UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

Bayesian variable selection using data driven reversible jump: an application to schizophrenia data

Djidenou Hans Amos Montcho

Dissertação de Mestrado do Programa Interinstitucional de Pós-Graduação em Estatística (PIPGEs)

ICMC USP
SÃO CARLOS

Djidenou Hans Amos Montcho

Bayesian variable selection using data driven reversible jump: an application to schizophrenia data

Master dissertation submitted to the Institute of Mathematics and Computer Sciences – ICMC-USP and to the Department of Statistics – DEs-UFSCar, in partial fulfillment of the requirements for the degree of the Master Interagency Program Graduate in Statistics. *FINAL VERSION*

Concentration Area: Statistics

Advisor: Prof. Dr. Luis Aparecido Milan
Co-advisor: Profa. Dra. Daiane Aparecida Zuanetti

USP – São Carlos
January 2022

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

```
        Montcho, Djidenou Hans Amos
M771b      Bayesian variable selection using data driven
        reversible jump: an application to schizophrenia
        data / Djidenou Hans Amos Montcho; orientador Luis
        Aparecido  Milan; coorientadora Daiane Aparecida
        Zuanetti. -- São Carlos, 2022.
           73 p.


           Dissertação (Mestrado - Programa
        Interinstitucional de Pós-graduação em Estatística) --
        Instituto de Ciências Matemáticas e de Computação,
        Universidade de São Paulo, 2022.


           1. inferência bayesiana. 2. reversible jump . 3.
        seleção de variáveis. 4. mcmc. I. Milan, Luis
        Aparecido , orient. II. Zuanetti, Daiane Aparecida
        , coorient. III. Título.
```

Bibliotecários responsáveis pela estrutura de catalogação da publicação de acordo com a AACR2:
Gláucia Maria Saia Cristianini - CRB - 8/4938
Juliana de Souza Moraes - CRB - 8/6176

Djidenou Hans Amos Montcho

Seleção Bayesiana de variáveis usando o algoritmo de saltos reversíveis direcionado pelos dados: uma aplicação a dados de esquizofrenia

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP e ao Departamento de Estatística – DEs-UFSCar, como parte dos requisitos para obtenção do título de Mestre em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística. *VERSÃO REVISADA*

Área de Concentração: Estatística

Orientador: Prof. Dr. Luis Aparecido Milan
Coorientadora: Profa. Dra. Daiane Aparecida Zuanetti

USP – São Carlos
Janeiro de 2022

UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa Interinstitucional de Pós-Graduação em Estatística

Folha de Aprovação

Defesa de Dissertação de Mestrado do candidato Djidenou Hans Amos Montcho, realizada em 17/12/2021.

Comissão Julgadora:

Prof. Dr. Luis Aparecido Milan (UFSCar)

Prof. Dr. Thierry Chekouo Tekougang (UCalgary)

Prof. Dr. Erlandson Ferreira Saraiva (UFMS)

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa Interinstitucional de Pós-Graduação em Estatística.

This work is dedicated to my parents Pulcherie and Mathias who are the best examples of sacrifice, dedication and inspiration that I know. I am grateful for the gift to explore the world through education.

ACKNOWLEDGEMENTS

I would like to seize this opportunity to thank all those people that supported me throughout these years of my master studies.

First, I thank my advisors Luis Aparecido Milan and Daiane Aparecida Zuanetti for their time, knowledge, patience, for being so supportive when I suggested this research topic and for allowing me to pursue this path. I was very lucky to have them as supervisors and I hope I could be a good advisor as they were for me. I would also like to thank the committee members Thierry Chekouo, Rafael Izbicki and Erlandson F. Saraiva for their time and contributions to this work that for sure improved the final results.

I owe a special debt of gratitude to Rafael Izbicki who introduced me the area of Statistical Machine learning, who broadened my perspective about research in statistics and also guided me on other research topics.

A Special Thanks goes to Dulcemira Mendes and her family for receiving me in São Carlos during my studies. Last but not least, I thank Mario Retamoso, my brothers Dietrich, Ulrich, Méléda, Gilles, for being such an emotional support and a special thank to my partner in life Caroline Porto Crisel for all her patience and support during this journey.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

"Humanity has already sacrificed too much in name of economic growth. Modern works need to be a sphere of self-realization and not an undesired sacrifice to be made just for an income."

RESUMO

MONTCHO, D. H. A. **Seleção Bayesiana de variáveis usando o algoritmo de saltos reversíveis direcionado pelos dados: uma aplicação a dados de esquizofrenia**. 2022. 73 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2022.

Diagnósticos médicos baseados em sintomas são conhecidos por suas limitações, especialmente no entendimento de distúrbios complexos como esquizofrenia. Abordagens modernas e complementares para predizer o risco de tais doenças integram dados genômicos e cerebrais. Nesta monografia, nosso objetivo é inferencial e preditivo. Na inferência, com base em dados de ressonância magnética funcional e de polimorfismo de nucleotídeo único obtidos de pessoas saudáveis e diagnosticadas com esquizofrenia, utilizamos um modelo probito Bayesiano para selecionar as variáveis mais importantes a fim de discriminar os pacientes. Para estimar o risco preditivo, os modelos mais promissores são combinados usando a ponderação bayesiana de modelos. Para estas finalidades, propomos o algoritmo de saltos reversíveis orientado pelos dados para realizar a seleção de variáveis, estimação de parâmetros dos modelos e predição para futuros pacientes.

Palavras-chave: Esquizofrenia, Genética, Algoritmo de saltos reversíveis, MCMC, Inferência Bayesiana, Seleção de variáveis.

ABSTRACT

MONTCHO, D. H. A. **Bayesian variable selection using data driven reversible jump: an application to schizophrenia data**. 2022. 73 p. Dissertação (Mestrado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2022.

Symptom based diagnosis are known to be limited specially concerning complex disorders such as schizophrenia. Modern attempts in providing predictive risk for such disease, to assist existing diagnosis tools, integrate genetic and brain information in what is known as imaging genetics. In this monography, our goal is both inferential and predictive. Regarding the inference, given the functional Magnetic Resonance Image and the Single Nucleotide Polymorphisms information of people diagnosed with schizophrenia and healthy people, we use a Bayesian probit model to select discriminating variables, while to estimate the predictive risk, the most promising models are combined using a Bayesian model averaging scheme. For these purposes, we propose an adaptive reversible jump markov chain monte carlo, named data driven reversible jump, for selecting the variables, estimating their effects and the predictive risk for future subjects.

Keywords: Schizophrenia, Genetics, Informed reversible jump, MCMC, Bayesian inference, Variable selection .

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

AIC	Akaike Information Criterion
AUC	Area Under the receiver operating characteristics Curve
BF	Bayes Factor
BIC	Bayesian Information Criterion
BMA	Bayesian model averaging
DDRJ	Data Driven Reversible Jump Markov Chain Monte Carlo
DIC	Deviance Information Criterion
fMRI	functional Magnetic Resonance Imaging
KL	Kullback-Leibler divergence
KW	Kruskal-Wallis
LASSO	Least Absolute Shrinkage and Selection Operator
MCE	Misclassification Error rate
MCIC	Mind Clinical Imaging Consortium
MCMC	Markov Chain Monte Carlo
MH	Metropolis-Hastings
QTL	Quantitative Trait Locus
RJ	Reversible Jump Markov Chain Monte Carlo
ROIs	Regions Of Interest
SNP	Single Nucleotide Polymorphism
WAIC	Widely Applicable Information Criterion or Watanabe Akaike Information Criterion

CONTENTS

CHAPTER 1

INTRODUCTION

Schizophrenia is a multifactorial disease whose etiology and pathophysiology aren't completely elucidated. It affects almost 1% of the population and commonly observed signs and symptoms are hallucinations, delusions, impairment of cognitive functions, disordering of thinking and akinesia (PATEL *et al.*, 2014). Furthermore, until today, there is no medical test for its diagnosis which is symptom-based and some limitations of such methods have been pointed in Jacob (2013). Moreover, early detection of prodromal symptoms could provide good insights for better targeted treatment and possibly prevent functional degradation, delaying or avoiding transition to psychosis (RUHRMANN; SCHULTZE-LUTTER; KLOSTERKÖTTER, 2003). Therefore, the development of new methods or tests that could assist existing medical tools is of great public health relevance.

Modern attempts in this direction integrate neurological and genetic information in what is known as imaging genetics, connectome genetics and variants of these terminologies. Imaging genetics is a growing field of research in which both neuroimaging and genetic dataset are integrated to unravel the impact of genetic variants on brain structure and function. The reaches of such integration are countless, going from better understanding of psychiatric disorders to their prevention. However, this integration brought new statistical challenges because of the complexity and high dimensionality of such dataset, both in number of covariates and size. Nathoo *et al.* (2019) and Pluta *et al.* (2018) provide a comprehensive review of statistical methods and challenges in imaging genetics, respectively.

In this work, we have available functional Magnetic Resonance Imaging (fMRI) and Single Nucleotide Polymorphism (SNP) information on healthy and patients diagnosed with schizophrenia. fMRI was mainly designed to identify brain's response to task by detecting regional neuronal activity captured by blood oxygenation level-dependent (BOLD) variations. Actually, it is at the core of neuroimaging for studying schizophrenia because of its low invasiveness, absence of radiation and relatively high quality resolution. A deeper introduction to fMRI usefulness and statistical analysis can be obtained in Glover (2011), Gur and Gur (2010) and

Lindquist *et al.* (2008). SNPs are substitution of a single nucleotide at a specific position in the genome that occur in at least 1% of the population. They are frequently used in Genome Wide Association Studies (GWAS) to find possible associations to disease and phenotypes (MAH; CHIA, 2007).

The available dataset was collected by the Mind Clinical Imaging Consortium (MCIC) (CHEN *et al.*, 2012) as an effort of deeper understanding of mental disorder and contains both imaging data on activation patterns using fMRI during a sensorimotor task and multiple SNPs allele frequencies which have previously been implicated in schizophrenia on 118 healthy controls and 92 subjects affected by this disorder. The goal of the MCIC study was to identify regions of interest (ROI) in the brain with discriminating activation patterns between cases and controls and relate them to a relevant set of SNPs able to explain these variations, a variable selection problem clearly. More description of the experimental study and pre-processing can be found in Chen *et al.* (2012) and Stingo *et al.* (2013).

Chen *et al.* (2012) used principal and independent component analysis and found evidence of significant association between fMRI and SNPs. Stingo *et al.* (2013) extended this inferential problem and developed an integrative Bayesian hierarchical mixture model and applied it to link brain connectivity, through fMRI, to genetic information from SNPs of healthy and schizophrenic patients. Their method was innovative because it includes a network that captures shared information between connected brain Regions Of Interest (ROIs). They identified ROIs and SNPs with discriminating activation patterns between case and control and used the model as a classifier for future subjects. In Chekouo *et al.* (2016) the objective was both inferential and predictive. The authors developed a Bayesian predictive model that includes ROIs based network and a new network capturing relations between SNPs and ROIs to quantify a subject's risk of being schizophrenic based on fMRI and SNPs information. Auxiliary indicator variables with spike-slab priors and a Bayesian model averaging were used for model selection and prediction, respectively.

In the Bayesian framework, model selection can be done using a variety of techniques. O'Hara and Sillanpää (2009) and Gelman, Hwang and Vehtari (2014) provide a review of some of those methods jointly with worked examples for illustration and comparison. Some of them are based on information criterion such as Akaike Information Criterion (AIC) (AKAIKE, 1973), Bayesian Information Criterion (BIC) (SCHWARZ *et al.*, 1978), Deviance Information Criterion (DIC) (SPIEGELHALTER *et al.*, 2002) and Widely Applicable Information Criterion or Watanabe Akaike Information Criterion (WAIC) (WATANABE, 2013), among others. Other fully Bayesian methodology such as Bayes factor (KASS; RAFTERY, 1995), selection with spike and slab priors (MITCHELL; BEAUCHAMP, 1988; GEORGE; MCCULLOCH, 1997; ISHWARAN; RAO *et al.*, 2005), selection with shrinkage prior (ERP; OBERSKI; MULDER, 2019) and Reversible Jump Markov Chain Monte Carlo (RJ) (GREEN, 1995) are also worth mentioning and will be theme of this work. However, when doing prediction, one does not need to restrict to only one model. Bayesian model averaging (HOETING *et al.*, 1999) takes model's uncertainty into account and compute the prediction averaging over all competing models with weights given by the posterior probability of each model leading to more trustworthy prediction.

In this work, we propose a new Bayesian predictive risk model for schizophrenia based on a sparse set of ROIs and SNPs, selected using an adaptive RJ. Though widely known for its ability in joint model selection and parameters estimation, traditional RJ lacks a straight way of designing efficient proposals for inter and intra models moves. Usually, models are proposed using uniform distribution which is not our best option if the model space is very large, for example when selecting covariates from a set of very large candidates and parameters are sampled from some Gaussian distribution. Furthermore, including information about the target distribution could increase the efficiency of Markov Chain Monte Carlo (MCMC) when compared with methods based on naive, uniform or random walk. For instance, this is done in Hamiltonian Monte Carlo (NEAL, 2011) using information from the gradient. In the special context of RJ, many works have been dedicated to try to overcome these limitations (BROOKS; GIUDICI; ROBERTS, 2003; JAIN; NEAL, 2004; SARAIVA; MILAN, 2012). Recently, Zanella (2020) proposed locally balanced proposals for discrete spaces on top of which Gagnon (2019) also creates another informed RJ.

A special informed RJ strategy proposed in Zuanetti and Milan (2016) and also used in Zuanetti and Milan (2017), named Data Driven Reversible Jump Markov Chain Monte Carlo (DDRJ) makes use of the data to inform about the next best candidate model and has been proposed for Quantitative Trait Locus (QTL) mapping i.e. for categorical covariates. Its name comes from the fact that the candidate proposal is designed to be driven by the data. This methodology leads to a better mixing, improves acceptance ratio and effective sample size. In this work, our main contribution is that we build on top of the DDRJ and extends it to the context where we have numerical or both categorical and numerical covariates. It is also worth mentioning that, the search for the next candidate could be done in parallel, can rely on batches of the dataset and could be spread on multiple threads to accelerate the search. Finally, we also combine the most visited models, using Bayesian model averaging, to create a classifier for future subjects and we compare its performance in terms of misclassification error and area under the receiver operating characteristic curve to our benchmark results in Chekouo *et al.* (2016), LASSO (TIBSHIRANI, 1996) and Random forest (BREIMAN *et al.*, 1984).

Our goals, selecting ROIs and SNPs, and assessing predictive risk for schizophrenia based on functional Magnetic Resonance Imaging (fMRI) and Single Nucleotide Polymorphism (SNPs) information have been reached. Most ROIs 35, 57, 61, 115 and SNP 22 that we selected were in accordance with results from other authors and also known to be related to the disease, even though some new findings ROI 96 and SNPs 32, 61 have been suggested and may be subject of deeper research. Compared to other methodologies as traditional LASSO and Random Forest, in terms of predictive accuracy, the DDRJ also perfoms well when predictions are done using the

Bayesian Model Averaging.

This monography is organized as follow: Chapter 2 reviews some of the existing Bayesian model selection approaches and in Chapter 3 we describe the Bayesian model under consideration to jointly select ROIs and SNPs. In Chapter 4, the DDRJ algorithm for doing efficient proposals and model selection is explained and Chapter 5 shows its efficiency on simulated data. Finally, Chapters 6 and 7 contain a real application to the MCIC dataset, a discussion on results and final considerations, respectively.

