On Bayesian high-dimensional regression with binary predictors: a simulation study

La regressione Bayesiana con previsori binari in contesti ad alta dimensionalità: uno studio di simulazione

Debora Slanzi, Valentina Mameli and Irene Poli

Abstract Aim of this work is to develop a comparative analysis to evaluate the performances of several Bayesian regression approaches in the high-dimensional context where the number of observations is very small with respect to the number of predictors. Moreover in this study we assume that the predictors can be expressed only as binary variables coding the presence or the absence of a particular characteristic of the system. This binary structure is very present in many real studies, in particular in laboratory experimentation.

Abstract *Lo scopo di questo lavoro è quello di sviluppare un'analisi comparativa per valutare il comportamento di alcuni metodi inferenziali di regressione Bayesiana in contesti di alta dimensonalità dove il numero di osservazioni è molto piccolo rispetto al numero dei predittori assunti per il modello. Lo studio considera solo predittori espressi in forma di variabili binarie in grado di codificare la presenza e l'assenza di una particolare caratteristica del sistema. Questa struttura del problema presente in molti studi di fenomeni reali e in particolare in ambito sperimentale.*

Key words: Bayesian regression, Binary predictors, Comparative analysis, High-dimensionality.

Debora Slanzi
Department of Management, Ca' Foscari University of Venice, San Giobbe, Cannaregio 873, Venice (IT) and European Centre for Living Technology, S. Marco 2940, Venice (IT), e-mail: debora.slanzi@unive.it

Valentina Mameli
European Centre for Living Technology, S. Marco 2940, Venice (IT) e-mail: valentina.mameli@unive.it

Irene Poli
Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, via Torino 155, Mestre (IT) and European Centre for Living Technology, S. Marco 2940, Venice (IT), e-mail: irenpoli@unive.it

1 Introduction

Bayesian regression models have been widely studied and adopted in the statistical literature [14, 10]. Many studies regard the development of efficient and effective priors to select the set of relevant variables and derive accurate posterior predictive distributions [6, 4]. Moreover in the context of high-dimensionality, when there are many predictors, sparsity is assumed and many parameters can be set to values very close to zero without affecting the fit of the model [11, 9]. The Bayesian penalized regression techniques for the analysis of high-dimensional data include, among others, the Bayesian Lasso [8, 7], the normal-gamma regression [5], the horse-shoe regression [1] and the Bayesian ridge regression [3, 12]. Generally the setup of the regression considers the standard multiple linear model assuming independent Normal error terms. Moreover it is usual to standardize both the response and the covariates to have zero mean and variance equals to one. While there are several studies conducted to compare the performances of the models when the predictors are continuous, these approaches are not very suited when the predictors are binary variables. This situation frequently occurs in many experimental fields, as for example in biochemical studies where the presence and absence of a component determines the results of the experimentation and affects the success of the study. In this paper we focus on this particular situation, and we conduct a simulation study to compare the performance of several high-dimensional Bayesian regression models when the predictors are expressed as binary variables.

The paper is organized as follows. In Section 2 we present the Bayesian multivariate regression model and we introduce the prior distributions considered in the analysis. Section 3 describes the characteristics of the simulation study and presents the results of the comparison by means of indicators of goodness of fitting and prediction. Finally in Section 4 we derive some concluding remarks.

2 Bayesian regression

Let consider the standard multiple linear regression model which assumes that a vector of responses $y = (y_1, y_2, \ldots, y_n)$ can be represented as

$$y = \alpha \mathbf{1} + \mathbf{X}\beta + \varepsilon$$

where the vector of errors $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$ are independent with $p(\varepsilon_i) = N(\varepsilon_i|0, \sigma^2)$ and $\mathbf{X}$ is an $n \times p$ matrix of predictor variables. The scalar α is the intercept, $\mathbf{1}$ a $n \times 1$ unit vector and the vector β represents the regression coefficients. In this work we adopt the Bayesian inferential approach which involves a choice of the prior distribution of the $(p \times 1)$-dimensional vector of regression coefficients β. Many approaches are proposed in literature to derive effective and efficient prior distributions with different characteristics. Among them, the sparsity inducting priors are commonly applied in the setting of high-dimensionality, where most of the predictors are

