Comparing Priors in Bayesian Logistic Regression for Sensorial Classification of Rice
Comparing Priors in Bayesian Logistic Regression for Sensory Classification of Rice

Geiziane Oliveira¹; George von Borries²; Priscila Zackzuk Bassinello³.

1. SAS Institute Inc., Brazil; 2. University of Brasília, Brazil; 3. Embrapa Rice and Beans, Brazil.

INTRODUCTION

Visual characteristics of rice grain are important in determination of quality and price of cooked rice. Stickiness is a visual characteristic of texture that is typically measured by an expensive and time consuming sensorial analysis of cooked rice. Here, PROC MCMC is used to automatically estimate rice stickiness (sticky or loose) with a Bayesian binomial logistic regression model applied to linear combinations of five viscosity measurements of cooked rice. The priors used in the Bayesian modelling were based on four different suggestions of literature: (1) a noninformative prior; (2) a default prior (Gelman et al., 2008); (3) a prior based on odds ratio (Sullivan and Greenland, 2012); and (4) power priors (Ibrahim and Chen, 2000). Then, the quality of each adjustment is evaluated by comparison of agreement between sensorial and predictive (model) classification of stickiness.

MATERIAL AND METHODS

Working data set (years and 2014):

<table>
<thead>
<tr>
<th>SAMPLE</th>
<th>YEAR</th>
<th>PEGAJB</th>
<th>C1</th>
<th>C2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2013</td>
<td>1</td>
<td>0.27599</td>
<td>-0.52857</td>
</tr>
<tr>
<td>2</td>
<td>2013</td>
<td>1</td>
<td>0.09544</td>
<td>-0.54368</td>
</tr>
<tr>
<td>3</td>
<td>2013</td>
<td>1</td>
<td>0.16147</td>
<td>-0.62299</td>
</tr>
<tr>
<td>4</td>
<td>2013</td>
<td>1</td>
<td>0.36966</td>
<td>-0.28279</td>
</tr>
<tr>
<td>5</td>
<td>2013</td>
<td>1</td>
<td>0.32899</td>
<td>-0.39975</td>
</tr>
<tr>
<td>6</td>
<td>2013</td>
<td>1</td>
<td>0.31115</td>
<td>-0.50216</td>
</tr>
<tr>
<td>7</td>
<td>2013</td>
<td>0</td>
<td>-1.12638</td>
<td>1.21886</td>
</tr>
<tr>
<td>8</td>
<td>2013</td>
<td>1</td>
<td>-1.09412</td>
<td>0.80034</td>
</tr>
<tr>
<td>9</td>
<td>2013</td>
<td>0</td>
<td>-1.08571</td>
<td>1.31549</td>
</tr>
<tr>
<td>10</td>
<td>2013</td>
<td>1</td>
<td>-1.09369</td>
<td>0.90455</td>
</tr>
</tbody>
</table>

The data has 144 samples from harvesting of rice for years 2013 (n = 72) and 2014 (n = 72). PEGAJB is the sensory classification of stickiness provided by researchers from EMBRAPA (0 = sticky; 1 = loose). Variables C1 and C2 are scores obtained from the first two principal components obtained from viscosity measures: apparent amylose content, gelatinization temperature, peak, breakdown and final viscosity.

BAYESIAN LOGISTIC REGRESSION

Stickiness as a response $Y_i$ with a binomial distribution,

$$y_i | p_i \sim \text{binomial}(n_i, p_i),$$

where $p_i$ is the success probability (sticky) and links to regression covariates $C_1$, $C_2$ through a logit transformation

$$\text{logit}(p_i) = \log \left( \frac{p_i}{1-p_i} \right) = \beta_0 + \beta_1 C_1 + \beta_2 C_2.$$ 

Once a prior distribution is defined for a parameter $\beta_1$, $p(\beta_1)$, the posterior distribution is found by

$$p(\beta_1 | y) = \frac{L(\beta_1 | y) p(\beta_1)}{\int L(\beta_1 | y) p(\beta_1) d\beta_1},$$

with $L(\beta_1 | y)$ the likelihood of $\beta_1$. The priors on $\beta_0$, $\beta_1$ and $\beta_2$ were defined according to four different prior distributions suggested in literature:

**NONINFORMATIVE PRIOR:** A noninformative prior distribution is proportional to the likelihood function, resulting in low (or none) impact in parameters of the posterior distribution. One common prior distribution is given by the uniform distribution.

**INFORMATIVE PRIOR A:** Sullivan and Greeland (2012) suggest the creation of a normal prior for each $\beta$ based on its corresponding odds ratio (Kleinbaum and Klein, 2010) obtained from original data.

**INFORMATIVE PRIOR B:** according to Gelman et al. (2014) propose independent Student-$t$ prior distribution for parameters of classical logistic regression models. The priors are applied after scaling all no binary variables to have mean 0 and standard deviation 0.5. In this study we used $C_1, C_2$ with mean 0 and corresponding variance. A conservative choice is to use independent Cauchy with center 0 and scale 10 to the constant term ($\beta_0$) and Cauchy with center 0 and scale 2.5 to each of the coefficients ($\beta_1, \beta_2$) in the logistic regression.

