able to reach the convergence to the true posterior distribution.

5.2 Binomial likelihood

5.2.1 Gaussian prior

Here we show the trace plots and auto correlation plots for the posterior distribution of the parameters with Gaussian prior. From Figure 18 we see how accepting a sample every 50 draws we have very low correlation between samples, indeed looking at the trace plots Figure 19 there is no evidence that the posterior distributions have not reached the convergence.

5.3 Binomial likelihood

5.3.1 Laplace prior

Here we show the trace plots and auto correlation plots for the posterior distributions of the parameters having Lasso prior ($\alpha = 0.1$, $\beta = 0.1$). Since changing the hyper-parameters of the priors does not change the behaviour of those plots, for brevity, we decide to show the plots (Figures 20, 24) only for this prior.

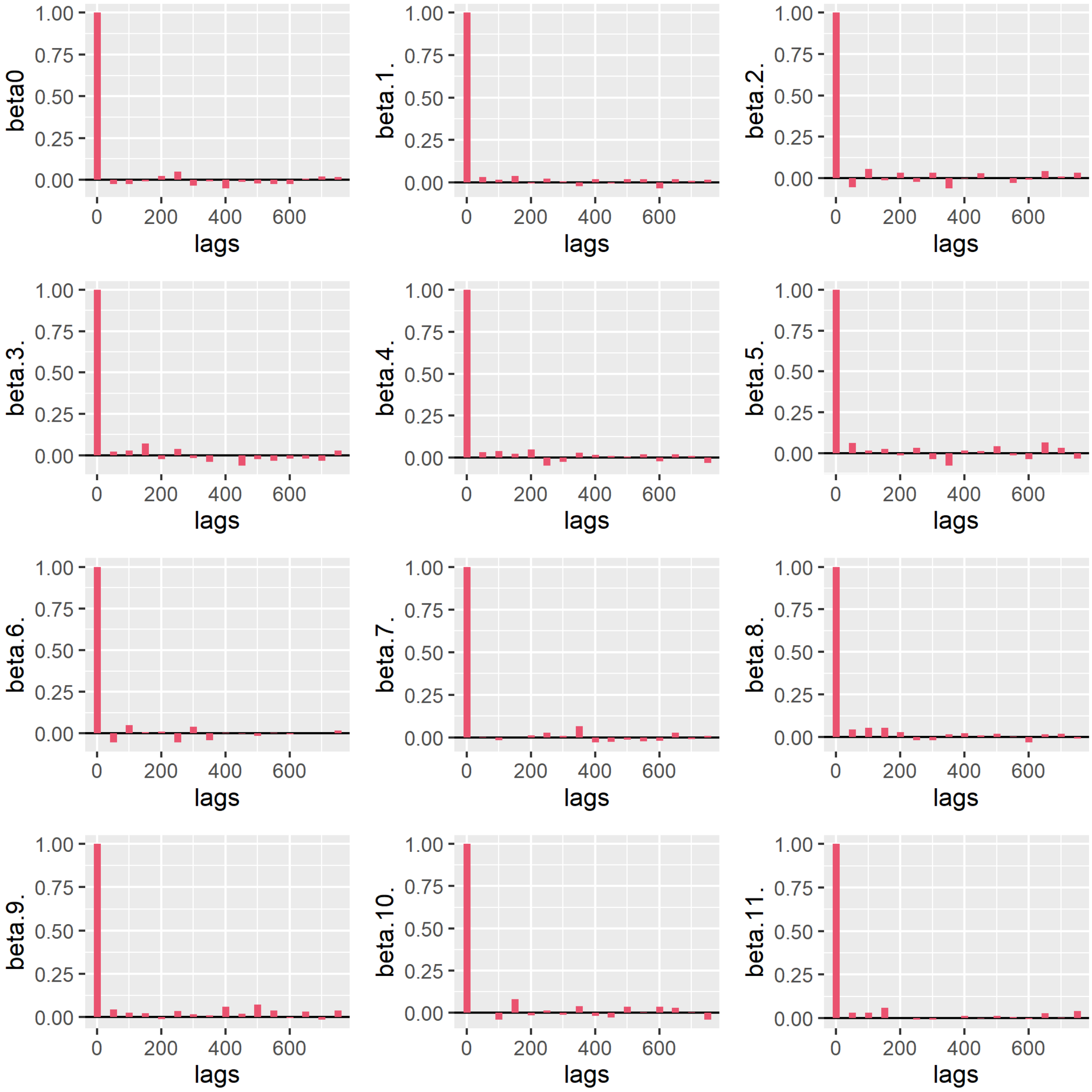

Figure 18: Auto correlation plots for parameters with prior $\mathcal{N}(0, \frac{1}{\lambda^2})$ and Binomial likelihood