assumed to be unassociated with the responses. The hierarchical representation of the full Bayesian regression model as proposed by [8], introduces the distributions of the parameters and the hyperparameters as follows:

$$y|\alpha,\mathbf{X},\beta,\sigma^2 \sim N_n(\alpha\mathbf{1}+\mathbf{X}\beta,\sigma^2\mathbf{I}_n),$$
$$\beta_j|\lambda_j^2,\sigma^2 \sim N(0,\lambda_j^2\sigma^2),$$
$$\sigma^2 \sim \pi(\sigma^2)d\sigma^2,$$
$$\lambda_j^2 \sim \pi(\lambda_j^2)d\lambda_j^2.$$

Different priors for σ^2 lead to different regression structures in terms of error distribution, while the hyperparameters $\lambda_1,\lambda_2,\ldots,\lambda_p$ are used to model the sparsity characteristics and control the amount of shrinkage in the coefficient estimates [11]. Usually σ^2 follows an improper prior distribution proportional to $1/\sigma^2$, while the distribution assumed for $\lambda_1,\lambda_2,\ldots,\lambda_p$ leads to different prior distributions for the regression coefficients β. Therefore, depending on the particular choices for the local shrinkage hyperparameters $\lambda_1,\lambda_2,\ldots,\lambda_p$ we can consider some of the most frequently used Bayesian regression models:

- *Bayesian Lasso regression*: the hyperparameters $\lambda_1,\lambda_2,\ldots,\lambda_p$ follow a joint exponential prior distribution which depends on further hyperparameters. Generally, this assumption is simplified by assuming that $\lambda_j^2 \sim Exp(1)$, $i = 1,\ldots,p$ [8, 7];
- *Horseshoe regression*: the prior distribution for the local shrinkage hyperparameters $\lambda_1,\lambda_2,\ldots,\lambda_p$ is the zero-mean half-Cauchy distribution [1];
- *Normal-Gamma regression*: the hyperparameters $\lambda_1,\lambda_2,\ldots,\lambda_p$ follow a Gamma distribution where both shape and scale parameters have an associated prior distribution. Then the marginal distribution of β_j is generally affected by these choices in a way that smaller value of shape parameter of the Gamma distribution is associated with larger amount of shrinkage for the betas [5];
- *Bayesian Ridge regression*: it can be obtained by assuming $\lambda_1^2 = \lambda_2^2 = \ldots = \lambda_p^2 = \lambda^2$ [3].

3 A comparative simulation study

We conduct a comparative simulation study to evaluate the performance of the Bayesian regression models under different prior distributions: Bayesian Lasso, the Bayesian horse-shoe regression, the Bayesian normal-gamma regression and the Bayesian ridge. The study has been conducted considering only binary variable predictors. The statistical analyses were performed using the R-project free software environment for statistical computing. In particular, we use the R-package `monomvn` to fit the regression models [13].

3.1 Experimental setting

The simulation is based on the linear regression model $y_i = X_i\beta + \varepsilon_i$, $i = 1,\ldots,n$, where $\varepsilon_i \sim N(0,1)$. Predictors are generated independently from a Bernoulli distribution with probability of success 0.1 to represent sparsity condition[1]. The simulation considers an increasing number of predictors, p =200, 500, 1000, 1500, 2000, 3000. The number of non-zero regression coefficients is assumed to be $p^\star = 10$ with values $\{-1,1\}$.

From this data generative process, we simulate $N = 2000$ observations (the full dataset) and we randomly select n =100, 200 and 500 sample points (corresponding to the 5%, 10% and 25% of the full space) as training set to estimate the models. The remaining data are considered as test set on which to evaluate the performance of the various Bayesian methods. Each simulation is repeated 50 MonteCarlo runs to evaluate the robustness of the approaches and we compute the Predictive Mean Square Error (PMSE) and the Sensitivity (the ratio between the number of selected important variables and the number of actual important variables) as defined in [2]. In particular, values of Sensitivity close to 1 means that the approach is able to select the relevant information for the regression.