POWER PRIORS: Ibrahim et al. (2015) suggested priors to be constructed from historical data. The formulation of the power prior is

$$\pi(\beta | a_0, a_1) \propto L(\beta | D_0)^{a_0} \pi_0(\beta | c_0),$$

with $c_0$ the hyperparameter for a initial prior ($\pi_0$) and $a_0$ is a scalar parameter (0 $\leq a_0 \leq 1$) that controls the influence of the historical data on weights the historical data relative to the likelihood of the current study ($L(\beta | D_0)$). When $a_0 = 1$, the posterior distribution is based completely on historical data. However, if $a_0 = 0$, the prior does not depend on historical data at all. Data from 2013 was used as historical data.

CLASSIFICATION ERROR

The apparent error rate (APR) was obtained as (Johnson and Wichern, 2007),

$$\text{APR} = \frac{n_d}{n},$$

where $n_d$ representing the number of discordances between researcher classification and model prediction and $n$ the total number of samples. The result was used as the main indicator of goodness of fit for each model.
Comparing Priors in Bayesian Logistic Regression for Sensorial Classification of Rice
Geiziane Oliveira¹; George von Borries²; Priscila Zackzuk Bassinello³.
1. SAS Institute Inc., Brazil; 2. University of Brasília, Brazil; 3. Embrapa Rice and Beans, Brazil.

Figure 1: Diagnostics for $\beta_0$.

Material and Methods Continued

Noninformative Prior:

With ODS statement, PROC MCMC produces three main graphics which aid in convergence diagnostic checking: (1) trace plots allow to check whether the mean of the Markov chain has stabilized and appears constant over the graph, and whether the chain has good mixing and is “dense”. The plots show if the chains appear to have reached their stationary distributions; (2) autocorrelation plots indicate the degree of autocorrelation for each of the posterior samples, with high autocorrelations indicating slow mixing; (3) kernel density plots estimate the posterior marginal distributions for each parameter. Figure 1 shows the diagnostic plots for parameter $\beta_0$ using priors $\beta_0 \sim U(-15,15)$, $\beta_1 \sim U(-15,15)$, and $\beta_2 \sim U(-15,15)$.

Power Priors:

SAS Implementation

The PARMS statement declares model parameters and optional starting values. PRIOR statement specifies prior distribution for each parameter. The programming statement does the logistic transformation and assigns it to P. The MODEL statement specifies that the response variable PEGAJB has distribution of Bernoulli with probability of success P. Finally, the ODS OUTPUT option creates a data set PARMSM1 with the summary statistics from the posterior distribution.

Informative Prior A:

PROC LOGISTIC DATA= RICECP14 DESCENDING;
MODEL PEGAJB = C1 C2;
ODS OUTPUT OddsRatios=ODSRATIOM2(KEEP=LowerCL UpperCL);
RUN;
The Bayesian model uses odds ratio of parameters with significant estimates. Then use same structure of the program for noninformative prior, but with normal priors given by statement PRIOR BETA ~ NORMAL(M,V).

Informative Prior B:

Here the only difference from the noninformative model is the prior distribution to be applied. It was assigned independent Cauchy with center 0 and scale 10 to the constant term ($\beta_0$) and Cauchy with center 0 and scale 2.5 to each of the coefficients ($\beta_1$, $\beta_2$) in the logistic regression. The prior specifications are given by PRIOR BETA ~ NORMAL(M,V), with M the center and V scale of the Cauchy.
Comparing Priors in Bayesian Logistic Regression for Sensorial Classification of Rice

Geiziane Oliveira¹; George von Borries²; Priscila Zackzuk Bassinello³.

1. SAS Institute Inc., Brazil; 2. University of Brasília, Brazil; 3. Embrapa Rice and Beans, Brazil.

The SAS programming statements define the probability to be used in the distribution for the log-likelihood function. LOGPDF function compute the logarithm of a Bernoulli probability density with parameter P applied at PEGAJB. Then, a different data is created considering the results for each weight of the historical data (2013). The GENERAL command in the MODEL statement defines the log-prior function to be used in the model. Note that separate code could be used for each value of A, but this code is faster to process and smaller than using separate codes.

RESULTS

The convergence diagnostic plots were very similar for all parameters and models. The plots indicate evidence of convergence for the chain, as observed in Figure 1 for β₀ with noninformative prior. The APR was the same in all models, only 5.56% of apparent error. From the 72 predicted samples only 3 were classified as loose when it was sticky and 1 was classified as sticky when it was loose. Display 1 has the results observed in all 4 models.

The main difference when using the four implemented priors was in the estimates and standard deviations of posterior densities for each parameter. The informative prior of Sullivan and Greeland (Model 2) resulted in a model with only two parameters (β₀, β₁) with lower standard error for each parameter (Figure 2). The result observed with the power prior with α₀ = 1, the one the posterior distribution is based completely on historical data, was the closest to the one presented by Model 2 (Figure 3). If one is interested in parameter estimates and interpretation, then we recommend the use of these models. This study focuse only on quality of stickiness prediction, and for this purpose all models had the same performance.

Figure 2: Posterior Mean and Credibility Intervals (α = 5%) for parameters of models with noninformative prior (1), informative prior A (2), informative prior B (3).

Figure 3: Posterior Mean and Credibility Intervals (α = 5%) for parameters of power priors models (α₀ = 4).

CONCLUSIONS

Sensorial prediction/classification of Rice Stickiness is very efficient when using Bayesian Logistic Regression with PROC MCMC. The use of different priors suggested in literature did produced the same quality of adjustment with difference only in the number of parameters used and size of credibility intervals. If one is interested in exploring parameters properties then we suggest the use of informative prior of Sullivan and Greeland (2012) or power prior based completely on historical data, if available. PROC MCMC is easy to use, fast to compute and very well documented.

MAIN REFERENCES


SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.