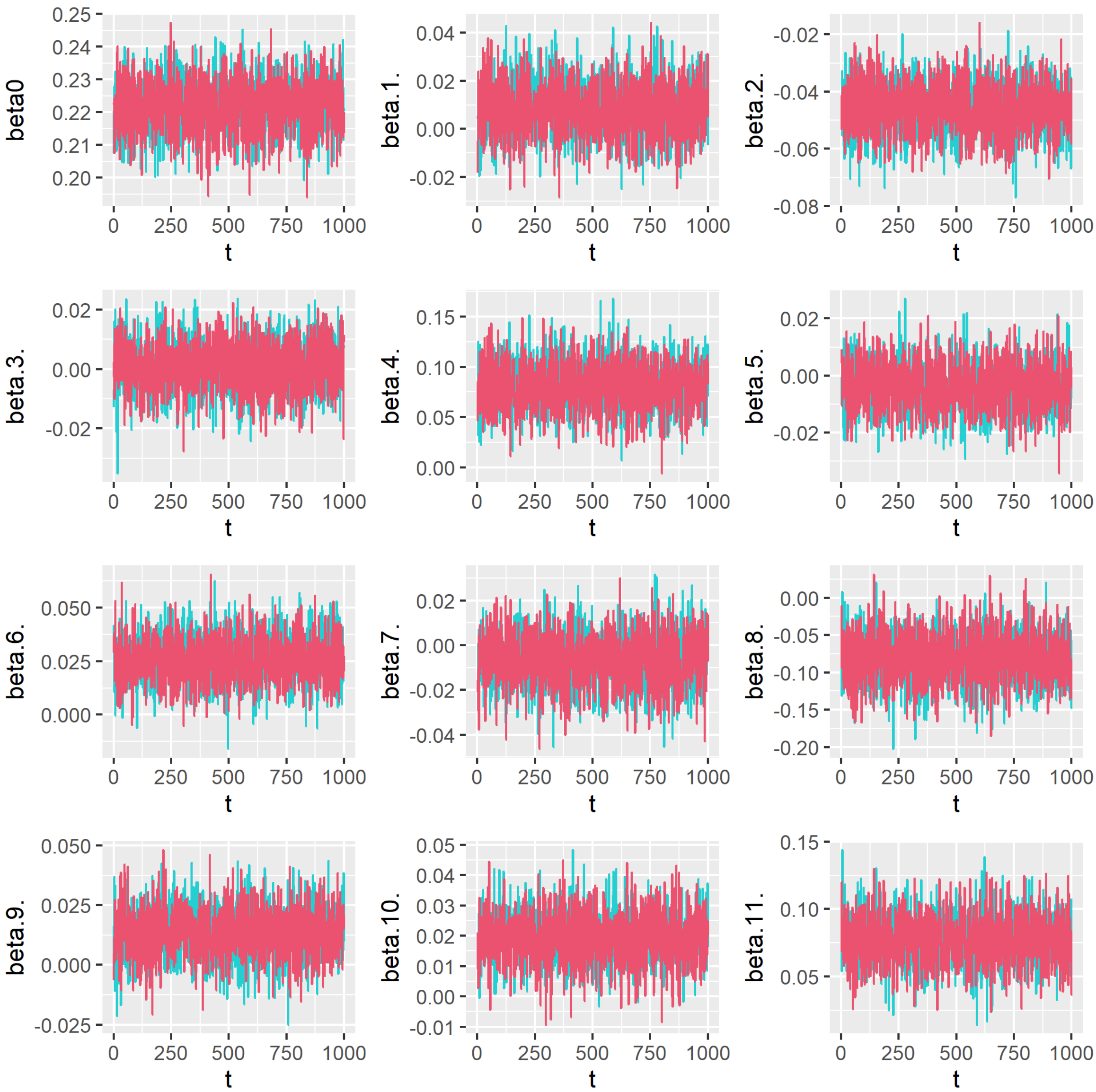

Figure 19: Trace plot for for parameters with prior $\mathcal{N}(0, \frac{1}{\lambda^2})$ and Binomial likelihood. Red traces belong to the first chain, blue traces belong to the second chain.