CHAPTER

2

A REVIEW OF BAYESIAN MODEL SELECTION

In model selection, we aim at selecting a given model from a set of competing ones. Each model could represent a specific probability distribution or density for the dataset or indexed by a subset of covariates, or the number of components in a mixture model, or even different architectures for a deep neural network (LECUN; BENGIO; HINTON, 2015). For instance, variable (covariates, features) selection aims at picking a subset of important features from a considerably large set of covariates. In classification context, It is equivalent to find the most discriminatory set of covariates whereas, in regression context, it aims at selecting those covariates able to explain a large fraction of the variability in the response variable. At least two reasons could motivate this problem. The first one refers to the underlying sparse nature of the process, in which case there are only a few numbers of covariates that explain the variation in the response variable. The second reason may be that of increasing the predictive capacity of the model using again a small set of variables. In this chapter, we briefly review some Bayesian model selection methods.

For this purpose, consider $Y = (Y_1, Y_2, \ldots, Y_n)$ be a data possibly generated from a set of competing models $M_1, M_2, \ldots, M_K$ indexed by $k \in \{1, \ldots, K\}$ and parameters $\theta_1, \theta_2, \ldots, \theta_K$.

2.1 Bayes Factor

The Bayesian solution to this problem is straight. Compute the posterior probability of each model and choose the one with the largest value. Formally stating, we assign a prior probability $\pi_k = P(M_k)$ to each model M_k. Then, the posterior probability is derived from Bayes theorem as

$$P(M_k|Y) = \frac{\pi_k P(Y|M_k)}{\sum_{k=1}^{K} \pi_k P(Y|M_k)} \tag{2.1}$$

where

$$P(Y|M_k) = \int P(Y,\theta_k|M_k)d\theta_k = \int P(Y|\theta_k,M_k)p(\theta_k|M_k)d\theta_k = \int L_k(\theta_k|Y)P_k(\theta_k)d\theta_k \tag{2.2}$$

with $L_k(\theta_k|Y)$ and $P_k(\theta_k)$ being the likelihood function and prior probability distribution for the parameters, respectively, under model M_k.

As we are only interested in comparing model's posterior probability and due to difficulties for computing the denominator in Equation (2.1), we could take the ratio $R_{kj} = \frac{P(M_k|Y)}{P(M_j|Y)} = \frac{\pi_k}{\pi_j}\frac{P(Y|M_k)}{P(Y|M_j)}$ and decide against model M_j if it is greater than 1. From this ratio, define the Bayes factor (KASS; RAFTERY, 1995), to quantify the increase of evidence from the data against model M_j as

$$BF_{kj} = \frac{P(Y|M_k)}{P(Y|M_j)} = \frac{P(M_k|Y)}{P(M_j|Y)}\frac{\pi_j}{\pi_k}. \tag{2.3}$$

Applicability of Bayes Factor (BF) only depends on being able to compute the integral in Equation (2.2). It is also sensitive to the prior distribution and for improper prior, we may have to deal with the marginalization paradox. Kass and Raftery (1995) suggested the following interpretation for BF_{kj} in Table 1.

Table 1 – Bayes factor interpretation.

$\log_{10}(BF_{kj})$	BF_{kj}	Evidence against M_j
$0-\frac{1}{2}$	1 - 3.2	not worth more than a bare mention
$\frac{1}{2}-1$	3.2 - 10	substantial
$1-2$	10 - 100	strong
>2	> 100	decisive

2.2 Information Criteria: AIC, BIC, DIC, WAIC

Information criterion measures aim, in general, at estimating expected out-of-sample prediction error from an adjusted within-sample error (GELMAN; HWANG; VEHTARI, 2014) and consequently select the best model under a given metric. The metric for selection is really important. For instance, in predictive modelling, one is usually interested in selecting the model with the largest predictive accuracy on future data whereas, in inferential modelling, where the interest lies in depicting or understanding relationships between covariates and the response variable, one is looking for the model that better approximates the real relations given by the true model. Usually, an information criterion is a balance of goodness of fit and a penalty function.

Goodness of fit could be measured by likelihood, deviance, or their expectation and the penalty as a measure of the complexity of the model, i.e., parameters dimension, for instance.

2.2.1 Akaike Information Criterion- AIC

In this section, we will outline the derivation and intuition behind AIC (AKAIKE, 1973) in the statistical context. Consider again a set of K models $M_1, M_2, \ldots, M_K$ representing probability distribution or densities $p_1(y|\theta_1), p_2(y|\theta_2), \ldots, p_K(y|\theta_K)$ and suppose available some dataset $Y = (Y_1, Y_2, \ldots, Y_n)$ from a probability distribution or density $p(y|\theta)$. Define by $\hat{\theta}_j$ the maximum likelihood estimator of θ under model M_j and an estimate for $p(y|\theta)$ is given by $\hat{p}_j(y|\theta) = p(y|\hat{\theta}_j)$. The quality of this estimate could be measured by the Kullback-Leibler divergence (KL) (KULLBACK; LEIBLER, 1951):

$$\begin{aligned} KL(p(y|\theta), \hat{p}_j(y|\theta)) &= \int p(y|\theta) \log\left(\frac{p(y|\theta)}{\hat{p}_j(y|\theta)}\right) dy \\ &= \int [p(y|\theta)\log(p(y|\theta)) - p(y|\theta)\log(\hat{p}_j(y|\theta))]dy. \qquad (2.4) \end{aligned}$$

The best estimate would be the one that minimizes Equation (2.4). The first part of this difference does not depend on $\hat{p}_j(y|\theta)$, then we could forget it and only focus on maximizing

$$KL_j = \int p(y|\theta)\log(\hat{p}_j(y|\theta))dy.$$

An intuitive estimate for KL_j is:

$$\overline{KL}_j = \frac{1}{n}\sum_{i=1}^{n} \log(\hat{p}_j(y_i|\theta)) = \frac{l_j(\hat{\theta})}{n}.$$

However, this estimate is biased and this bias was corrected by Akaike who derived an approximately unbiased estimate of KL_j as

$$\hat{KL}_j = \frac{l_j(\hat{\theta})}{n} - \frac{dim(\theta_j)}{n}$$

and the AIC is finally given by:

$$\mathrm{AIC}_j = -2n\hat{KL}_j = -2l_j(\hat{\theta}) + 2dim(\theta_j). \qquad (2.5)$$

According to KL, and then to AIC, the best model would be that with the smallest AIC. Some caveats are worth mentioning. From this derivation, one can see that AIC was designed to select model looking at predictive accuracy, so it might not select the true model if this were part of our competing models. Moreover, AIC is not Bayesian because we only need the likelihood to compute it. It is in fact a penalized likelihood, with the penalty given by the dimension of the model. Nonetheless, we have decided to include it for its popularity and historical purpose for being the first information criterion to select model. Last but not least, for hierarchical models where the model's dimension is not really clear or for likelihood free methods, AIC becomes impracticable.

2.2.2 Bayesian Information Criterion- BIC

Bayesian Information Criterion-BIC was proposed by Schwarz *et al.* (1978) to be an alternative of AIC adding harsher penalty to the log-likelihood as in AIC. It is given by the formula in Equation (2.6) below

$$\mathrm{BIC}_j = -2l_j(\hat{\theta}) + dim(\theta_j)\log(n). \tag{2.6}$$

BIC's penalty is harsher than AIC due to this $\log(n)$ factor and this penalty increases with the sample size. As one can see, contrary to its name, BIC is not really Bayesian, because the posterior distribution or density is not used for its computation, but this metric is justified under a Bayesian framework. Indeed, one can show that the BIC is consistent i.e. it, asymptotically, approximates the Bayes factor when the prior's effect of the model is negligible compared to the likelihood function and assuming a uniform distribution over the set of models. Thus, the BIC is trying to find the model with the largest posterior probability.

It should be noted that BIC and AIC are solving different problems. AIC is trying to select the model with the largest accuracy or generalization power and BIC is trying to select the true model, if this were part of our candidate models.

2.2.3 Deviance Information Criterion- DIC

This is probably the most used criterion for model selection in Bayesian context. The main reasons are that it is easy to compute and it is available in most standard softwares as JAGS by Plummer *et al.* (2003) and BUGS by Lunn *et al.* (2000). DIC defined by Spiegelhalter *et al.* (2002) also contains two parts. The goodness of fit using the posterior mean deviance

$$\mathbf{E}_{\theta|y}(D(\theta)) = \bar{D}(\theta) = -2\int \log[p(y|\theta)]p(\theta|y)d\theta \tag{2.7}$$

and the complexity term named effective number of parameters

$$\begin{aligned} p_D &= \bar{D}(\theta) - D(E_{\theta|y}[\theta]) \\ p_D &= -2\int \log[p(y|\theta)]p(\theta|y)d\theta + 2\log[p(y|E_{\theta|y})[\theta])]. \end{aligned} \tag{2.8}$$

DIC is defined as the sum of Equation (2.7) and Equation (2.8)

$$\begin{aligned}
\text{DIC} &= \bar{D}(\theta) + p_D \\
&= \bar{D}(\theta) + \bar{D}(\theta) - D(E_{\theta|y}[\theta]) \\
\text{DIC} &= 2\bar{D}(\theta) - D(E_{\theta|y}[\theta]). \qquad (2.9)
\end{aligned}$$

A Monte Carlo approximation can be used to compute the DIC using a sample from the posterior distribution obtained via any mechanism, Markov Chain Monte Carlo (MCMC) for instance. Again, DIC intends to be a generalization of AIC and according to this metric, the best model should be the one with smallest DIC.

2.2.4 Watanabe Akaike Information Criterion- WAIC

WAIC also known as Widely Applicable Akaike Information Criterion was proposed by Watanabe and Opper (2010) as a fully Bayesian information criterion that generalizes AIC. Again, It is a balance of goodness of fit and model complexity. The goodness of fit is computed by the log pointwise predictive density (lppd)

$$\begin{aligned}
\text{lppd} &= \log \prod_{i=1}^{n} \int p(y_i|\theta) p(\theta|y) d\theta \\
&= \sum_{i=1}^{n} \log \left[\int p(y_i|\theta) p(\theta|y) d\theta \right] \qquad (2.10)
\end{aligned}$$

and one of the complexity measure p_{waic} is

$$p_{waic} = \sum_{i=1}^{n} Var_{\theta|y} \left[\log(p(y_i|\theta)) \right]. \qquad (2.11)$$

Finally the WAIC is defined by a combination of Equation (2.10) and Equation (2.11) as:

$$\begin{aligned}
\text{WAIC} &= -2\text{lppd} + 2p_{waic} \qquad (2.12) \\
&= -2 \sum_{i=1}^{n} \log \left[\int p(y_i|\theta) p(\theta|y) d\theta \right] + 2 \sum_{i=1}^{n} Var_{\theta|y} \left[\log(p(y_i|\theta)) \right].
\end{aligned}$$

Gelman, Hwang and Vehtari (2014) advocates that WAIC should be preferred to AIC, BIC and DIC not only for the fact that it is fully Bayesian, but also for being an approximation to cross-validation even when the Fisher information matrix, usually positive definite, has zero eigenvalues. Moreover, Watanabe (2013) also introduced WBIC (Widely Bayesian Information Criterion) as a generalization of BIC. The author argues that WAIC should be used when searching for selecting model according to predictive loss whereas WBIC should be preferred when looking for the true model.

There is a plethora of information criteria that has not been mentioned here, almost all of them balancing goodness of fit and complexity. From this section, it should be clear that, information criterion measures are impracticable when the set of models is large. Imagine yourself having to compute the metric for each model and then comparing them all. A typical example relates to variable selection in high dimensional setting $(p >> n)$ where the number of possible models is at least 2^p if one doesn't include interactions. Definitely, we need other techniques for such problems, subject of the next topics.

2.3 Shrinkage prior

One of the most used method for variable selection under a non Bayesian framework is the Least Absolute Shrinkage and Selection Operator (LASSO) (TIBSHIRANI, 1996). It finds the least squares estimate with an L1 penalty on the parameters. In that seminal work, the author has also provided a Bayesian interpretation of his methodology. Indeed, the Lasso's solution is equivalent to the maximum a posteriori estimate with an independent Laplace or double-exponential prior for each parameter. The result is that covariates with large effects are maintained in the model while those with small effects are shrunk to zero. Therefore, the selected model is the one containing covariates with non-zero coefficients. The same could be said on ridge regression (HOERL; KENNARD, 1970) corresponding to a Bayesian regression with independent normal priors. Such ideas are easily extended to other shrinkage priors leading to Bayesian penalized methods. In this section, we will outline some useful priors for model selection. Most of what follows was inspired by Erp, Oberski and Mulder (2019), O'Hara and Sillanpää (2009) that provide a thorough review of Bayesian penalized methods for variable selection and a general review of Bayesian variable selection, respectively.

A general framework for defining shrinkage priors was proposed by Ishwaran, Rao *et al.* (2005) and will be used here. For this purpose, consider the fully Bayesian regression below

$$\begin{aligned} y_i|\beta, x_i, \sigma^2 &\sim N(x_i\beta, \sigma^2); \quad x_i = (x_{i1}, \ldots, x_{ip});\ i = 1, \ldots, n \\ \beta|\tau &\sim N(0, Diag(\tau)); \quad \beta = (\beta_0, \ldots, \beta_p) \\ \tau &\sim p(\tau) \\ \sigma &\sim p(\sigma). \end{aligned} \tag{2.13}$$

Shrinkage prior are defined as the marginal distribution of β. The main point of Ishwaran, Rao *et al.* (2005) is that one can obtain most common shrinkage prior, i.e. a marginal distribution for β, using the hierarchical framework in Equation (2.13) controlled by the distribution on the hyperparameter τ.

2.3.1 Ridge and Lasso

- Ridge

  The Ridge prior assumes an independent normal distribution for each coefficient. In a hierarchical setup, assuming a uniform distribution for the scale parameter τ and marginalizing over it, leads to a marginal normal distribution for each coefficient, as below

$$\begin{aligned} \beta_j|\tau &\sim Normal(0,\tau), \quad j=1,\dots,p \\ \tau &\sim Unif(0,\lambda) \\ \beta_j &\sim Normal(0,\frac{1}{\lambda}). \end{aligned}$$

- Lasso

  The Lasso prior assumes an independent Laplace prior for each coefficient. In a hierarchical setup, assuming a specific gamma distribution for the scale parameter τ and marginalizing over it, leads to the double exponential distribution, as below

$$\begin{aligned} \beta_j|\tau &\sim N(0,\tau), \quad j=1,\dots,p \\ \tau &\sim Exp\left(\frac{\lambda^2}{2}\right), \quad \lambda>0 \\ \beta_j &\sim Laplace(0,\lambda). \end{aligned}$$

Using the same hierarchical setup and with a bit of algebra one can derive fancier marginal distribution for βs. It is also worth mentioning some priors that have been gaining popularity like the Horseshoe prior by Carvalho, Polson and Scott (2009), the penalizing model complexity priors, PC-priors by Simpson *et al.* (2017) and product moment prior (pMOM) by Johnson and Rossell (2012) used in Chekouo *et al.* (2016) for features selection.

We could not end this section without citing a seminal paper on Bayesian variable selection named Stochastic Search Variable Selection (SSVS) by George and McCulloch (1997) that inspired many modern variable selection techniques. For this purpose, we will briefly review their approach.