3.2 Comparative results

The results for increasing number of predictors are presented, i.e. p= 200, 500, 1000, 1500, 2000, 3000. Note that for n=200, we did not run simulations for p=200, and for n=500 we did not run simulations for p=200 and 500. In Table 1 we report the evaluation of the prediction capacity by means of PMSE for the different regression models. We can see that with regard to the predictive power of the models, they perform almost in the same way producing accurate predictions in particular when the number of covariates don't exceed too much the number of observations. This predictive power tends to decrease when the ratio between the number of observations and variables tends to decrease. We notice that there is the same trend for all the different approaches, however, the Bayesian Ridge regression presents the poorest performance for all the different values of n here considered. A different situation emerges if we consider the sensitivity measure by which the power of selecting important variables of the simulation is expressed. In Figure 1 we show how this measure evolves through the increasing values of p assuming different values of n. We notice that again the approaches show a very similar performance, but the Bayesian Ridge regression presents values of Sensitivity higher than the other regressions. Therefore, comparing the different indicators of performance of these regressions we see that they all have a good prediction accuracy but the Bayesian Ridge regression presents an higher capacity (Sensitivity) to detect the relevant vari-

[1] We plan to develop the simulation also for other usual values of probability of success representing sparsity condition.

Table 1 Predictive Mean Square Errors for regression: BL=Bayesian Lasso; HS= Horseshoe regression; NG= Normal-Gamma regression; BR= Bayesian Ridge (standard errors of MonteCarlo runs in parentheses).

n	p	BL	HS	NG	BR
100	200	1.52 (0.10)	1.55 (0.14)	1.52 (0.10)	1.81 (0.15)
100	500	1.80 (0.11)	1.86 (0.16)	1.81 (0.11)	1.85 (0.16)
100	1000	1.88 (0.09)	1.91(0.14)	1.88 (0.09)	1.87 (0.13)
100	1500	1.99 (0.07)	2.00 (0.10)	1.99 (0.09)	2.08 (0.11)
100	2000	1.87 (0.06)	1.87 (0.07)	1.89 (0.10)	1.99 (0.11)
100	3000	1.99 (0.05)	1.98 (0.05)	1.99 (0.07)	2.20 (0.14)
200	500	1.40 (0.13)	1.42 (0.17)	1.41 (0.14)	1.53 (0.15)
200	1000	1.54 (0.16)	1.57 (0.15)	1.54 (0.14)	1.73 (0.30)
200	1500	1.75 (0.20)	1.77 (0.19)	1.75 (0.17)	1.94 (0.24)
200	2000	1.78 (0.10)	1.77 (0.11)	1.78 (0.10)	2.06 (0.22)
200	3000	1.93 (0.10)	1.91 (0.11)	1.92 (0.12)	2.28 (0.23)
500	1000	1.15 (0.07)	1.12 (0.09)	1.13 (0.08)	1.20 (0.09)
500	1500	1.35 (0.15)	1.34 (0.17)	1.35 (0.16)	1.36 (0.18)
500	2000	1.54 (1.13)	1.52 (0.14)	1.55 (0.14)	1.60 (0.16)
500	3000	1.80 (0.11)	1.76 (0.15)	1.79 (0.16)	1.85 (0.19)