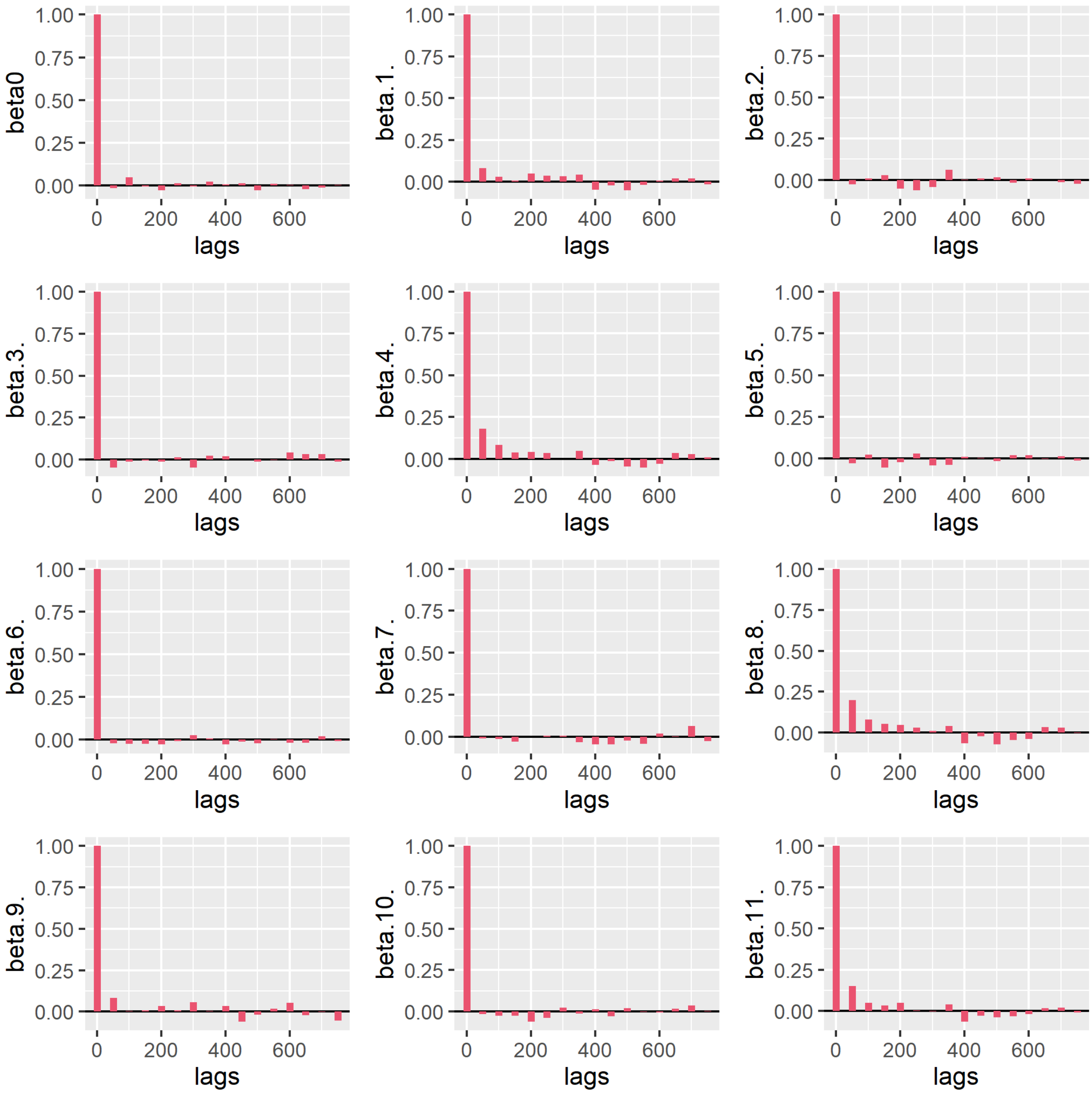

Figure 20: Auto correlation plots for parameters with Laplace prior and Binomial likelihood