2.3.2 Stochastic Search Variable Selection-SSVS

SSVS, which probably has some inspiration in the work from Mitchell and Beauchamp (1988), is a Bayesian methodology for variable selection coined initially in regression context. However, extensions to generalized linear models, time series, etc already exist. The main idea of SSVS is to define a latent indicator variable γ_j for each covariate x_j that controls its presence or absence in the model. Being strict, a covariate is absent from the model if and only if its coefficient β_j equals zero. However and usually, one consider a covariate to be absent if its coefficient lies in a neighborhood of zero with small radius, i.e. having distribution concentrated around 0. This strategy also avoids dealing with models of different dimensions. The authors defined the following hierarchical Bayesian model

$$\begin{aligned}
y_i|\beta,x_i,\sigma^2 &\sim N(x_i\beta,\sigma^2), \quad i=1,\dots,n \\
\beta_j|\gamma_j &\sim (1-\gamma_j)\underbrace{N(0,\tau_j)}_{spike}+\gamma_j\underbrace{N(0,c_j\tau_j)}_{slab} \\
\pi(\gamma=(\gamma_1,\dots,\gamma_p)) &= \prod_{j=1}^{p}\pi_j^{\gamma_j}(1-\pi_j)^{(1-\gamma_j)}, \quad \pi_j=P(\gamma_j=1) \\
\pi_j &\sim \text{Beta}(a,b) \\
\sigma^2|\gamma &\sim \text{Inverse Gamma}(\nu_\gamma/2,\lambda_\gamma\nu_\gamma/2). \qquad (2.14)
\end{aligned}$$

In summary, when $\gamma_j=1$, a slab prior, i.e. a normal distribution with large variance, controlled by c_j is assumed for β_j and this occurs with probability π_j. Otherwise, a spike prior, i.e. a normal concentrated around zero, controlled by τ_j, for β_j and this occurs with probability $1-\pi_j$. The method is sensitive to choice for τ_j and c_j and a dedicated analysis and practical guidelines have been done and suggested by Scott and Berger (2010). Being able to write this mixture of spike and slab prior as a multivariate normal distribution and assuming conjugate prior on π_j and σ^2 permits to derive all full conditional posteriors (normal for β, inverse gamma for σ^2 and binomial for γ) and use a Gibbs sampling scheme to obtain L samples from the posterior distribution. For variable selection, the marginal posterior probability for inclusion of X_j could be estimated by $\hat{P}(\gamma_j|\mathbf{X},Y)=\frac{1}{L}\sum_{l=1}^{L}\mathbb{I}(\gamma_j^l=1)$, and the final model could be containing only those variables with posterior probability for inclusion above a user defined threshold.

2.4 Reversible Jump MCMC

The Reversible Jump Markov Chain Monte Carlo was proposed by Green (1995) as a generalization of the Metropolis-Hastings (MH) algorithm (HASTINGS, 1970; METROPOLIS *et al.*, 1953) for obtaining samples from a distribution on spaces of varying dimensions. In this section we briefly review RJ, we point out its usefulness in model selection context and its main practical limitation. This section was inspired from Lopes (2006) and Fan and Sisson (2011).

Let's reconsider our model selection problem from a set of competing models $M=\{M_1,M_2,\dots,M_k,\dots\}$ indexed by $\{1,2,\dots,k,\dots\}$ and parameters $\theta=\{\theta_1,\theta_2,\dots,\theta_k,\dots\}$. Under model M_k, the joint posterior distribution of θ_k and the model index k is given by:

$$p(\theta_k,k|y)\propto p(k)p(\theta_k|k)p(y|\theta_k,k). \qquad (2.15)$$

RJ is a broader version of MH such that one can sample from the joint distribution in Equation (2.15) where moves can be done through Markov chain between models (k,θ_k) and $(k^{'},\theta_{k^{'}})$ having different dimensions k and $k^{'}$. Suppose that the current state of the Markov chain is (k,θ_k), RJ works as follow

1. Propose to visit model $M_{k'}$ with probability $p(k \to k^{'})$.

   As k may not be equal to $k^{'}$, a dimension matching procedure is needed.

2. Sample $u \sim q(u)_{k \to k'}$ and $u^{'} \sim q(u^{'})_{k' \to k}$ such that $k + dim(u) = k^{'} + dim(u^{'})$.

   The new set of parameters could be obtained from a bijective transformation of (θ_k, u) as follow

3. $(\theta_{k^{'}}, u^{'}) = g_{k,k'}(\theta_k, u)$.

4. Having the new set of parameters, we accept this move $k \to k^{'}$ with probability $min(1, \alpha(k,k^{'}))$, where

$$\alpha(k,k^{'}) = \frac{p(\theta_{k^{'}},k^{'})p(y|\theta_{k^{'}},k^{'})}{p(\theta_K,k)p(y|\theta_{k^{'}},k^{'})} \frac{p(k^{'} \to k)}{p(k \to k^{'})} \frac{q(u^{'})_{k^{'} \to k}}{q(u)_{k \to k^{'}}} \left| \frac{\partial g_{k,k'}(\theta_K,u)}{\partial(\theta_K,u)} \right| \tag{2.16}$$

   where the last term in Equation (2.16) is the Jacobian of $g_{.,.}(.,.)$.

As, one can see there are some tuning functions to be chosen: the proposal to visit the next model, the proposal to match the dimension, and the bijective function to obtain the new set of parameters. Thus, RJ's efficiency strongly depends on them and our attempt with this work is to propose clever way of choosing some of them. Running this procedure L times leads to a sample of (k,θ_k). Any subsequent inference can be done from it. For instance, the posterior probability of model M_k could be estimated by $\hat{p}(M_k|y) = \frac{1}{L}\sum_{l=1}^{L} \mathbb{I}(k_l = k)$, with $\mathbb{I}(k_t = k)$ being the indicator function. Finally, the model with the largest posterior could be the selected one.

2.5 Bayesian Model Averaging

A typical statistical framework will choose the model with the largest probability, estimate functions of parameters, create standard errors, credible intervals and make predictions. However, making decisions relying solely on this model ignores the uncertainty in the model selection. What if we could include this uncertainty in our whole posterior inference? This is exactly what is intended with Bayesian model averaging (BMA). As we saw in the previous section on RJ, a byproduct is the posterior probability of models under evaluation. Thus, these probabilities could be used as weights for further estimation remembering ensembles methods in machine learning.

For instance, a posterior mean of a parameter η could be estimated as $E(\eta|y) = \sum_t \hat{\eta}_t p(M_t|y)$, where $\hat{\eta}_t = E(\eta|y,M_t$. Usually, t the number of visited models could be very large and the equation above would be computationally intensive. To reduce this cost, one could reduce the number of models in certain way. One possible strategy only includes models with posterior probability greater than a threshold c defined by the user and we could approximate the posterior mean of η as $E(\eta|y) \approx \sum_{t:p(M_t|y)\geq c)} \hat{\eta}_t p(M_t|y)$. However, this choice clearly doesn't account for some models, as they are assigned small posterior probability and may impact on prediction.

In summary, in our schizophrenia context where we intend to select ROIs, SNPs, and create a classifier for future subjects, DDRJ will be used to solve the first part of variable selection and BMA will be used for doing prediction. In the next section, we will describe the Bayesian models under consideration. Hoeting *et al.* (1999) provides a great introduction to BMA and a more recent work by Fragoso, Bertoli and Louzada (2018) reviewed thoroughly BMA.

CHAPTER 3

BAYESIAN MODEL DESCRIPTION

Given n individuals, let $Y = (Y_1, \ldots, Y_n)$ be a set of binary random variables characterizing the disease status, healthy or diagnosed with schizophrenia, of an individual. Also define $X = [X_{ij}]_{n\times g}$, $Z = [Z_{ij}]_{n\times m}$ as the matrix of g ROI-based summaries of BOLD intensity and the matrix of m genetic covariates (SNPs), respectively. For our purpose, we consider three Probit models. A first model analyzing the relation between schizophrenia and ROIs, a second model analyzing the SNP's effect on schizophrenia and a last model that considers the joint effect of ROIs and SNPs on schizophrenia. With Model 1 and 2 being a special case, this section will concentrate on describing in detail Model 3 only. Results concerning Model 1 and 2 can be derived with minor adjustments.

3.1 Bayesian Probit Model

To introduce the idea of the Probit model from Albert and Chib (1993), consider the biological process of information, i.e. electric signal, exchange between neurons (LOVINGER, 2008). When an electrical signal reaches a neuron at rest, through the axon, its values must be above a threshold for the neuron to fire, otherwise the neuron will keep resting. Imagine, we only have access to the output (fire - no fire) without knowing the underlying process. The Probit data augmentation approach relies on a latent random variable, unseen electrical signal, normally distributed and classify the output according to its value being above a threshold or not. On the same way, we introduce a latent continuous variable Y_i^*, viewed as a hidden process that depends on ROIs and SNPs, such that when its value is positive, the patient is classified as schizophrenic and healthy otherwise. Opting for the Gaussian distribution to model this underlying process leads us to nice conditional distributions for all other random variables and allows us to use Gibbs Sampling to estimate their parameters. Depending on the distribution of Y_i^*, we can have more elaborate data augmentation method and probably Metropolis-Hasting within Gibbs to sample from the posterior distribution. For instance, one could have followed the data augmentation model proposed in Polson, Scott and Windle (2013) and then use a Bayesian logit model. In summary, our model assume there is a latent variable Y_i^* such that:

$$Y_i^* = \beta_0 + \sum_{p \in \mathscr{G}} \beta_p X_{ip} + \sum_{k \in \mathscr{M}} \alpha_k Z_{ik} + \sum_{k \in \mathscr{M}} \delta_k (1 - |Z_{ik}|) + \varepsilon_i, \quad \varepsilon_i \sim N(0,1), \tag{3.1}$$

$$Y_i = \mathbb{I}(Y_i^* > 0)$$

$$Z_{ik} \in \{-1,0,1\}, \text{ for SNPs having genotype aa, aA, AA, respectively}$$

$$\mathscr{G} = \{ROIs\ in\ the\ model\}, \quad \mathscr{M} = \{SNPs\ in\ the\ model\},$$

$$P = \text{Cardinality}(\mathscr{G}), \quad K = \text{Cardinality}(\mathscr{M}).$$

The goal of this work is to select, under a Bayesian framework, a set of discriminatory ROIs and SNPs indexed by P and K from the set of available ROIs (g) and SNPs (m), respectively. We also aim at providing estimates for the coefficients β_0, β_ps and for the additive and dominant effects α_k, δ_k, respectively for the selected features. Let us denote the unknown parameters by $\theta = (\gamma, K, P)$ with $\gamma = (\beta = (\beta_0, \beta_1, \ldots, \beta_P), \alpha = (\alpha_1, \ldots, \alpha_K), \delta = (\delta_1, \ldots, \delta_K))$. The likelihood function for θ is given by

$$\begin{aligned} L(\theta|Y^*, X, Z) &= \prod_{i=1}^{n} P(Y_i^*|\theta, X, Z) \\ &= \frac{1}{(\sqrt{2\pi})^n} exp^{-\sum_{i=1}^{n} \varepsilon_i^2}, \end{aligned} \tag{3.2}$$

where

$$\varepsilon_i = y_i^* - \beta_0 - \sum_{p=1}^{P} \beta_p x_{ip} - \sum_{k=1}^{K} \alpha_k x_{ik} - \sum_{k=1}^{K} \delta_k (1 - |z_{ik}|).$$

Generally, statisticians use dummy variables to encode categorical variables, and we could have also done the same here. However, in biology the genetic interpretation is easier when the SNPs are encoded as we have done above.

We complete our model assigning independent prior distribution to each parameter and the joint prior distribution is defined as follow:

$$\pi(\theta) = \pi(K)\pi(P) \underbrace{\pi(\beta|P)\pi(\alpha|K)\pi(\delta|K)}_{\pi(\gamma|K,P)}, \tag{3.3}$$

where

$$K \sim Unif(g), \quad P \sim Unif(m), \quad \beta \sim N_{P+1}(0, \sigma_\beta^2 \mathbb{I}), \quad \alpha \sim N_K(0, \sigma_\alpha^2 \mathbb{I}), \quad \delta \sim N_K(0, \sigma_\delta^2 \mathbb{I})$$

with all hyperparameters fixed.

3.2 Conditional Posteriors

Model in Equation (4.7) is a usual regression model in which we are assuming gaussian priors for the coefficients. Hence all conditional posteriors can be easily derived as:

$$\begin{aligned}\pi(\beta|Y^*,X,Z,\alpha,\delta) &= N(\beta^*,\Gamma_1)\\ \beta^* &= \Gamma_1\,[1|X]^{'}\,\{Y^*-Z\alpha^{'}-[1-|Z|]\delta^{'}\}\\ \Gamma_1 &= \left\{\frac{1}{\sigma^2_\beta}\mathbb{I}_{P+1}+[1|X]^{'}[1|X]\right\}^{-1}, \end{aligned} \tag{3.4}$$

$$\begin{aligned}\pi(\alpha|Y^*,X,Z,\beta,\delta) &= N(\alpha^*,\Gamma_2)\\ \alpha^* &= \Gamma_2\,Z^{'}\,\{Y^*-[1|X]\beta^{'}-[1-|Z|]\delta^{'}\}\\ \Gamma_2 &= \left\{\frac{1}{\sigma^2_\alpha}\mathbb{I}_K+Z^{'}Z\right\}^{-1}, \end{aligned} \tag{3.5}$$

$$\begin{aligned}\pi(\delta|Y^*,X,Z,\beta,\alpha) &= N(\delta^*,\Gamma_3)\\ \delta^* &= \Gamma_3[1-|Z|]^{'}\{Y^*-[1|X]\beta^{'}-Z\alpha^{'}\}]\\ \Gamma_3 &= \left\{\frac{1}{\sigma^2_\delta}\mathbb{I}_K+[1-|Z|]^{'}[1-|Z|]\right\}^{-1}, \end{aligned} \tag{3.6}$$

$$\begin{aligned}\pi(Y_i^*|,\theta,y_i,x_i,z_i) &= Nt([1|X_i]\beta^{'}+Z_i\alpha^{'}+(1-|Z_i|)\delta^{'},1,\text{left}=0) \quad \text{if} \quad \text{y}_i=1\\ \pi(Y_i^*|\theta,y_i,x_i,z_i) &= Nt([1|X_i]\beta^{'}+Z_i\alpha^{'}+(1-|Z_i|)\delta^{'},1,\text{right}=0) \quad \text{if} \quad \text{y}_i=0, \end{aligned} \tag{3.7}$$

with $\mathbb{I}_N$ being the identity matrix of size N, Nt being the truncated Normal Distribution and $[1|X]$ being a matrix of dimension $n\times(P+1)$ having 1's in the first column..

For intra model movement, a Gibbs sampling scheme can be used to update the parameters iteratively given K and P. In the next section, we describe the Data Driven Reversible Jump (DDRJ) to propose efficient ROIs and SNPs to be included (birth) into the current model or excluded (death) from it.

CHAPTER 4

DATA DRIVEN REVERSIBLE JUMP

Despite its generalization, RJ also has some drawbacks. Its performance relies on the probability of visiting the next model and the proposal distribution to obtain the next set of parameters. Indeed, bad proposals will usually lead to high rejection rate, low mixing and consequently more iterations would be needed for convergence. In summary, the consequences are inefficient moves and delay in convergence, even lack of convergence for a fixed number of iterations. One reason to understand these points may be that of trying to move from a parameter having high density in a bad model to a parameter of low density in a good model. There is a high probability of this move being rejected and if the proposals are bad, this kind of move may often happen and not be accepted.

Data Driven Reversible Jump (ZUANETTI; MILAN, 2016; ZUANETTI; MILAN, 2017) is an attempt to solve this problem and will be matter of discussion in this chapter. Much of the literature and recent works on RJ have been devoted to strategy of designing good proposals. Brooks, Giudici and Roberts (2003), Ehlers and Brooks (2008), Papathomas, Dellaportas and Vasdekis (2009), Gagnon (2019), Zanella (2020) have suggested some methods for designing proposals in a variety of context. Our approach here follows that of Zuanetti and Milan (2016) named Data Driven Reversible Jump (DDRJ) in sequel where the methodology was applied in genetics context for Quantitative Locus Trait (QTL) mapping. The main idea is to design a proposal driven by the data that leads to lower rejection rate, better mixing and higher effective sample size.