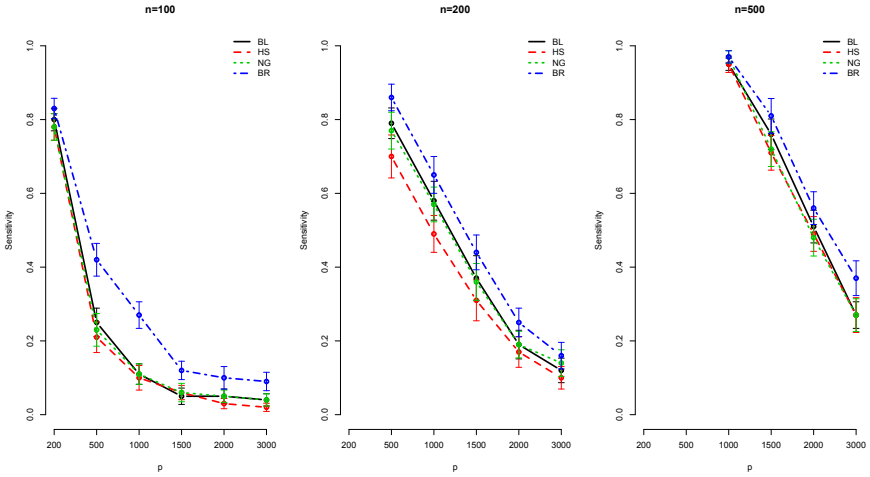

Fig. 1 Sensitivity measures for regression: BL=Bayesian Lasso; HS= Horseshoe regression; NG= Normal-Gamma regression; BR= Bayesian Ridge.

ables of the system. The results of this simulation can be helpful when choosing the structure of the regression model to adopt in a particular study.

4 Concluding remarks

In this work we have developed a comparative analysis to study the performance of several Bayesian regressions with binary predictors in terms of predictive accuracy and variables selection. Further analyses will be conducted to strengthen these preliminary results and to identify the relation between n and p in deriving reliable inferential results in particular when binary predictors are present in the model.

References

1. Carvalho, C., Polson, N., Scott, J. The horseshoe estimator for sparse signals. Biometrika **97**, 465–480 (2010)
2. Geng,Z.,Wang,S.,Yu,M.,Monahan,P.O.,Champion,V.,Wahba,G.: Group variable selection via convex log-exp-sum penalty with application to a breast cancer survivor study. Biometrics **71(1)**, 53–62 (2015)
3. Geweke, J.: Variable selection and model comparison in regression. In: Bernardo,J.M. et al. (eds.) Bayesian Statistics, pp. 609-620. Oxford Press (1996)
4. Griffin, J.E., Brown, P.J.: Hierarchical Shrinkage Priors for Regression Models. Bayesian Analysis **12(1)**,135–159 (2017)
5. Griffin, J.E., Brown, P.J.: Inference with Normal-Gamma prior distributions in regression problems. Bayesian Analysis **5**, 171–188 (2010)
6. Griffin, J.E., Brown, P.J.: Some priors for sparse regression modelling. Bayesian Analysis **8**, 691–702 (2013)
7. Hans, C.: Bayesian lasso regression. Biometrika **96**, 835–845 (2009)
8. Park, T., Casella, G. The Bayesian Lasso. Journal of the American Statistical Association **103**, 672–680 (2008)
9. Piironen, J., Vehtari, A.: Sparsity information and regularization in the horseshoe and other shrinkage priors. Electrononic Journal of Statististics **11(2)**, 5018–5051 (2017)
10. Piironen, J., Vehtari, A.: Comparison of Bayesian predictive methods for model selection. Statistical Computing **27**, 711–735 (2017)
11. Polson, N.G., Scott, J.G.:Local shrinkage rules, L?evy processes and Regularized Regression. Journal of the Royal Statistical Society, Series B **74**, 287–311 (2012)
12. Polson, N.G., Scott, J.G., Windle, J.: The Bayesian bridge. Journal of the Royal Statistical Society: Series B **76**, 713–733 (2014)
13. Robert B., Gramacy: monomvn: Estimation for Multivariate Normal and Student-t Data with Monotone Missingness. R package version 1.9-7 (2017)
14. Sinay, M.S., Hsu, J.S.J.: Bayesian Inference of a Multivariate Regression Model. Journal of Probability and Statistics, vol. 2014, Article ID 673657, 13 pages (2014)