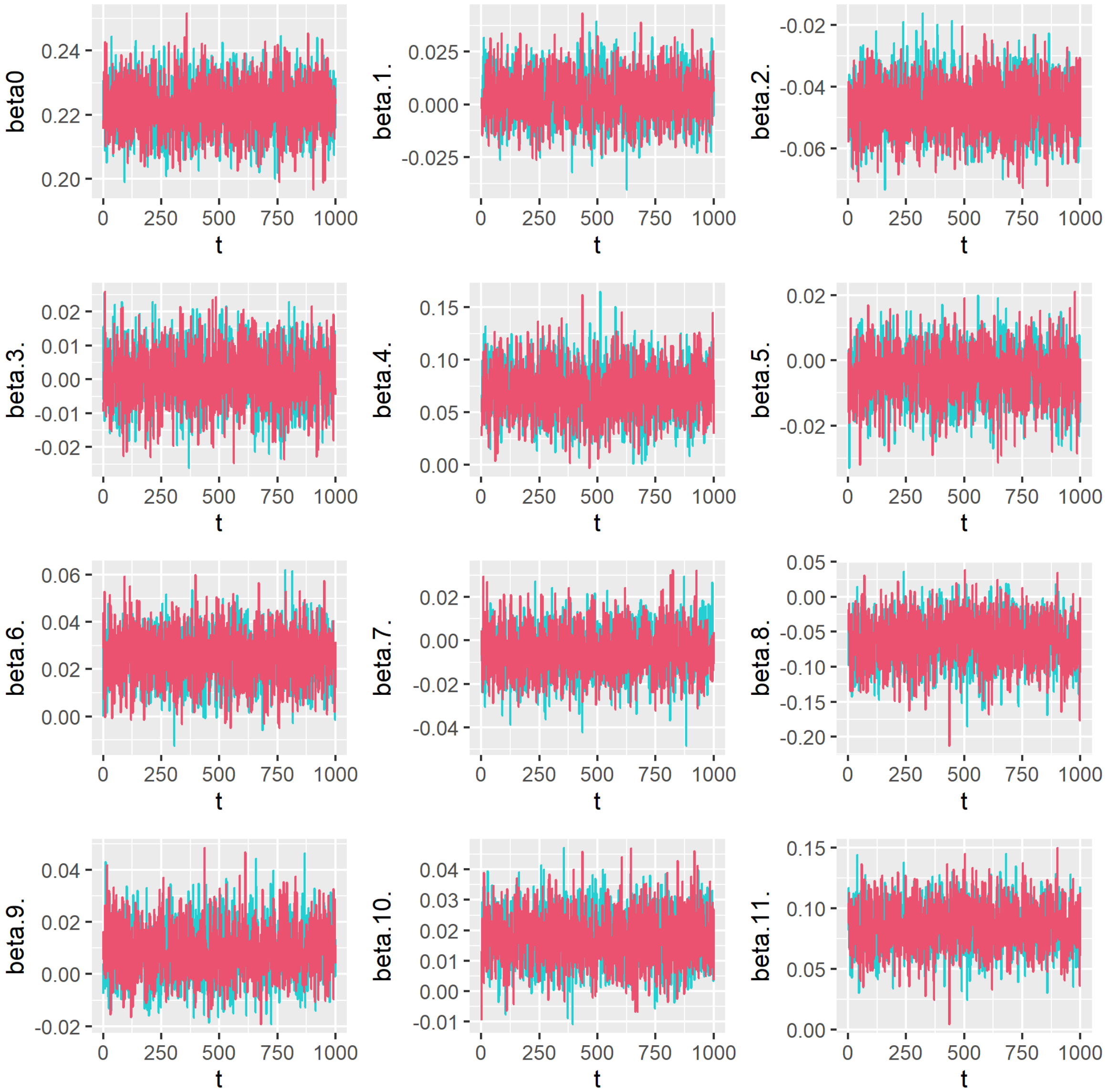

Figure 21: Trace plot for for parameters with Laplace prior and Binomial likelihood. Red traces belong to the first chain, blue traces belong to the second chain.