Just for analogy to understand DDRJ, think about the standard greedy method when creating classification trees (BREIMAN *et al.*, 1984). First of all, one has to choose a loss or negative loss function, usually Information Gain or Gini Impurity, to decide on the best split, i.e. the combination of a node and a splitting point. For each combination of a candidate node and splitting point, Information Gain or Gini Impurity is computed and the combination that maximizes or minimizes our function is selected as the best split. DDRJ is quite similar with this approach, in the sense that it introduces a metric for deciding on the next model but holding the random essence from RJ.

In particular, in traditional RJ for variable selection, the next model to be visited is chosen uniformly from the set of candidate models, which is not the best we can do if our search space is large as in many combinatorial problems. Therefore, instead of using a uniform distribution, we propose a strategy to assign higher weights to more promising candidate models .

Our strategy relies on, at each iteration, trying to include or exclude a single covariate from the current. Thus, consider the current model with some covariates and First, we decide if we will include a new covariate (birth) or exclude (death) one that is present in the model. Obviously, in the case where we don't have any covariate, we would opt for a birth move and at the other extreme when the model is saturated with all the covariates we would opt for a death move. Then, we define a metric roughly understood as a criterion to choose the next candidate and not a rigorous definition of metric in mathematics[1]. After obtaining the candidate model, we sample the new set of parameters using any mechanism and we then test its acceptance. From a practical perspective, this step should be fast and hopefully parallelizable and may not even use the full dataset if possible to avoid more time consumption in our MCMC scheme. We also argue that the DDRJ could be coupled with recent advances in scalable Bayes (ANGELINO; JOHNSON; ADAMS, 2016), specially in the regime of big p and n.

[1] a metric in math should have some properties as identity, symmetry, triangle inequality.

4.1 Model 1: Selecting ROIs

In this section, we define the DDRJ for ROIs selection where the model under consideration is

$$Y_i^* = \beta_0 + \sum_{p \in \mathscr{G}} \beta_p X_{ip} + \varepsilon_i, \quad \varepsilon_i \sim N(0,1), \quad Y_i = \mathbb{I}(Y_i^* > 0), \quad \mathscr{G} = \{ROIs\, in\, the\, model\}. \tag{4.1}$$

Assume that the current model in Equation (4.1) contains $Card(\mathscr{G})^2 = P \in \{0, 1, \ldots, g\}$ ROIs. Thus, the next move is decided as follow

1. if $P = 0$, no ROI is in the model, then a birth (b) movement is proposed with probability $p(b|P = 0) = 1$;
2. if $0 < P < g$, some ROIs are in the model, then a birth or death movement is proposed with probability $p(b|P) = p(d|P) = \frac{1}{2}$;
3. if $P = g$, all the ROIs are in the model, then a death (d) movement is proposed with probability $p(d|P = g) = 1$.

- **Birth**

  Let us suppose that the current model has P ROIs, denote by X_P its design matrix and that a birth move has been chosen. We propose to choose the next candidate from the remaining ROIs X_{-P} with probability given $p_{bj} = \frac{|cor(\xi_P, X_j)|}{\sum_{X_{-P}} |cor(\xi_P, X_j)|}$, where $cor(\xi_P, X_j)$ is the correlation between a candidate ROI X_j and the residuals ξ_P from the current model. Any other metric that could gather how relevant each candidate is in respect to the current residual can be used to easier the Markov chain's direction.

  After selecting this ROI, we sample β^b from the full conditional $\pi(\beta | X_{P+1}, Y^*)$, X_{P+1} being the design matrix with the new candidate, and test its acceptance with probability $\psi^b = min(1, A^b)$, where

$$A^b = \frac{L(\beta^b | X_{P+1}, Y^*)\pi(\beta^b) q(\beta | \beta^b)}{L(\beta | X_P, Y^*)\pi(\beta) q(\beta^b | \beta)} \tag{4.2}$$

$$q(\beta^b | \beta) = p(b|P) p_{bj} \pi(\beta^b | Y^*, X_{P+1}) \text{ and}$$

$$q(\beta | \beta^b) = p(d|P+1) p_{dj} \pi(\beta | Y^*, X_P).$$

[2] Card : cardinality

- **Death**

  A possible way of choosing the candidate ROI to be deleted is by comparing the size of their coefficients after standardizing and scaling the design matrix. Then, we propose to select a ROI to be excluded with probability $p_{dj} = \frac{\frac{1}{|\beta_j|}}{\sum_{j=1}^{P} \frac{1}{|\beta_j|}}$. The greater the coefficient of a given ROI, the lesser is its probability to be deleted. Again, any other metric to choose the ROI to be deleted is valid.

  After selecting this ROI, we sample β^d from the full conditional $\pi(\beta|X_{P-1}, Y^*)$, X_{P-1} being the design matrix without the candidate covariate to be deleted, and test its acceptance with probability $\psi^d = min(1, A^d)$, where

$$A^d = \frac{L(\beta^d|X_{P-1}, Y^*)\pi(\beta^d)q(\beta|\beta^d)}{L(\beta|X_P, Y^*)\pi(\beta)q(\beta^d|\beta)} \tag{4.3}$$

$$q(\beta^d|\beta) = p(d|P)p_{dj}\pi(\beta^d|Y^*, X_{P-1}), \text{ and}$$

$$q(\beta|\beta^d) = p(b|P-1)p_{bj}\pi(\beta|Y^*, X_P).$$

4.2 Model 2: Selecting SNPs

In this section, we define the DDRJ for SNPs selection where the model under consideration is

$$Y_i^* = \beta_0 + \sum_{k\in\mathscr{M}} \alpha_k Z_{ik} + \sum_{k\in\mathscr{M}} \delta_k(1-|Z_{ik}|) + \varepsilon_i, \quad \varepsilon_i \sim N(0,1), \quad Y_i = \mathbb{I}(Y_i^* > 0), \tag{4.4}$$

$$\mathscr{M} = \{SNPs\ in\ the\ model\}.$$

Assume that at a stage of the process, the model in Equation (4.4) contains $Card(\mathscr{M}) = K \in \{0, 1, \ldots, m\}$ SNPs. Thus, the next step is decided as follow:

1. if $K = 0$, no SNP in the model, then a birth (b) movement is proposed with probability $p(b|K=0) = 1$;

2. if $0 < K < m$, some SNPs are in the model, then a birth or death movement is proposed with probability $p(b|K) = p(d|K) = \frac{1}{2}$;

3. if $K = m$, all the SNPs are in the model, then a death (d) movement is proposed with probability $p(d|K=m) = 1$.

- **Birth**

  Assume that the current model contains K SNPs. The choice of the next SNP is guided by its association with the residuals ξ_K from model in Equation (4.4). Each SNP Z_k is a factor with 3 levels, thus the association with the current residuals can be measured using the Kruskal-Wallis (KW) statistics and Z_k is selected with probability $p_{bk} = \frac{\mathrm{KW}(\xi_k, Z_k)}{\sum_{Z_{-K}} \mathrm{KW}(\xi_k, Z_k)}$, where $\mathrm{KW}(\xi_K, Z_k)$ is the KW-statistics summarizing the association between the residuals ξ_K and the candidate Z_k, and Z_{-K} is the set of remaining candidate SNPs to be included. It's worth mentioning that we are not testing hypothesis but only using the test's statistic as a metric to quantify levels of association. Again, any other measure of association could have been used, F-statistic for instance just to cite one.

  After selecting this SNP, we sample $(\beta_0, \alpha, \delta)^b$ from $\pi(\beta_0, \alpha, \delta | Z_{K+1}, Y^*)$, Z_{K+1} being the design matrix with the new candidate SNP. This sampling will occur in one step using a joint matrix representation with a Gibbs sampling scheme and test its acceptance with probability $\psi^b = min(1, A^b)$.

$$A^b = \frac{L((\beta_0, \alpha, \delta)^b | Z_{K+1}, Y^*)\pi((\beta_0, \alpha, \delta)^b) q((\beta_0, \alpha, \delta)|(\beta_0, \alpha, \delta)^b)}{L((\beta_0, \alpha, \delta))|Z_K, Y^*)\pi(\beta_0, \alpha, \delta) q((\beta_0, \alpha, \delta)^b|(\beta_0, \alpha, \delta))}, \quad (4.5)$$

$$q((\beta_0, \alpha, \delta)^b|(\beta_0, \alpha, \delta)) = p(b|K) p_{bk} \pi((\beta_0, \alpha, \delta)^b | Y^*, Z_{K+1}) \text{ and}$$

$$q((\beta_0, \alpha, \delta)|(\beta_0, \alpha, \delta)^b) = p(d|K+1) p_{dk} \pi((\beta_0, \alpha, \delta)| Y^*, Z_K).$$

- **Death**

  As Z_k only takes value in $\{-1, 0, 1\}$, the absolute value of the coefficients α_k and δ_k in Equation (4.4) gives a measure of its importance. We propose to select a SNP to be excluded with probability $p_{dk} = \frac{\frac{1}{|\alpha_k| + |\delta_k|}}{\sum_{k=1}^{K} \frac{1}{|\alpha_k| + |\delta_k|}}$. The higher the effect of the SNP, the lesser is its probability to be deleted.

  After selecting this SNP, we sample $(\beta_0, \alpha, \delta)^d$ from $\pi(\beta_0, \alpha, \delta | Z_{K-1}, Y^*)$, Z_{K-1} being the design matrix without the candidate SNP to be deleted. This sampling will occur in one step and we test its acceptance with probability $\psi^d = min(1, A^d)$, where

$$A^d = \frac{L((\beta_0, \alpha, \delta)^d | Z_{K-1}, Y^*)\pi((\beta_0, \alpha, \delta)^d) q((\beta_0, \alpha, \delta)|(\beta_0, \alpha, \delta)^d)}{L((\beta_0, \alpha, \delta))|Z_K, Y^*)\pi(\beta_0, \alpha, \delta) q((\beta_0, \alpha, \delta)^d|(\beta_0, \alpha, \delta))}, \quad (4.6)$$

$$q((\beta_0, \alpha, \delta)^d|(\beta_0, \alpha, \delta)) = p(d|K) p_{dk} \pi((\beta_0, \alpha, \delta)^d | Y^*, Z_{K-1}) \text{ and}$$

$$q((\beta_0, \alpha, \delta)|(\beta_0, \alpha, \delta)^d) = p(b|K-1) p_{bk} \pi((\beta_0, \alpha, \delta)| Y^*, Z_K).$$

4.3 Model 3: Selecting ROIs and SNPs

Model 3 aims at selecting ROIs and SNPs in an integrative manner, i.e. jointly. Let's remember the model under study defined by

$$Y_i^* = \beta_0 + \sum_{p \in \mathscr{G}} \beta_p X_{ip} + \sum_{k \in \mathscr{M}} \alpha_k Z_{ik} + \sum_{k \in \mathscr{M}} \delta_k (1 - |Z_{ik}|) + \varepsilon_i, \quad \varepsilon_i \sim N(0,1), \tag{4.7}$$

$$Y_i = \mathbb{I}(Y_i^* > 0), \quad Z_{ik} \in \{-1, 0, 1\},$$

$$\mathscr{G} = \{ROIs\ in\ the\ model\}, \quad \mathscr{M} = \{SNPs\ in\ the\ model\}.$$

We could think of, mainly, three alternatives to perform this joint variable selection, all worth testing to draw conclusions about which is better. The final choice relies on whether one sees ROIs or SNPs to be the most relevant part of the available information. The alternatives are

- **Alternative 1:** Select all possible ROIs and then select SNPs. In other words, run Model 1 and then run Model 2 conditional on selected ROIs;
- **Alternative 2:** Select all possible SNPs and then select ROIs i.e. In other words, run Model 2 and then run Model 1 conditional on selected SNPs;
- **Alternative 3:** Randomly alternate between Model 1 and Model 2.

Alternatives 1 and 2 are just a combination of what have been discussed in the previous sections. The last option is more challenging, mainly for one reason. In Model 1 and 2, all the features were of the same type, either numerical or categorical, and thinking of a data driven metric to include or exclude candidate was simple. However, in the joint framework, this is not so obvious and the goal here is to propose a way to tackle this problem. First of all, when trying to include or exclude ROIs we will keep using correlation and coefficients size as it was the case in Model 1, and we will also keep using KW statistic and coefficients size to include or delete SNPs as in Model 2.

At each stage of the process, we will randomly decide to work ROIs and SNPs in the following manner:

1. Decide with probability $s = \frac{g}{m+g}$ and $1 - s = \frac{m}{m+g}$ to work on ROIs or SNPs, respectively. This step allows us to jump into ROIs or SNPs space and then work on them separately. This is fair enough if $m \approx g$ as $s \approx 0.5$. However, if one dimension dominates the other, it may be better to select variables separately using model 1 and 2 or simply design an adaptive probability to favor any desired space.
2. If ROIs space has been selected, then we apply Model 1 strategy conditional on already selected SNPs at this moment.

3. If SNPs space has been selected, then we apply Model 2 strategy conditional on already selected ROIs at this moment.

The next two subsections can be skipped if one has already understood all the process, though we decided to include all necessary details for better understanding.

4.3.1 What if we jump into ROIs space?

Suppose that the current model contains P ROIs and K SNPs, with parameters $\theta = (\beta, \alpha, \delta)$ and we decide to jump to ROIs space. Then:

1. if $P = 0$ then a birth (b) movement is proposed with probability $p(b|P = 0) = 1$;
2. if $0 < P < g$ then a birth or death movement is proposed with probability $p(b|P) = p(d|P) = \frac{1}{2}$;
3. if $P = g$ then a death (d) movement is proposed with probability $p(d|P = g) = 1$.

- **Birth**

  Let's suppose that a birth move has been chosen. We propose to choose the next candidate from the remaining ROIs X_{-P} with probability $p_{bj} = \frac{|cor(\xi_P,X_j)|}{\sum_{X_{-P}}|cor(\xi_P,X_j)|}$,where $cor(\xi_p, X_j)$ is the correlation between a candidate ROI X_j and the residuals ξ_P from the current model.

  After selecting a ROI, we sample θ^b from the full conditional $\pi(\beta, \alpha, \delta|X_{P+1}, Z_K, Y^*)$ and test its acceptance with probability $\psi^b = min(1, A^b)$, where

$$A^b = \frac{L(\theta^b|X_{P+1}, Z_K, Y^*)\pi(\theta^b)q(\theta|\theta^b)}{L(\theta|X_P, Z_K, Y^*)\pi(\theta)q(\theta^b|\theta)}, \tag{4.8}$$

$$q(\theta^b|\theta) = p(b|P)p_{bj}\pi(\theta^b|X_{P+1}, Z_K, Y^*) \text{ and}$$

$$q(\theta|\theta^b) = p(d|P+1)p_{dj}\pi(\theta|X_P, Z_K, Y^*).$$

- **Death**

  A possible way of choosing the candidate ROI to be deleted is by comparing the size of their coefficients after normalizing the design matrix. Then, we propose to select a ROI to be excluded with probability $p_{dj} = \frac{\frac{1}{|\beta_j|}}{\sum_{j=1}^{P}\frac{1}{|\beta_j|}}$. The greater the coefficient of a given ROI, the lesser is its probability to be deleted.

After selecting a ROI, we sample θ^d from the full conditional $\pi(\beta, \alpha, \delta | X_{P-1}, Z_K, Y^*)$ and test its acceptance with probability $\psi^d = min(1, A^d)$, where

$$A^d = \frac{L(\theta^d | X_{P-1}, Z_K, Y^*)\pi(\theta^d)q(\theta|\theta^d)}{L(\theta|X_P, Z_K, Y^*)\pi(\theta)q(\theta^d|\theta)}, \tag{4.9}$$

$$q(\theta^d|\theta) = p(d|P)p_{dj}\pi(\theta^d|Y^*, X_{P-1}, Z_K) \text{ and}$$

$$q(\theta|\theta^d) = p(b|P-1)p_{bj}\pi(\theta|Y^*, X_P, Z_K).$$

4.3.2 What if we jump into SNPs space?

Suppose that the current model contains P ROIs and K SNPs, with parameters $\theta = (\beta, \alpha, \delta)$ and we decide to jump into SNPs space. Then:

1. if $K = 0$ then a birth (b) movement is proposed with probability $p(b|K = 0) = 1$;
2. if $0 < K < m$ then a birth or death movement is proposed with probability $p(b|K) = p(d|K) = \frac{1}{2}$;
3. if $K = m$ then a death (d) movement is proposed with probability $p(d|K = m) = 1$.

- **Birth**

  The choice of the next SNP is guided by its association with the residuals ξ_K from model in Equation (4.7). Z_k is a factor with 3 levels, so the association with the current residuals can be measured using the Kruskal-Wallis(KW) statistics and Z_k is selected with probability $p_{bk} = \frac{\text{KW}(\xi_k, Z_k)}{\sum_{Z_{-K}} \text{KW}(\xi_k, Z_k)}$, where $\text{KW}(\xi_k, Z_k)$ is the KW-statistics summarizing the association between the residuals ξ_K and the candidate Z_k and Z_{-K} is the set of remaining candidate SNPs to be included. It's worth mentioning that we're not testing hypothesis but only using the test's statistic as a metric to quantify levels of association.

  After selecting a SNP, we sample θ^b from the full conditional $\pi(\beta, \alpha, \delta | X_P, Z_{K+1}, Y^*)$ and test its acceptance with probability $\psi^b = min(1, A^b)$, where

$$A^b = \frac{L(\theta^b | X_P, Z_{K+1}, Y^*)\pi(\theta^b)q(\theta|\theta^b)}{L(\theta|X_P, Z_K, Y^*)\pi(\theta)q(\theta^b|\theta)}, \tag{4.10}$$

$$q(\theta^b|\theta) = p(b|K)p_{bk}\pi(\theta^b|X_P, Z_{K+1}, Y^*) \text{ and}$$

$$q(\theta|\theta^b) = p(d|K+1)p_{dk}\pi(\theta|X_P, Z_K, Y^*).$$

- **Death**

  As Z_k only takes value in $\{-1,0,1\}$, the absolute value of the coefficients α_k and δ_k in Equation (4.7) gives a measure of its importance. We propose to select a SNP to be excluded with probability $pd_k = \frac{\frac{1}{|\alpha_k|+|\delta_k|}}{\sum_{k=1}^{K}\frac{1}{|\alpha_k|+|\delta_k|}}$. The higher the effect of the SNP, the lesser is its probability to be deleted.

  After selecting a SNP, we sample θ^d from the full conditional $\pi(\beta,\alpha,\delta|X_P,Z_{K-1},Y^*)$ and test its acceptance with probability $\psi^d = min(1,A^d)$.

$$A^d = \frac{L(\theta^d|X_P,Z_{K-1},Y^*)\pi(\theta^d)q(\theta|\theta^d)}{L(\theta|X_P,Z_K,Y^*)\pi(\theta)q(\theta^d|\theta)}, \tag{4.11}$$

$$q(\theta^d|\theta) = p(d|K)p_{dk}\pi(\theta^d|Y^*,X_P,Z_{K-1}) \quad \text{and}$$

$$q(\theta|\theta^d) = p(b|K-1)p_{bk}\pi(\theta|Y^*,X_P,Z_K).$$

4.4 Algorithms

4.4.1 Model 1: Selecting ROIs

This section describes the algorithm for performing ROIs selection.

Algorithm for selecting ROIs

1. Setup $P = 0$ to start without any ROI and sample Y^* from the truncated normal distribution in Equation (3.7).
2. For the lth iteration $l = 1, \dots, L$
   a) Choose a birth or death movement
   b) If a birth has been chosen then
      i. Select the ROI to be included using p_{bj}
      ii. Sample the candidate value β^b from the full conditional in Equation (3.4)
      iii. Accept the proposal with probability ψ^b given the new candidate parameters
      iv. If the candidate is accepted, update the model size $P^{(l)} = P^{(l-1)} + 1$, the new parameters are β^b and sample Y_b^* from $\pi(Y^*|\beta^b, X_{P+1}, Y)$ given by the full conditional in Equation (3.7)
      v. If the candidate is not accepted, do $P^{(l)} = P^{(l-1)}$ and the new set of parameters β and Y^* are just updated given the current X_P using a Gibbs Sampling.
   c) If a death has been chosen then
      i. Select the ROI to be excluded using p_{dj}
      ii. Sample the candidate value β^d from the full conditional Equation (3.4)
      iii. Accept the proposal with probability ψ^d given the candidate parameters
      iv. If the candidate is accepted, update the model size $P^{(l)} = P^{(l-1)} - 1$, the new parameters are β^d and sample Y_d^* from $\pi(Y^*|\beta^d, X_{P-1}, Y)$ given by the full conditional in Equation (3.7)
      v. If the candidate is not accepted, do $P^{(l)} = P^{(l-1)}$ and the new set of parameters β and Y^* are just updated given the current X_P and Y using a Gibbs Sampling.

4.4.2 Model 2: Selecting SNPs

This section describes the algorithm for performing SNPs selection.

Algorithm for selecting SNPs

```
1. Setup $K = 0$ to start without SNPs and sample $Y^*$ from the truncated normal distribution in Equation (3.7)
2. For the $l$th iteration $l = 1, \dots, L$
    a) Choose a birth or death movement
    b) If a birth has been chosen then
        i. Select the SNP to be included using $p_{bk}$
        ii. Sample the candidate value $(\beta_0,\ \alpha,\ \delta)^b$ from the full conditional $\pi(\beta_0, \alpha, \delta | Z_{K+1}, Y^*)$
        iii. Accept the proposal with probability $\psi^b$ in Equation (4.5) given the candidate parameters
        iv. If the candidate is accepted, update the model size $K^{(l)} = K^{(l-1)} + 1$, the new parameters are $(\beta_0, \alpha, \delta)^b$ and sample $Y_b^*$ from $\pi(Y^* | (\beta_0, \alpha, \delta)^b, Z_{K+1}, Y)$ given by the full conditional in Equation (3.7)
        v. If the candidate is not accepted, do $K^{(l)} = K^{(l-1)}$ and the new set of parameters $(\beta_0, \alpha, \delta)$, $Y^*$ are just updated given the current $Z_K$ using a Gibbs Sampling.
    c) If a death has been chosen then
        i. Select the SNP to be excluded using $p_{dk}$
        ii. Sample the candidate value $(\beta_0,\ \alpha,\ \delta)^d$ from the full conditional $\pi(\beta_0, \alpha, \delta | Z_{K-1}, Y^*)$
        iii. Accept the proposal with probability $\psi^d$ in Equation (4.6) given the candidate parameters
        iv. If the candidate is accepted, update the model size $K^{(l)} = K^{(l-1)} - 1$, the new parameters are $(\beta_0, \alpha, \delta)^b$ and sample $Y_d^*$ from $\pi(Y^* | (\beta_0, \alpha, \delta)^b, Z_{K-1}, Y)$ given by the full conditional in Equation (3.7)
        v. If the candidate is not accepted, do $K^{(l)} = K^{(l-1)}$ and the new set of parameters $(\beta_0, \alpha, \delta)$, $Y^*$ are just updated given the current $Z_K$ and $Y$ using a Gibbs Sampling.
```

4.4.3 Model 3: Selecting ROIs and SNPs

This section describes the algorithm for performing ROIs and SNPs selection.

Algorithm for jointly selecting SNPs and ROIs

```
1. Setup $P = K = 0$ to start without ROIs or SNPs and sample $Y^*$ from the truncated normal distribution in Equation (3.7)
2. For the $l$th iteration $l = 1, \dots, L$
   a) Choose a jump into ROIs or SNPs space
   b) If a jump into ROIs has been accepted
      i. Choose a birth or death movement
      ii. If a birth has been chosen then:
         A. Select the ROI to be included using $p_{bj}$
         B. Sample the candidate value $\theta^b$ from the full conditionals using Equations (3.4), (3.5), (3.6)
         C. Accept the proposal with probability $\psi^b$ in Equation (4.8) given the candidate parameters
         D. If the candidate is accepted, update the model size $P^{(l)} = P^{(l-1)} + 1$, $K^{(l)} = K^{(l-1)}$, the new parameters are $\theta^b$ and sample $Y_b^*$ from $\pi(Y^*|\theta^b, X_{P+1}, Z_K, Y)$ given by the full conditional in Equation (3.7)
         E. If the candidate is not accepted, do $P^{(l)} = P^{(l-1)}$, $K^{(l)} = K^{(l-1)}$, and the new parameters $\theta$ and $Y^*$ are updated given $Z_{K^{(l-1)}}$, $X_{P^{(l-1)}}$ and $Y$.
      iii. If a death has been chosen then
         A. Select the ROI to be excluded using $p_{dj}$
         B. Sample the candidate value $\theta^d$ from the full conditionals using Equations (3.4), (3.5), (3.6)
         C. Accept the proposal with probability $\psi^d$ in Equation (4.9) given the candidate parameters
         D. If the candidate is accepted, update the model size $P^{(l)} = P^{(l-1)} - 1$, $K(l) = K^{(l-1)}$, the new parameters are $\theta^d$ and sample $Y_d^*$ from $\pi(Y^*|\theta^d, X_{P-1}, Z_K, Y)$ given by the full conditional in Equation (3.7)
         E. If the candidate is not accepted, do $P^{(l)} = P^{(l-1)}$, $K^{(l)} = K^{(l-1)}$, and the new parameters $\theta$ and $Y^*$ are updated given $Z_{K^{(l-1)}}$, $X_{P^{(l-1)}}$ and $Y$.
   c) If a jump into SNPs space has been accepted
      i. Choose a birth or death movement
      ii. If a birth has been chosen then
```

A. Select the SNP to be included using p_{bk} from Equation (4.3.2)

B. Sample the candidate value θ^b from the full conditionals from Equations (3.4), (3.5), (3.6) with minor change

C. Accept the proposal with probability ψ^b in Equation (4.10) given the candidate parameters

D. If the candidate is accepted, update the model size $P^{(l)} = P^{(l-1)}$, $K^{(l)} = K^{(l-1)} + 1$, the new parameters are θ^b and sample Y_b^* from $\pi(Y^*|\theta^b, X_P, Z_{K+1}, Y)$ given by the full conditional in Equation (3.7)

E. If the candidate is not accepted, do $P^{(l)} = P^{(l-1)}$, $K^{(l)} = K^{(l-1)}$, and the new parameters θ and Y^* are updated given $Z_{K^{(l-1)}}$, $X_{P^{(l-1)}}$ and Y.

iii. If a death has been chosen then:

A. select the SNP to be excluded using p_{dk}

B. sample the candidate value θ^b from the full conditional Equation (3.4), (3.5),(3.6) with minor change in one step using a joint structure

C. Accept the proposal with probability ψ^d in Equation (4.11) given the candidate parameters

D. Update the model size $P^{(l)} = P^{(l-1)}$, $K^{(l)} = K^{(l-1)} - 1$, the new parameters are θ^d and sample Y_d^* from $\pi(Y^*|\theta^d, X_{P-1}, Z_K, Y)$ given by the full conditional in Equation (3.7)

E. If the candidate is not accepted, do $P^{(l)} = P^{(l-1)}$, $K^{(l)} = K^{(l-1)}$, and the new parameters θ and Y^* are updated given $Z_{K^{(l-1)}}$, $X_{P^{(l-1)}}$ and Y.

4.5 Variable selection and Prediction

As stated at the beginning of the work, our goal is two-fold: variable selection and prediction that can be achieved using two different strategies. On one hand, this goal can be attained in two separate steps with variable selection done using the full dataset and prediction done with a cross-validation approach to access the model's quality. On the other hand, the full process can be done in only one step using a cross validation approach for both variable selection and prediction. In any case, we can decide to select only those variables with posterior probability of inclusion, estimated as the relative frequency of the coefficient being non null, above a threshold (0.5 for instance).

Here, we opt for the first strategy where the full dataset is used for variable selection , which allows us to have a greater sample size that benefits our method, and we keep all features with marginal posterior probability of inclusion greater than 0.5. For prediction, we use a 5-folds cross-validation with the variable selection algorithm run on each fold. As it has been said through out this work, the manuscripts from Chekouo *et al.* (2016) and Stingo *et al.* (2013) will be our benchmark. Thus, following the same approach there, the 5-folds cross-validation with 94 healthy controls and 74 patients for the training set, and 24 healthy controls and 18 patients for the validation set will be used for predictive performance analysis.

To predict the disease status for a new subject having ROIs and SNPs X_{new}, Z_{new}, the latent variable $Y^*_{new}|\cdots = \sum_t \left(\hat{\beta_0}^t + \sum_{p\in\mathscr{G}} \hat{\beta_p}^t X^{new}_{ip} + \sum_{k\in\mathscr{M}} \hat{\alpha_k}^t Z^{new}_{ik} + \sum_{k\in\mathscr{M}} \hat{\delta_k}^t (1 - |Z^{new}_{ik}|)\right) P(M_t|y)$ is first computed with parameters set to the posterior mean and the posterior predictive probability of disease is computed as $P(Y_{new} = 1|\ldots) = \Phi(Y^*_{new}|\ldots)$, $\Phi(.)$ being the standard normal cumulative distribution function. Note how a Bayesian Model Averaging is employed to compute the latent variable $Y^*_{new}|\ldots$ using all the t visited models M_t. From these posterior probabilities and latent variables, we can compute the Area Under the receiver operating characteristics Curve (AUC) and Misclassification Error rate (MCE).

CHAPTER 5

SIMULATION STUDY

This section summarizes a simulation study to demonstrate the efficiency of our method for performing variable selection using Data Driven Reversible Jump and for making prediction for future subjects. For each scenario, 35.000 MCMC iterations were run with a burn-in period of 5.000 iterations holding one sample of ten. To access convergence, monitored through log posterior, we run two chains with randomly chosen initial points. Each of the following section contains two types of studies: one in which we test our method on a simulated dataset that mimics the real dataset with the same number of ROIs and SNPs and in the second study we increase the number of ROIs and SNPs. We also use the posterior probability of each model to compare our method to the traditional Reversible jump with uniform proposals between models, expecting that our methodology will assign a higher posterior probability to the true model, due to faster convergence. Finally, we compare our model and inference methodology to the LASSO and Random forest in terms of missclassification error and area under the ROC curve-AUC using a 5-fold cross-validation. All the results were run using the R software (RStudio Team, 2020) on a *Intel(R) Core(TM) i7-8565U CPU 1.80GHz* with the KW statistics being coded in C++ to accelerate the proposal computation.

5.1 Model 1: Selecting ROIs

To mimic the real ROI dataset, we simulate 116 covariates from a multivariate normal distribution with empirical mean and covariance matrix retrieved from the real design matrix for 210 individuals. The second group of datasets is simulated from a standard multivariate normal distribution with fixed sample size $n = 300$ and increased number of ROIs. From these covariates, we select some ROIs with non-null effects and their coefficients were assigned to maintain the healthy and diagnosed with schizophrenia proportion (43.8%). The disease status was generated from the probit model in Equation (4.1) with coefficients summarized in Table 2 and the prior variance is set to $\sigma^2_\beta = 100$. We decide to select a ROI if its marginal posterior probability of inclusion (mppi) is greater than 0.5. Our algorithm performed well in all the scenarios, selecting all the relevant variables as well as providing good estimates and small standard errors summarized in Table 3. Furthermore, in all the scenarios our methodology always select and assign a higher posterior probability to the true model compared to the RJ with uniform proposals as it is shown in Table 4, with those differences probably due to the faster convergence of DDRJ and better mixing of DDRJ chain. Finally, in Table 5, MCE and AUC computed from the Bayesian model averaging used for prediction show that our methodology generally outperforms the Random Forest and is comparable to the LASSO, another well established method for variable selection.