5.4 Statistics for Lasso

5.5 Statistics for Spikes & Slab

Here we show the trace plots and auto correlation plots for the posterior distributions of the parameters having Spikes & Slab prior. We ran two chains but the results are the same so here we plot only one chain. We can see how the gamma variables influence the beta_temp densities (the normal distributions before applying the spikes & slab) to create the beta densities. Some betas have all the mass in zero because their corresponding gamma has most of the mass in zero. Here the trace plots Figure 25, posterior densities in Figure 26 and auto correlation plots in Figure 27.

	statistics.Mean	statistics.SD	statistics.Naive.SE	statistics.Time.series.SE
fixed.acidity	0.00401620864678915	0.0108736544851446	0.000243142305926288	0.000250594207974404
volatile.acidity	-0.0460548577227604	0.00893219235295482	0.000199729892893108	0.00021438094558184
citric.acid	0.000659304959427967	0.00796753779394011	0.000178159561205488	0.000178171800096149
residual.sugar	0.0695830749805464	0.024796088530022	0.00055445739529232	0.000624946212107999
chlorides	-0.00401067630572898	0.00782382482187717	0.000174946041457675	0.000167222392601723
free.sulfur.dioxide	0.0250109110707778	0.0104049071115074	0.000232660796009015	0.000232325177288802
total.sulfur.dioxide	-0.00494111896722177	0.0105881137808565	0.000236757421674974	0.000247989921331528
density	-0.0643132682358503	0.0348053088509562	0.000778270365686132	0.000876966219713693
pH	0.00883687629024039	0.0105638005946895	0.000236213762304785	0.000244227840467444
sulphates	0.0175045166095831	0.00855127134274807	0.000191212240164306	0.000197446438844992
alcohol	0.0848426748447465	0.0185841805573802	0.000415554910324322	0.000452009260761468
intercept	0.222110335566105	0.00789496772104631	0.000176536845044261	0.000175229654833597
lambda	0.055010926677227	0.0189766822966903	0.000424331516028162	0.000424436749666934

Figure 22: Summary for the posterior distributions of the binomial lasso model associated to $\alpha = 0.05$, showcasing the features that have positive and negative effects on the quality.

	statistics.Mean	statistics.SD	statistics.Naive.SE	statistics.Time.series.SE
fixed.acidity	0.00339279173807158	0.0110776616482006	0.000247704044771189	0.000270547803243199
volatile.acidity	-0.0460373440521979	0.00861112189486983	0.000192550539194657	0.000184334865642392
citric.acid	0.000824225226165448	0.00811871813760189	0.000181540056458383	0.000177415332200903
residual.sugar	0.0688350308105808	0.0244954639737089	0.000547735225856103	0.000671071823524475
chlorides	-0.00426887871477146	0.00790183589124429	0.000176690421998699	0.000174102653548959
free.sulfur.dioxide	0.0252925746907623	0.0108534017221247	0.000242689440377841	0.00024969658187972
total.sulfur.dioxide	-0.00521594440717227	0.0106782095699555	0.000238772024764093	0.000234589614414152
density	-0.0630817846745401	0.0344931318567188	0.000771289875884867	0.000916981388036701
pH	0.00877748262745901	0.0103348163175054	0.00023109351820916	0.000250967095683053
sulphates	0.0174485527686765	0.00855110954153503	0.000191208622179194	0.000191187437795322
alcohol	0.085245493326621	0.0189831219555042	0.000424475511176762	0.000481056619728815
intercept	0.222271768134419	0.00801967644785842	0.000179325416949655	0.000179367470122157
lambda	0.0530882151623218	0.0179611342271083	0.000401623170848123	0.00040160364025844

Figure 23: Summary for the posterior distributions of the binomial lasso model associated to $\alpha = 0.1$, showcasing the features that have positive and negative effects on the quality.