Table 2 – Coefficients in the simulation scenario for ROIs.

	Relevant effect	Non relevant effects
$n = 210,\ g = 116$	$\beta_0 = 1,\ \beta_1 = -2,\ \beta_3 = -2.5,\ \beta_{115} = 3$	$Normal(0, 0.1)$
$n = 300,\ g = 300$	$\beta_0 = 1,\ \beta_1 = -1,\ \beta_3 = -1.5,\ \beta_{299} = 2$	$Normal(0, 0.1)$
$n = 300,\ g = 500$	$\beta_0 = 1,\ \beta_1 = -1,\ \beta_3 = 0.8,\ \beta_4 = -1.5,\ \beta_{499} = 2$	$Normal(0, 0.1)$
$n = 300,\ g = 1000$	$\beta_0 = 1,\ \beta_1 = -2,\ \beta_3 = -2.5,\ \beta_{115} = 3$	$Normal(0, 0.1)$

Table 4 – Comparing the DDRJ and RJ using the three most visited models with their posterior probability for ROIs selection, where the real model column shows the true ROIs in the simulated model.

	Real model	DDRJ	RJ
$n = 210,\ g = 116$	1 3 115	**1 3 115 (0.342)** 1 3 70 115 (0.045) 1 3 52 115 (0.039)	1 3 115 (0.304) 1 3 70 115 (0.112) 1 3 52 115 (0.042)
$n = 300,\ g = 300$	1 3 299	**1 3 299 (0.341)** 1 3 34 299 (0.026) 1 3 32 299 (0.020)	1 3 299 (0.329) 1 3 269 299 (0.029) 1 3 32 299 (0.023)
$n = 300,\ g = 500$	1 2 3 499	**1 2 3 499 (0.0635)** 1 2 3 486 499 (0.0607) 1 2 3 177 486 499 (0.045)	1 2 3 486 499 (0.039) 1 2 3 129 302 393 486 499 (0.014) 1 2 3 176 486 499 (0.011)
$n = 300,\ g = 1000$	1 2 3 4 1000	**1 2 3 4 1000 (0.083)** 1 2 3 4 752 1000 (0.013) 1 2 3 4 353 1000 (0.009)	1 2 3 4 1000 (0.076) 1 2 3 4 752 1000 (0.041) 1 3 4 752 1000 (0.034)

Table 3 – Marginal posterior probability of inclusion, estimates and standard errors for selected ROIs on simulated datasets.

		mppi	Estimate	Real
$n = 210,\ g = 116$	β_0	1.000	0.794 (0.184)	1.000
	ROI 1	0.999	-2.020 (0.376)	-2.000
	ROI 3	0.999	-2.640 (0.526)	-2.500
	ROI 115	0.999	3.068 (0.496)	3.000
$n = 300,\ g = 300$	β_0	1.000	0.833 (0.156)	1.000
	ROI 1	0.999	-0.992 (0.164)	-1.000
	ROI 3	0.999	-1.770 (0.234)	-1.500
	ROI 299	0.999	1.968 (0.272)	2.000
$n = 300,\ g = 500$	β_0	1.000	1.202 (0.212)	1.000
	ROI 1	0.999	-1.306 (0.248)	-1.000
	ROI 3	0.999	0.887 (0.184)	0.800
	ROI 4	0.999	-1.535 (0.233)	-1.500
	ROI 486	0.627	-0.340 (0.291)	0.007
	ROI 499	0.999	2.145 (0.331)	2.000
$n = 300,\ g = 1000$	β_0	1.000	0.957 (0.278)	1.000
	ROI 1	0.999	1.272 (0.330)	1.200
	ROI 2	0.999	0.903 (0.266)	0.800
	ROI 3	0.999	-1.728 (0.461)	-1.500
	ROI 4	0.999	-1.206 (0.351)	-1.000
	ROI 1000	0.999	2.840 (0.692)	2.300

Table 5 – Comparing the predictive performance in terms of Misclassification error (MCE) and Area under the ROC curve (AUC) on simulated ROIs dataset.

		DDRJ	LASSO	RF
$n = 210,\ g = 116$	MCE	0.114 (0.061)	0.137 (0.073)	0.228 (0.064)
	AUC	0.956 (0.034)	0.944 (0.057)	0.838 (0.054)
$n = 300,\ g = 300$	MCE	0.126 (0.055)	0.129 (0.047)	0.289 (0.035)
	AUC	0.959 (0.021)	0.944 (0.028)	0.874 (0.018)
$n = 300,\ g = 500$	MCE	0.109 (0.025)	0.149 (0.031)	0.349 (0.016)
	AUC	0.962 (0.020)	0.935 (0.029)	0.800 (0.061)
$n = 300,\ g = 1000$	MCE	0.133 (0.042)	0.109 (0.022)	0.369 (0.021)
	AUC	0.942 (0.022)	0.951 (0.015)	0.820 (0.061)

5.2 Model 2: Selecting SNPs

Regarding the genetic dataset, we simulate 81 features from independent discrete distributions with empirical probabilities retrieved from the real SNP dataset, while the second group of dataset is simulated from independent discrete distribution with fixed sample size $n = 300$ and increased number of SNPs. Then, we select some SNPs with non null effects and coefficients assigned to maintain the healthy and diagnosed with schizophrenia proportion. The disease status was generated from the probit model in Equation (4.4) with coefficients summarized in Table 6 and the prior variance is set to $\sigma^2_{\beta_0} = \sigma^2_{\alpha} = \sigma^2_{\delta} = 100$. Again, we decide to select a SNP if its mppi is greater than 0.5. Our algorithm performed well in all the scenarios, selecting all the relevant variables as well as providing good estimates and small standard errors summarized in Table 7. Moreover, in almost all the scenarios our methodology always select and assign a higher posterior probability to the true model compared to the RJ with uniform proposals as it is shown in Table 8, as a result of faster convergence of the DDRJ. Finally, in Table 9, MCE and AUC computed from the Bayesian model averaging used for prediction show that our methodology generally outperforms the Random Forest and is comparable and even better, in some cases, than the LASSO, another well established method for variable selection.

Table 6 – Coefficients in the simulation scenario for SNPs.

	Relevant effect	Non Relevant effects
$n = 210,\ m = 81$	$\beta_0 = 1.7$ $(\alpha_1 = 1.3, \alpha_2 = 1, \alpha_3 = -1.5, \alpha_4 = -1.2)$ $(\delta_1 = -1, \delta_2 = -1.4, \delta_3 = -1.4, \delta_4 = -2)$	$Normal(0, 0.1)$
$n = 300,\ m = 300$	$\beta_0 = 2$ $(\alpha_1 = 1.3, \alpha_2 = 1.2, \alpha_3 = -1, \alpha_4 = -1.5)$ $(\delta_1 = -1, \delta_2 = -1.4, \delta_3 = -1.5, \delta_4 = -2)$	$Normal(0, 0.1)$
$n = 300,\ m = 500$	$\beta_0 = 1.3$ $(\alpha_1 = 1.3, \alpha_2 = 1.2, \alpha_3 = -1, \alpha_4 = -0.5)$ $(\delta_1 = -1, \delta_2 = -1.4, \delta_3 = -1.5, \delta_4 = -2)$	$Normal(0, 0.1)$
$n = 300,\ m = 1000$	$\beta_0 = 1.3$ $(\alpha_1 = 1.3, \alpha_2 = 1.2, \alpha_3 = -1, \alpha_4 = -0.5)$ $(\delta_1 = -1, \delta_2 = -1.4, \delta_3 = -1.5, \delta_4 = -2)$	$Normal(0, 0.1)$

Table 7 – Marginal posterior probability of inclusion, estimates and standard errors for selected SNPs on simulated datasets.

		mppi	$\hat{\alpha}$	α	$\hat{\delta}$	δ
$n = 210,\ m = 81$	β_0	1.000	1.640 (0.293)	1.700	-	-
	SNP 1	0.999	1.479 (0.235)	1.300	-0.538 (0.353)	-1.000
	SNP 2	0.999	1.025 (0.199)	1.000	-1.596 (0.409)	-1.400
	SNP 3	0.998	-1.545 (0.248)	-1.500	-1.627 (0.431)	-1.400
	SNP 4	0.999	-0.954 (0.180)	-1.200	-1.682 (0.437)	-2.000
$n = 300,\ m = 300$	β_0	1.000	2.129 (0.273)	2.000	-	-
	SNP 1	0.999	1.324 (0.189)	1.300	-1.560 (0.399)	-1.000
	SNP 2	0.999	1.320 (0.185)	1.200	-1.107 (0.294)	-1.400
	SNP 3	0.998	-0.956 (0.174)	-1.000	-1.633 (0.382)	-1.500
	SNP 4	0.999	-1.662 (0.212)	-1.500	-1.919 (0.340)	-2.000
$n = 300,\ m = 500$	β_0	1.000	1.260 (0.187)	1.300	-	-
	SNP 1	0.999	1.135 (0.150)	1.300	-0.721 (0.274)	-1.000
	SNP 2	0.998	0.933 (0.139)	1.200	-1.772 (0.316)	-1.400
	SNP 3	0.994	-0.912 (0.143)	-1.000	-1.087 (0.293)	-1.500
	SNP 4	0.996	-0.414 (0.119)	-0.500	-1.631 (0.301)	-2.000
$n = 300,\ m = 1000$	β_0	1.000	1.213 (0.179)	1.300	-	-
	SNP 1	0.998	1.291 (0.166)	1.300	-1.336 (0.286)	-1.000
	SNP 2	0.998	1.001 (0.147)	1.200	-1.393 (0.341)	-1.400
	SNP 3	0.998	-0.743 (0.142)	-1.000	-1.390 (0.308)	-1.500
	SNP 4	0.999	-0.475 (0.122)	-0.500	-2.092 (0.396)	-2.000

Table 8 – Comparing the DDRJ and RJ using the three most visited models with their posterior probability for SNPs selection, where the real model column shows the true SNPs in the simulated model.

	Real model	DDRJ	RJ
$n = 210,\ m = 81$	1 2 3 4	1 2 3 4 (0.932) 1 2 3 4 75 (0.058) 1 2 3 4 30 (0.002)	**1 2 3 4 (0.969)** 1 2 3 4 75 (0.012) 1 2 3 4 58 (0.005)
$n = 300,\ m = 300$	1 2 3 4	**1 2 3 4 (0.988)** 1 2 3 4 167 (0.004) 1 2 3 4 217 (0.002)	1 2 3 4 (0.984) 1 2 3 4 258 (0.008) 1 2 3 4 17 (0.003)
$n = 300,\ m = 500$	1 2 3 4	**1 2 3 4 (0.989)** 1 2 3 4 261(0.002) 1 2 3 4 274 (0.001)	1 2 3 4 (0.987) 1 2 3 4 492 (0.001) 1 2 3 4 417 (0.001)
$n = 300,\ m = 1000$	1 2 3 4	**1 2 3 4 (0.962)** 1 2 3 4 833 (0.006) 1 2 3 4 990 (0.006)	1 2 3 4 (0.807) 1 3 (0.081) 1 2 3 (0.074)

Table 9 – Comparing the predictive performance in terms of Misclassification error (MCE) and Area under the ROC curve (AUC) on simulated SNPs dataset.

		DDRJ	LASSO	RF
$n = 210,\ m = 81$	MCE	0.104 (0.031)	0.175 (0.024)	0.251 (0.041)
	AUC	0.942 (0.020)	0.924 (0.015)	0.851 (0.030)
$n = 300,\ m = 300$	MCE	0.143 (0.049)	0.190 (0.062)	0.346 (0.021)
	AUC	0.934 (0.033)	0.911 (0.033)	0.872 (0.053)
$n = 300,\ m = 500$	MCE	0.166 (0.065)	0.195 (0.055)	0.396 (0.015)
	AUC	0.907 (0.004)	0.866 (0.039)	0.730 (0.007)
$n = 300,\ m = 1000$	MCE	0.176 (0.060)	0.229 (0.032)	0.400 (0.011)
	AUC	0.905 (0.035)	0.864 (0.031)	0.719 (0.028)

5.3 Model 3: Selecting ROIs and SNPs

For the joint selection of ROIs and SNPs, the first dataset is a simulation of 116 ROIs from a multivariate normal distribution with empirical mean and covariance matrix retrieved from the design matrix and we simulate 81 SNPs from independent discrete distributions with probabilities retrieved from the real SNP dataset for 210 individuals and the second group of dataset contains a simulation from a standard multivariate normal and independent discrete distribution with increased number of ROIs and SNPs respectively. Non null effects, summarized in Table 10, for ROIs and SNPs were chosen to keep the proportion of healthy and diagnosed with schizophrenia. The disease status was generated using the probit model in Equation (4.7) with prior variance set to $\sigma^2_\beta = \sigma^2_\alpha = \sigma^2_\delta = 25$. As the number of candidate models under consideration grows for joint selection, we decided for a two steps procedure. In the first step, we apply Algorithm 1 and 2 separately for a pre-filtering to reduce the numbers of ROIs and SNPs, and we use a small threshold (0.1) for selection. From our studies, this strategy reduces the number of covariates to approximately $10-15\%$, in average. The selected variables are then used together in the second step under Algorithm 3 for joint selection and prediction. Our algorithm performed well in all the scenarios, selecting all the relevant variables as well as providing good estimates and small standard errors summarized in Table 11. Moreover, in almost all the scenarios our methodology always select and assign a higher posterior probability to the true model compared to the RJ with uniform proposals as it is shown in Table 12, as a result of faster convergence of the DDRJ. Finally, in Table 13, MCE and AUC computed from the Bayesian model averaging used for prediction show that our methodology generally outperforms the Random Forest and is comparable and even better, in some cases, than the LASSO, another well established method for variable selection.

Table 10 – Coefficients in simulation scenario for SNPs and ROIs.