	statistics.Mean	statistics.SD	statistics.Naive.SE	statistics.Time.series.SE
fixed.acidity	0.00397162530491971	0.010873203099601	0.000243132212638693	0.000249990864154947
volatile.acidity	-0.0462494669628807	0.00889481052324053	0.000198894009769463	0.000198727471535092
citric.acid	0.000605492725823206	0.00789535975889642	0.000176545611277087	0.000176585861078361
residual.sugar	0.0698669542272547	0.0249482962947809	0.000557860864379361	0.00063704483814627
chlorides	-0.00424200032897855	0.00798406415810871	0.000178529101942507	0.000188390039598659
free.sulfur.dioxide	0.0249808016626179	0.0107959996110368	0.000241405890153396	0.000241462673592528
total.sulfur.dioxide	-0.00545364100111029	0.01050636654029	0.000234929497806178	0.000218020072445697
density	-0.0651521352324584	0.0349656140366115	0.000781854898608841	0.000907585496220849
pH	0.00887877546144941	0.0106983297313999	0.000239221925251174	0.000245198826236324
sulphates	0.0175578240096781	0.00821997683087899	0.000183804269673187	0.000183850243223463
alcohol	0.0841093577872781	0.0191165210640115	0.000427458405924364	0.000470748244305589
intercept	0.221964533964602	0.00832525254167839	0.000186158306130458	0.000190805133905964
lambda	0.0519437575301952	0.0167899866686393	0.000375435515323928	0.000375346575593098

Figure 24: Summary for the posterior distributions of the binomial lasso model associated to $\alpha = 0.01$, showcasing the features that have positive and negative effects on the quality.

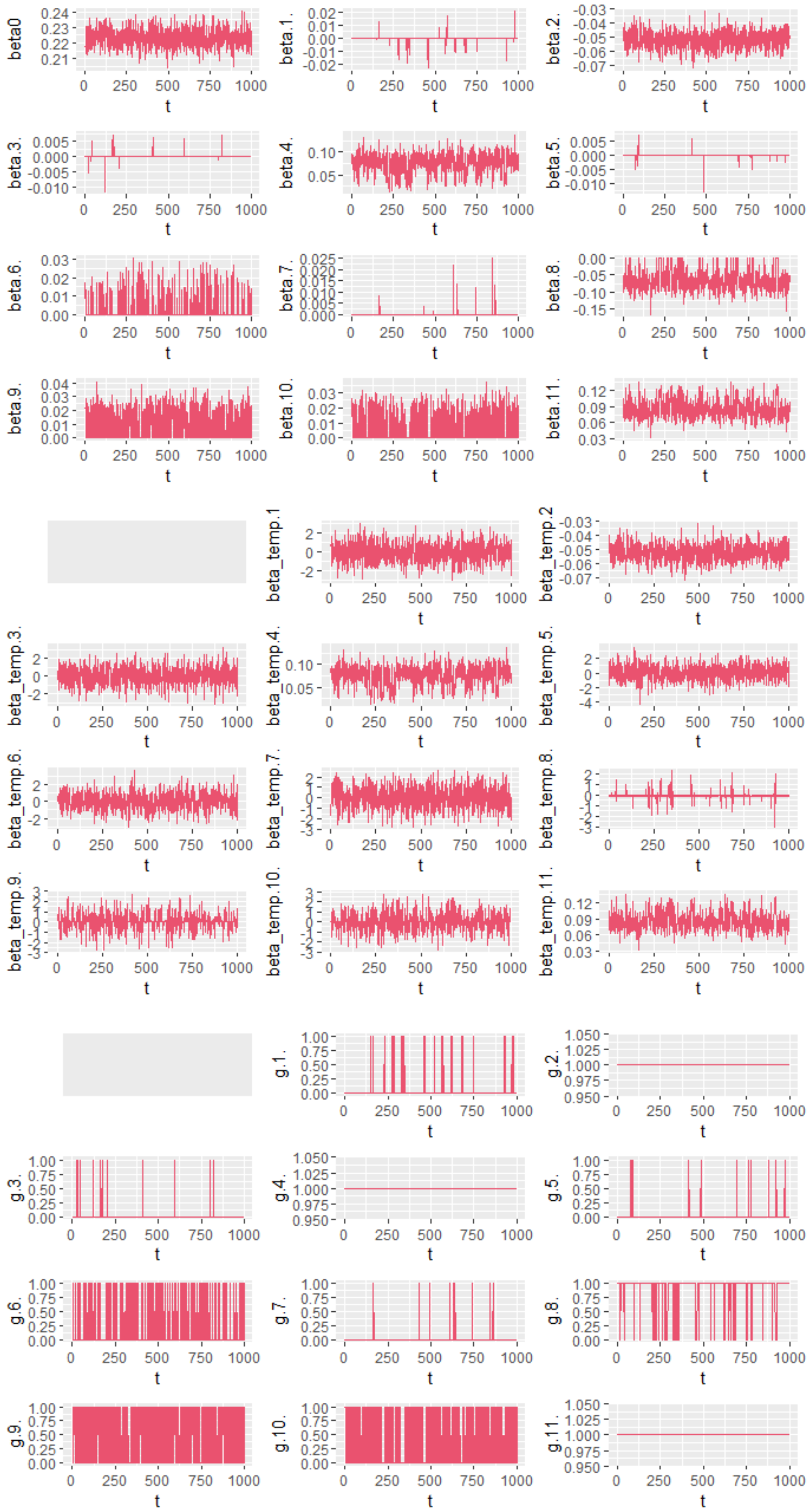

Figure 25: Trace plots for for parameters with Spikes & Slab prior and Binomial likelihood.

Figure 26: Densities for for parameters with Spikes & Slab prior and Binomial likelihood.

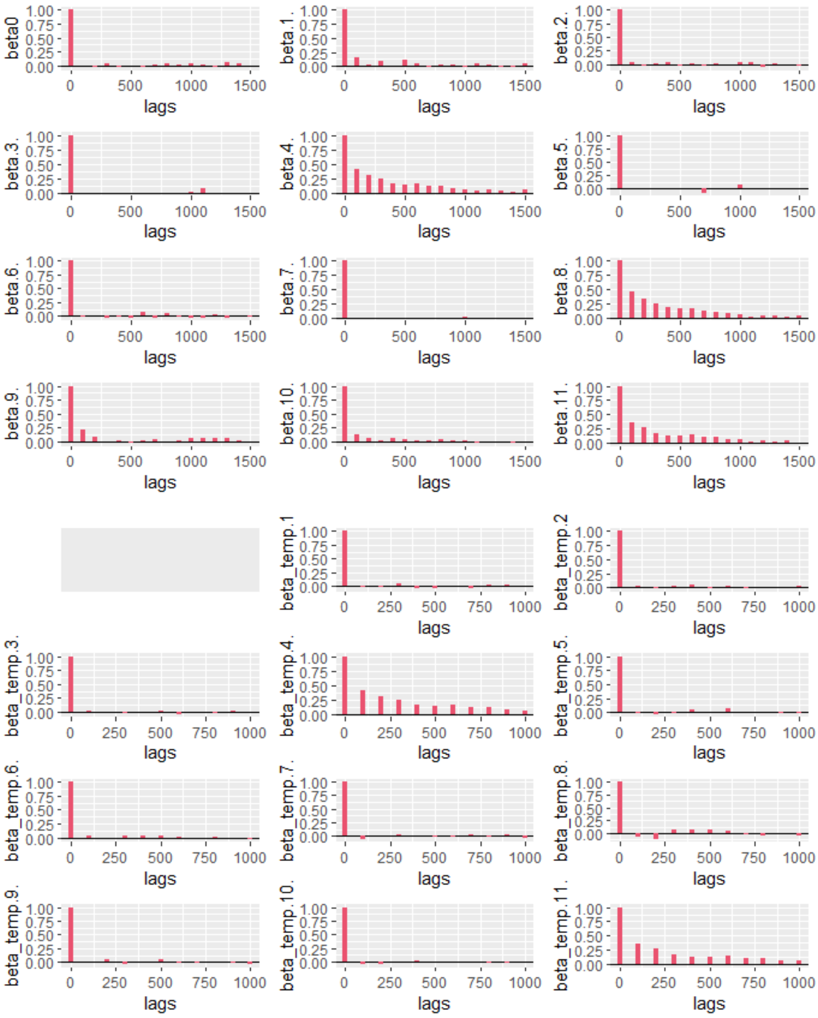

Figure 27: Auto correlation plots for for parameters with Spikes & Slab prior and Binomial likelihood.