	Relevant effect	Non relevant effects
$n = 210,\ g = 116,\ m = 81$	$\beta_0 = 1$ $(\beta_1 = 1.3, \beta_3 = 1.5, \beta_{115} = 1)$ $(\alpha_1 = 1.3, \alpha_2 = -1, \alpha_3 = 1.5, \alpha_4 = 1)$ $(\delta_1 = -1.2, \delta_2 = -1, \delta_3 = -1.3, \delta_4 = -2)$	$Normal(0, 0.1)$
$n = 300,\ g = m = 300$	$\beta_0 = 1$ $(\beta_1 = 1.3, \beta_3 = 1.5, \beta_{299} = 1)$ $(\alpha_1 = 1.3, \alpha_2 = -1, \alpha_3 = 1.5, \alpha_4 = 1)$ $(\delta_1 = -1.2, \delta_2 = -1, \delta_3 = -1.3, \delta_4 = -2)$	$Normal(0, 0.1)$
$n = 300,\ g = m = 500$	$\beta_0 = 1$ $(\beta_1 = 1.3, \beta_3 = 1.5, \beta_{499} = 1)$ $(\alpha_1 = 1.3, \alpha_2 = -1, \alpha_3 = 1.5, \alpha_4 = 1)$ $(\delta_1 = -1.2, \delta_2 = -1, \delta_3 = -1.3, \delta_4 = -2)$	$Normal(0, 0.1)$
$n = 300,\ g = m = 1000$	$\beta_0 = 1$ $(\beta_1 = 1.3, \beta_3 = 1.5, \beta_{999} = 1)$ $(\alpha_1 = 1.3, \alpha_2 = -1, \alpha_3 = 1.5, \alpha_4 = 1)$ $(\delta_1 = -1.2, \delta_2 = -1, \delta_3 = -1.3, \delta_4 = -2)$	$Normal(0, 0.1)$

Table 11 – Marginal posterior probability of inclusion, estimates and standard errors for selected ROIs and SNPs on simulated datasets.

		mppi	$\hat{\beta}$	β	$\hat{\alpha}$	α	$\hat{\delta}$	δ
	β_0		1.289 (0.280)	1.000	-	-	-	-
	ROI 1	1.000	1.215 (0.238)	1.300	-	-	-	-
	ROI 3	0.999	1.669 (0.284)	1.500	-	-	-	-
	ROI 115	0.999	1.490 (0.292)	1.000	-	-	-	-
$n = 210$	*SNP* 1	0.999	-	-	1.401(0.301)	1.300	-2.614 (0.566)	-1.200
$g = 116$	*SNP* 2	0.999	-	-	-1.105 (0.243)	-1.000	-1.186 (0.513)	-1.000
$m = 81$	*SNP* 3	0.999	-	-	1.871 (0.336)	1.500	-0.840 (0.420)	-1.300
	SNP 4	0.999	-	-	1.184 (0.230)	1.000	-2.439 (0.720)	-2.000
	β_0	1.000	1.436 (0.377)	1.000	-	-	-	-
	ROI 1	0.998	1.303 (0.285)	1.300	-	-	-	-
	ROI 3	0.998	1.886 (0.406)	1.500	-	-	-	-
	ROI 299	0.998	1.323 (0.313)	1.000	-	-	-	-
$n = 300$	*SNP* 1	0.999	-	-	1.492 (0.351)	1.300	-1.605 (0.567)	-1.200
$g = 300$	*SNP* 2	0.878	-	-	-0.964 (0.421)	-1.000	-1.362 (0.660)	-1.000
$m = 300$	*SNP* 3	0.999	-	-	1.820 (0.388)	1.500	-1.579 (0.485)	-1.300
	SNP 4	0.999	-	-	1.441 (0.323)	1.000	-3.063 (0.670)	-2.000
	β_0	1.000	1.368 (0.292)	1.000	-	-	-	-
	ROI 1	0.999	1.716 (0.276)	1.300	-	-	-	-
	ROI 3	0.999	1.999 (0.318)	1.500	-	-	-	-
	ROI 499	0.998	1.131 (0.217)	1.000	-	-	-	-
$n = 300$	*SNP* 1	0.999	-	-	1.343 (0.247)	1.300	-2.480 (0.542)	-1.200
$g = 500$	*SNP* 2	0.999	-	-	-1.101 (0.232)	-1.000	-1.240 (0.390)	-1.000
$m = 500$	*SNP* 3	0.999	-	-	2.035 (0.323)	1.500	-1.761 (0.458)	-1.300
	SNP 4	0.999	-	-	1.379 (0.258)	1.000	-2.834 (0.556)	-2.000
	β_0	1.000	1.319 (0.243)	1.000	-	-	-	-
	ROI 1	0.998	1.361 (0.214)	1.300	-	-	-	-
	ROI 3	0.998	1.663 (0.253)	1.500	-	-	-	-
	ROI 999	0.998	1.001 (0.170)	1.000	-	-	-	-
$n = 300$	*SNP* 1	0.999	-	-	1.438 (0.233)	1.300	-1.426 (0.376)	-1.200
$g = 1000$	*SNP* 2	0.878	-	-	-1.243 (0.208)	-1.000	-1.413 (0.386)	-1.000
$m = 1000$	*SNP* 3	0.999	-	-	1.685 (0.230)	1.500	-2.193 (0.470)	-1.300
	SNP 4	0.999	-	-	1.047 (0.196)	1.000	-2.292 (0.408)	-2.000

Table 13 – Comparing the predictive performance in terms of Misclassification error (MCE) and Area under the ROC curve (AUC) on simulated ROIs-SNPs dataset.

		DDRJ	LASSO	RF
$n = 210, m = 81$	MCE	0.193 (0.061)	0.208 (0.026)	0.347 (0.054
	AUC	0.880 (0.053)	0.890 (0.023)	0.758 (0.037)
$n = 300, m = 300$	MCE	0.113 (0.026)	0.149 (0.042)	0.302 (0.070)
	AUC	0.960 (0.017)	0.944 (0.025)	0.791 (0.074)
$n = 300, m = 500$	MCE	0.156 (0.069)	0.182 (0.047)	0.409 (0.0.04)
	AUC	0.926 (0.04)	0.899 (0.033)	0.673 (0.044)
$n = 300, m = 1000$	MCE	0.183 (0.035)	0.136 (0.059)	0.349 (0.028)
	AUC	0.902 (0.029)	0.945(0.036)	0.743 (0.021)

Table 12 – Comparing the DDRJ and RJ using the three most visited models with their posterior probability for ROIs and SNPs selection, where the real model column shows the true ROIs and SNPs in the simulated model.

	Real model	DDRJ	RJ
$n = 210$		*ROI* (1,3,115)-*SNP* (1,2,3,4) **(0.920)**	*ROI* (1,2,115)-*SNP* (1,2,3,4) (0.901)
$m = 81$	*SNP* (1,2,3,4)	(1,3,107,115)-(1,2,3,4) (0.018)	(1,3,7,115)-(1,2,3,4) (0.025)
$g = 116$	*ROI* (1,3,115)	(1,3,7,115)-(1,2,3,4) (0.013)	(1,3,49,115)-(1,2,3,4) (0.019)
$n = 300$		*ROI* (1,3,299)-*SNP* (1,2,3,4) **(0.903)**	*ROI* (1,3,299)-SNP(1,2,3,4) (0.873)
$m = 300$	*SNP* (1,2,3,4)	(1,3,75,115)-(1,2,3,4) (0.086)	(1,3,75,115)-(1,2,3,4) (0.102)
$g = 300$	*ROI* (1,3,299)	(1,3,16,115)-(1,2,3,4) (0.006)	(1,3,16,115)-(1,2,3,4) (0.003)
$n = 300$		*ROI* (1,3,499)-*SNP*(1,2,3,4) **(0.834)**	*ROI* (1,3,499)-*SNP* (1,2,3,4) (0.807)
$m = 500$	*SNP* (1,2,3,4)	(1,3,499)-(1,2,3,4,63) (0.113)	(1,3,499)-(1,2,3,4,63) (0.049)
$g = 500$	*ROI* (1,3,499)	(1,3,499)-(1,3,4,63) (0.011)	(1,3,499)-(1,3,4,63) (0.003)
$n = 300$		*ROI* (1,3,999)-*SNP* (1,2,3,4) (0.770)	*ROI* (1,3,999)-*SNP* (1,2,3,4) **(0.773)**
$m = 1000$	*SNP* (1,2,3,4)	(1,3,528,999)-(1,2,3,4) (0.220)	(1,3,528,999)-(1,2,3,4) (0.221)
$g = 1000$	*ROI* (1,3,999)	(1,3,999)-(1,3,4) (0.005)	(1,3,999)-(1,3,4) (0.001)

CHAPTER 6

MCIC SCHIZOPHRENIA DATASET

The available dataset was collected by the Mind Clinical Imaging Consortium (MCIC) (CHEN *et al.*, 2012) as an effort of deeper understanding of mental disorder and contains both imaging data on activation patterns using fMRI during a sensorimotor task and multiple SNPs allele frequencies which have previously been implicated in schizophrenia on 118 healthy controls and 92 subjects affected by this disorder, with no history of substance abuse and free of any medical, neurological or psychiatric illnesses.

The goal of the MCIC study, a joint effort of four research teams from Boston, Iowa, Minnesota and New Mexico, was to identify regions of interest (ROI) in the brain with discriminating activation patterns between cases and controls and relate them to a relevant set of SNPs able to explain these variations, a model selection problem clearly. fMRI was mainly designed to identify brain's response to stimulus by detecting regional neuronal activity captured by blood oxygenation level-dependent (BOLD) variations. The original fMRI data were collected during sensorimotor task as response to auditory stimulation.

The data were then preprocessed in SPM5 <http://www.fil.ion.ucl.ac.uk/spm>, realigned to correct for the subjects movements, spatially normalized to correct for anatomic variability, spatially smoothed to improve signal to noise ratio. For each of the 116 ROIs, the activation level was summarized as the median of the statistical parametric map values (FRISTON *et al.*, 1994) for that region. The genetic information of the available dataset is given by 81 SNPs, already known to be related to schizophrenia retrieved from the Schizophrenia Research Forum <http://www.schizophreniaforum.org/> information. In the original dataset, the SNP information was coded as the number of minor allele for those with genotype aa, aA and AA respectively. More details of the experimental study and preprocessing can be found in Chen *et al.* (2012) and Stingo *et al.* (2013).

Our baseline comparison for model selection and prediction will be the results obtained in Stingo *et al.* (2013) and Chekouo *et al.* (2016). Thus, for model selection, the full dataset will be used while for prediction, we will keep the same 5-fold cross-validation configuration and use AUC (Area under ROC curve), MCE (Misclassification error) as predictive performance metrics.

6.1 Model selection

For each scenario, 35.000 MCMC iterations were run with a burn-in period of 5.000 iterations holding each sample of 10. The prior variance is set to $\sigma_\alpha^2 = \sigma_\beta^2 = \sigma_\delta^2 = 25$ to ensure that the prior is not too informative but also not too much large for our method to be effective.

6.1.1 Model 1: Selecting ROIs

For model selection, the full dataset have been used and the selected variables are ROIs **61** and **115** with mppi 0.837 and 0.932, but also suggesting more investigation on ROI **35** with mppi 0.416 and estimated effects $\hat{\beta} = (0.183, -0.181, -0.514, -0.607)$ summarised in Table 14. ROIs **35** (left posterior cingulate region) and **61** (left inferior parietal region) were also selected by Stingo *et al.* (2013) and Chekouo *et al.* (2016) and are known to be related to schizophrenia. In special, ROI **115** (posterior inferior vermis- lobule IX) was a new findings that could narrow future research on lobules I to X. Chekouo *et al.* (2016) found one more ROI **57** that has not been selected here but was present in the top 3 models as shown in Table 15. A more careful scientist may rely on this rule, include every covariate that appears in the top t models, here 3, to select the ROIs and thus will select ROIs $\mathbf{35}, \mathbf{57}, \mathbf{61}, \mathbf{96}, \mathbf{115}$. As a result of faster convergence, we were expecting the DDRJ to assign a higher probability to the most visited model compared to the traditional RJ which didn't happen in this case as shown in Table 15 and is probably due to the multicollinearity between candidate ROIs and the correlation metric for selecting the next model to be visited. If, for instance, there is a high collinearity between two ROIs, the probability of selecting any of them is almost the same, according to our correlation metric, and our algorithm may not be batter than traditional RJ in these cases. Regarding prediction, in Table 16 we show that our model, though simple, and methodology perform well in terms of predictive performance compared to the results from Chekouo *et al.* (2016), LASSO and Random Forest, with a misclassification error and area under the ROC curve of **0.399 (0.05)** and **0.619 (0.06)** respectively.

Table 14 – Marginal posterior probability of inclusion, estimates and standard errors for selected ROIs on the real dataset.

	mppi	$\hat{\beta}$
β_0	1.000	0.183 (0.095)
ROI 35	0.416	-0.181 (0.239)
ROI 61	0.837	-0.514 (0.286)
ROI 115	0.932	-0.607 (0.233)

Table 15 – Comparing the DDRJ and RJ using the three most visited models with their posterior probability for ROIs selection on the real dataset.

DDRJ	RJ
ROI 61 115 (0.061)	*ROI* 61 115 **(0.076)**
35 61 96 115 (0.055)	35 61 115 (0.051)
35 57 61 (0.021)	35 57 96 (0.019)

Table 16 – Comparing the predictive performance on the real ROI dataset in terms of Misclassification error (MCE) and Area under ROC curve (AUC).

	Benchmark	DDRJ	LASSO	RF
MCE	0.37 (0.02)	0.399 (0.05)	0.379 (0.06)	0.351 (0.05)
AUC	0.66 (0.02)	0.619 (0.06)	0.653 (0.06)	0.684 (0.06)

6.1.2 Model 2: Selecting SNPs

In Table 17, we summarize the selected SNPs through the marginal posterior probability of inclusion (mppi) using the full dataset. The selected variable would be SNP $\mathbf{22}, \mathbf{61}$ with mppi **0.96** and **0.72** respectively. Although having a mppi **0.34** lesser than 0.5, we would also suggest SNP **32**. SNP **22** (rs3737597) is located in gene DISC1 (chromosome 1), a gene known to be strongly associated to schizophrenia and was also found by Stingo *et al.* (2013) and Chekouo *et al.* (2016) who also found SNPs **10** and **38** to be discriminatory. Table 18 shows that DDRJ is also consistent with traditional RJ in the selected variable but also converges faster. Finally, in Table 19 we show that our model, though simple, and methodology perform well in terms of predictive performance compared to the results from Chekouo *et al.* (2016), LASSO and Random Forest, with a misclassification error and area under the ROC curve of **0.475 (0.03)** and **0.570 (0.02)** respectively.

Table 17 – Marginal posterior probability of inclusion, estimates and standard errors for selected SNPs on the real dataset.

	mppi	$\hat{\alpha}$	$\hat{\delta}$
β_0	1.000	2.511 (0.349)	-
SNP 22	0.957	-1.513 (0.248)	3.842 (0.664)
SNP 32	0.345	0.874 (0.482)	0.817 (0.462)
SNP 61	0.719	-2.159 (0.926)	-1.960 (0.844)

Table 18 – Comparing the DDRJ and RJ using the three most visited models with their posterior probability for SNPs selection on the real dataset.

DDRJ	RJ
SNP 22 61 **(0.096)**	*SNP* 22 61 (0.083)
22 32 61 (0.046)	22 (0.074)
22 (0.043)	22 32 61 (0.020)

Table 19 – Comparing the predictive performance on the real SNP dataset in terms of Misclassification error (MCE) and Area under ROC curve (AUC).

	Benchmark	DDRJ	LASSO	RF
MCE	0.45 (0.01)	0.475 (0.031)	0.446 (0.037)	0.437 (0.031)
AUC	0.64 (0.02)	0.570 (0.02)	0.558 (0.04)	0.557 (0.05)

6.1.3 Model 3: Selecting SNPs and ROIs

In this section, a joint selection of ROIs and SNPs is performed using Algorithm 3. Again ROIs **35, 61, 115** and SNP **22** are suggested as discriminatory variables with mppi **0.291, 0.794, 0.968, 0.955** respectively. In Table 20, we summarise the mppi, estimates and standard errors for each coefficients and Table 21 shows that our result is also consistent with traditional RJ, though assigning a lower posterior probability. Moreover, in Table 22 we show that our model and algorithm perform well in terms of predictive performance compared to the results from Chekouo *et al.* (2016), LASSO and Random Forest, with a misclassification error and area under the ROC curve of **0.427 (0.017)** and **0.672 (0.05)** respectively.

Table 20 – Marginal posterior probability of inclusion, estimates and standard errors for selected ROIs and SNPs on the real dataset.

	mppi	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\delta}$
β_0	1.000	2.945 (0.447)	-	-
ROI 35	0.291	-0.119 (0.203)	-	-
ROI 61	0.794	-0.479 (0.296)	-	-
ROI 115	0.968	0.619 (0.196)	-	-
SNP 22	0.955	-	-1.602 (0.635)	2.607 (0.592)

Table 21 – Comparing the DDRJ and RJ using the three most visited models with their posterior probability for ROIs and SNPs selection on the real dataset.

DDRJ	RJ
ROI (61,115)-SNP (22) 0.067	ROI (61,115)-SNP (22) **0.082**
(35,61,115)-(22) 0.044	(61,62,115)-(22) 0.055
(61,62,115)-(22) 0.017	(35,61,115)-(22) 0.026

Table 22 – Comparing the predictive performance on the real ROI-SNP dataset in terms of Misclassification error (MCE) and Area under ROC curve (AUC).

	Benchmark	DDRJ	LASSO	RF
MCE	0.33 (0.02)	0.427 (0.017)	0.406 (0.037)	0.404 (0.01)
AUC	0.69 (0.03)	0.672 (0.05)	0.617 (0.04)	0.633 (0.06)

CHAPTER 7

CONCLUSIONS

In this work, we have developed a Data Driven Reversible Jump for variable selection using a Bayesian probit model. Our goals, selecting ROIs and SNPs, and assess predictive risk for schizophrenia based on functional Magnetic Resonance Imaging (fMRI) and Single Nucleotide Polymorphism (SNPs) information have been reached. Most ROIs 35, 57, 61, 115 and SNP 22 that we selected were in accordance with results from other authors and also known to be related to the disease, even though some new findings ROI 96 and SNPs 32, 61 have been suggested and may be subject of deeper research. Compared to other methodologies as traditional LASSO and Random Forest, in terms of predictive accuracy, the DDRJ also perfoms well when predictions are done using the Bayesian Model Averaging. We have also noticed that assuming the same prior variance for ROIs and SNPs is quite restrictive. Thus, future work may use different prior variance or assign a hyperprior to the variance.

From a methodological perspective, we noticed that the metric used inside the DDRJ can improve or degrade the efficiency of the algorithm. For instance, in high dimensional setting where features are usually correlated, using correlation between residuals and candidate variable may not be efficient, because the probabilities of selecting any candidate could be similar and then recast our algorithm into the traditional RJ with uniform jumps. Moreover, a metric as correlation only captures linear relation, thus a more suitable idea may be a metric, like a kernel, that captures some non-linearity.

Regarding extensions, another direction of studies would be testing other priors such as those shrinkage priors introduced earlier to improve our current methodology. As we have also mentioned, a distance matrix between ROIs is available and has not been used in this work. This information could be included either as part of the DDRJ to make better jumps, or assume a Markov Random Field type of prior for ROIs and apply the DDRJ to perform variable selection and prediction for future subjects. Other extension of this work that is worth investigating is to perform clustering while selecting discriminating ROIs and SNPs, and again the DDRJ could be used to select the number of cluster and estimate parameters.

BIBLIOGRAPHY

AKAIKE, H. Information theory and an extension of the maximum likelihood principle. **Proceedings of the 2nd International Symposium on Information Theory**, Budapest: Akademiai Kiado, p. 267–281, 1973. Citations on pages 22 and 27.

ALBERT, J. H.; CHIB, S. Bayesian analysis of binary and polychotomous response data. **Journal of the American Statistical Association**, Taylor & Francis Group, v. 88, n. 422, p. 669–679, 1993. Citation on page 35.

ANGELINO, E.; JOHNSON, M. J.; ADAMS, R. P. Patterns of scalable Bayesian inference. **arXiv preprint arXiv:1602.05221**, 2016. Citation on page 40.

BREIMAN, L.; FRIEDMAN, J.; STONE, C. J.; OLSHEN, R. A. **Classification and Regression Trees**. [S.l.]: CRC press, 1984. Citations on pages 23 and 39.

BROOKS, S. P.; GIUDICI, P.; ROBERTS, G. O. Efficient construction of Reversible Jump Markov Chain Monte Carlo proposal distributions. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Wiley Online Library, v. 65, n. 1, p. 3–39, 2003. Citations on pages 23 and 39.

CARVALHO, C. M.; POLSON, N. G.; SCOTT, J. G. Handling sparsity via the horseshoe. In: **Artificial Intelligence and Statistics**. [S.l.: s.n.], 2009. p. 73–80. Citation on page 31.

CHEKOUO, T.; STINGO, F. C.; GUINDANI, M.; DO, K.-A. A Bayesian predictive model for imaging genetics with application to schizophrenia. **The Annals of Applied Statistics**, The Institute of Mathematical Statistics, v. 10, n. 3, p. 1547–1571, 09 2016. Available: <https://doi.org/10.1214/16-AOAS948>. Citations on pages 22, 23, 31, 52, 63, 64, 65, and 66.

CHEN, J.; CALHOUN, V. D.; PEARLSON, G. D.; EHRLICH, S.; TURNER, J. A.; HO, B.-C.; WASSINK, T. H.; MICHAEL, A. M.; LIU, J. Multifaceted genomic risk for brain function in schizophrenia. **Neuroimage**, Elsevier, v. 61, n. 4, p. 866–875, 2012. Citations on pages 22 and 63.

EHLERS, R. S.; BROOKS, S. P. Adaptive proposal construction for Reversible jump MCMC. **Scandinavian Journal of Statistics**, Wiley Online Library, v. 35, n. 4, p. 677–690, 2008. Citation on page 39.

ERP, S. V.; OBERSKI, D. L.; MULDER, J. Shrinkage priors for Bayesian penalized regression. **Journal of Mathematical Psychology**, Elsevier, v. 89, p. 31–50, 2019. Citations on pages 22 and 30.

FAN, Y.; SISSON, S. A. Reversible jump MCMC . **Handbook of Markov Chain Monte Carlo**, Chapman and Hall/CRC, p. 67–92, 2011. Citation on page 32.

FRAGOSO, T. M.; BERTOLI, W.; LOUZADA, F. Bayesian model averaging: A systematic review and conceptual classification. **International Statistical Review**, Wiley Online Library, v. 86, n. 1, p. 1–28, 2018. Citation on page 34.

FRISTON, K. J.; HOLMES, A. P.; WORSLEY, K. J.; POLINE, J.-P.; FRITH, C. D.; FRACKOWIAK, R. S. Statistical parametric maps in functional imaging: a general linear approach. **Human Brain Mapping**, Wiley Online Library, v. 2, n. 4, p. 189–210, 1994. Citation on page 63.

GAGNON, P. Informed Reversible Jump algorithms. **arXiv preprint arXiv:1911.02089**, 2019. Citations on pages 23 and 39.

GELMAN, A.; HWANG, J.; VEHTARI, A. Understanding predictive information criteria for Bayesian models. **Statistics and Computing**, Springer, v. 24, n. 6, p. 997–1016, 2014. Citations on pages 22, 26, and 29.

GEORGE, E. I.; MCCULLOCH, R. E. Approaches for Bayesian variable selection. **Statistica Sinica**, JSTOR, p. 339–373, 1997. Citations on pages 22 and 31.

GLOVER, G. H. Overview of functional magnetic resonance imaging. **Neurosurgery Clinics**, Elsevier, v. 22, n. 2, p. 133–139, 2011. Citation on page 21.

GREEN, P. J. Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. **Biometrika**, Oxford University Press, v. 82, n. 4, p. 711–732, 1995. Citations on pages 22 and 32.

GUR, R. E.; GUR, R. C. Functional magnetic resonance imaging in schizophrenia. **Dialogues in Clinical Neuroscience**, Les Laboratoires Servier, v. 12, n. 3, p. 333, 2010. Citation on page 21.

HASTINGS, W. K. Monte Carlo sampling methods using Markov Chains and their applications. Oxford University Press, 1970. Citation on page 32.

HOERL, A. E.; KENNARD, R. W. Ridge regression: Biased estimation for nonorthogonal problems. **Technometrics**, Taylor & Francis Group, v. 12, n. 1, p. 55–67, 1970. Citation on page 30.

HOETING, J. A.; MADIGAN, D.; RAFTERY, A. E.; VOLINSKY, C. T. Bayesian model averaging: a tutorial. **Statistical Science**, JSTOR, p. 382–401, 1999. Citations on pages 23 and 34.

ISHWARAN, H.; RAO, J. S. *et al.* Spike and slab variable selection: frequentist and Bayesian strategies. **The Annals of Statistics**, Institute of Mathematical Statistics, v. 33, n. 2, p. 730–773, 2005. Citations on pages 22 and 30.

JACOB, A. Limitations of clinical psychiatric diagnostic measurements. **Journal of Neurological Disorders**, Citeseer, 2013. Citation on page 21.

JAIN, S.; NEAL, R. M. A split-merge Markov Chain Monte Carlo procedure for the dirichlet process mixture model. **Journal of Computational and Graphical Statistics**, Taylor & Francis, v. 13, n. 1, p. 158–182, 2004. Citation on page 23.

JOHNSON, V. E.; ROSSELL, D. Bayesian model selection in high-dimensional settings. **Journal of the American Statistical Association**, Taylor & Francis Group, v. 107, n. 498, p. 649–660, 2012. Citation on page 31.

KASS, R. E.; RAFTERY, A. E. Bayes factors. **Journal of the American Statistical Association**, Taylor & Francis, v. 90, n. 430, p. 773–795, 1995. Citations on pages 22 and 26.

KULLBACK, S.; LEIBLER, R. A. On information and sufficiency. **Annals of Mathematical Statistics**, The Institute of Mathematical Statistics, v. 22, n. 1, p. 79–86, 03 1951. Available: <https://doi.org/10.1214/aoms/1177729694>. Citation on page 27.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 2015. Citation on page 25.

LINDQUIST, M. A. *et al.* The statistical analysis of fmri data. **Statistical Science**, Institute of Mathematical Statistics, v. 23, n. 4, p. 439–464, 2008. Citation on page 22.

LOPES, H. F. A note on Reversible Jump Markov Chain Monte Carlo. 2006. Citation on page 32.

LOVINGER, D. M. Communication networks in the brain: neurons, receptors, neurotransmitters, and alcohol. **Alcohol Research & Health**, Superintendent of Documents, 2008. Citation on page 35.

LUNN, D. J.; THOMAS, A.; BEST, N.; SPIEGELHALTER, D. Winbugs-a Bayesian modelling framework: concepts, structure, and extensibility. **Statistics and Computing**, Springer, v. 10, n. 4, p. 325–337, 2000. Citation on page 28.

MAH, J. T.; CHIA, K. A gentle introduction to SNP analysis: resources and tools. **Journal of Bioinformatics and Computational Biology**, World Scientific, v. 5, n. 05, p. 1123–1138, 2007. Citation on page 22.

METROPOLIS, N.; ROSENBLUTH, A. W.; ROSENBLUTH, M. N.; TELLER, A. H.; TELLER, E. Equation of state calculations by fast computing machines. **The Journal of Chemical Physics**, American Institute of Physics, v. 21, n. 6, p. 1087–1092, 1953. Citation on page 32.

MITCHELL, T. J.; BEAUCHAMP, J. J. Bayesian variable selection in linear regression. **Journal of the American Statistical Association**, Taylor & Francis Group, v. 83, n. 404, p. 1023–1032, 1988. Citations on pages 22 and 31.

NATHOO, F. S.; KONG, L.; ZHU, H.; INITIATIVE, A. D. N. A review of statistical methods in imaging genetics. **Canadian Journal of Statistics**, Wiley Online Library, v. 47, n. 1, p. 108–131, 2019. Citation on page 21.

NEAL, R. M. MCMC using Hamiltonian dynamics. In: ____. **Handbook of Markov Chain Monte Carlo**. [S.l.]: CRC Press, 2011. v. 2. Citation on page 23.

O'HARA, R. B.; SILLANPÄÄ, M. J. A review of bayesian variable selection methods: what, how and which. **Bayesian Analysis**, International Society for Bayesian Analysis, v. 4, n. 1, p. 85–117, 2009. Citations on pages 22 and 30.

PAPATHOMAS, M.; DELLAPORTAS, P.; VASDEKIS, V. G. A general proposal construction for Reversible jump MCMC. **Technical Paper, Department of Statistics, Athens University of Economics and Business**, Citeseer, 2009. Citation on page 39.

PATEL, K. R.; CHERIAN, J.; GOHIL, K.; ATKINSON, D. Schizophrenia: overview and treatment options. **Pharmacy and Therapeutics**, MediMedia, USA, v. 39, n. 9, p. 638, 2014. Citation on page 21.

PLUMMER, M. *et al.* Jags: A program for analysis of bayesian graphical models using gibbs sampling. In: VIENNA, AUSTRIA. **Proceedings of the 3rd international workshop on distributed statistical computing**. [S.l.], 2003. v. 124, n. 125.10, p. 1–10. Citation on page 28.

PLUTA, D.; YU, Z.; SHEN, T.; CHEN, C.; XUE, G.; OMBAO, H. Statistical methods and challenges in connectome genetics. **Statistics & Probability Letters**, Elsevier, v. 136, p. 83–86, 2018. Citation on page 21.

POLSON, N. G.; SCOTT, J. G.; WINDLE, J. Bayesian inference for logistic models using pólya–gamma latent variables. **Journal of the American statistical Association**, Taylor & Francis, v. 108, n. 504, p. 1339–1349, 2013. Citation on page 36.

RStudio Team. **RStudio: Integrated Development Environment for R**. Boston, MA, 2020. Available: <http://www.rstudio.com/>. Citation on page 53.

RUHRMANN, S.; SCHULTZE-LUTTER, F.; KLOSTERKÖTTER, J. Early detection and intervention in the initial prodromal phase of schizophrenia. **Pharmacopsychiatry**, © Georg Thieme Verlag Stuttgart· New York, v. 36, n. S 3, p. 162–167, 2003. Citation on page 21.

SARAIVA, E. F.; MILAN, L. A. Clustering gene expression data using a posterior split-merge-birth procedure. **Scandinavian Journal of Statistics**, Wiley Online Library, v. 39, n. 3, p. 399–415, 2012. Citation on page 23.

SCHWARZ, G. *et al.* Estimating the dimension of a model. **The Annals of Statistics**, Institute of Mathematical Statistics, v. 6, n. 2, p. 461–464, 1978. Citations on pages 22 and 28.

SCOTT, J. G.; BERGER, J. O. Bayes and empirical-Bayes multiplicity adjustment in the variable-selection problem. **The Annals of Statistics**, JSTOR, p. 2587–2619, 2010. Citation on page 32.

SIMPSON, D.; RUE, H.; RIEBLER, A.; MARTINS, T. G.; SØRBYE, S. H. Penalising model component complexity: A principled, practical approach to constructing priors. **Statistical Science**, Institute of Mathematical Statistics, v. 32, n. 1, p. 1–28, 2017. Citation on page 31.

SPIEGELHALTER, D. J.; BEST, N. G.; CARLIN, B. P.; LINDE, A. V. D. Bayesian measures of model complexity and fit. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Wiley Online Library, v. 64, n. 4, p. 583–639, 2002. Citations on pages 22 and 28.

STINGO, F. C.; GUINDANI, M.; VANNUCCI, M.; CALHOUN, V. D. An integrative Bayesian modeling approach to imaging genetics. **Journal of the American Statistical Association**, Taylor & Francis, v. 108, n. 503, p. 876–891, 2013. Citations on pages 22, 52, 63, 64, and 65.

TIBSHIRANI, R. Regression shrinkage and selection via the Lasso. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 58, n. 1, p. 267–288, 1996. Citations on pages 23 and 30.

WATANABE, S. A widely applicable Bayesian information criterion. **Journal of Machine Learning Research**, v. 14, n. Mar, p. 867–897, 2013. Citations on pages 22 and 29.

WATANABE, S.; OPPER, M. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. **Journal of Machine Learning Research**, v. 11, n. 12, 2010. Citation on page 29.

ZANELLA, G. Informed proposals for local MCMC in discrete spaces. **Journal of the American Statistical Association**, Taylor & Francis, v. 115, n. 530, p. 852–865, 2020. Citations on pages 23 and 39.

ZUANETTI, D. A.; MILAN, L. A. Data-driven Reversible Jump for QTL mapping. **Genetics**, Genetics Soc America, v. 202, n. 1, p. 25–36, 2016. Citations on pages 23 and 39.

____. A generalized mixture model applied to diabetes incidence data. **Biometrical Journal**, Wiley Online Library, v. 59, n. 4, p. 826–842, 2017. Citations on pages 23 and 39.

ICMC USP
SÃO CARLOS