University of Michigan School of Public Health

The University of Michigan Department of Biostatistics Working Paper Series

Year 2004   *Paper* 48

A Bayesian method for finding interactions in genomic studies

Wei Chen[*]   Debashis Ghosh[†]

Trivellore E. Raghuanthan[‡]   Sharon Kardia[**]

[*]University of Michigan Biostatistics, lisachen@umich.edu

[†]University of Michigan, ghoshd@psu.edu

[‡]University of Michigan, teraghu@umich.edu

[**]University of Michigan

A Bayesian method for finding interactions in genomic studies

Wei Chen, Debashis Ghosh, Trivellore E. Raghuanthan, and Sharon Kardia

Abstract

An important step in building a multiple regression model is the selection of predictors. In genomic and epidemiologic studies, datasets with a small sample size and a large number of predictors are common. In such settings, most standard methods for identifying a good subset of predictors are unstable. Furthermore, there is an increasing emphasis towards identification of interactions, which has not been studied much in the statistical literature. We propose a method, called BSI (Bayesian Selection of Interactions), for selecting predictors in a regression setting when the number of predictors is considerably larger than the sample size with a focus towards selecting interactions. Latent variables are used to infer subset choices based on the posterior distribution. Inference about interactions is implemented by a constraint on the latent variables. The posterior distribution is computed using the Gibbs Sampling methods. The finite-sample properties of the proposed method are assessed by simulation studies. We illustrate the BSI method by analyzing data from a hypertension study involving Single Nucleotide Polymorphisms (SNPs).

A Bayesian method for finding interactions in genomic studies

Wei Chen[1]*

Debashis Ghosh[1]

Trivellore Raghunathan[1]

Sharon Kardia[2]

[1] Department of Biostatistics

[2] Department of Epidemiology

University of Michigan

1420 Washington Heights, Ann Arbor, MI 48109-2029

*To Whom correspondence should be addressed

Contact: lisachen@umich.edu

Tel:(734) 647-3721

Fax:(734) 996-0193

A Bayesian method for finding interactions in genomic studies

Abstract

An important step in building a multiple regression model is the selection of predictors. In genomic and epidemiologic studies, datasets with a small sample size and a large number of predictors are common. In such settings, most standard methods for identifying a good subset of predictors are unstable. Furthermore, there is an increasing emphasis towards identification of interactions, which has not been studied much in the statistical literature. In this article, we propose a method, called BSI (Bayesian Selection of Interactions), for selecting predictors in a regression setting when the number of predictors is considerably larger than the sample size with a focus towards selecting interactions. Latent variables are used to infer subset choices based on the posterior distribution. Inference about interactions is implemented by a constraint on the latent variables. The posterior distribution is computed using the Gibbs Sampling methods. The finite-sample properties of the proposed method are assessed by simulation studies. We illustrate the proposed method by analyzing data from a hypertension study involving Single Nucleotide Polymorphisms (SNPs).

Keywords: Bayesian variable selection; Conditional prior distribution; Constrained Bayes inference; Gibbs sampling; Latent mixture modeling.

1

1 Introduction

The selection of predictors is an important step in building a multiple regression model. Given potential predictors, there are 2^p possible regression models to consider. When the number of potential predictors is larger than the number of samples, conventional methods will yield unstable predictor subsets. The limited sample size undermines the ability to explore all the models. When interaction terms are to be considered as well, identifying predictors becomes even more problematic.

In practice, this problem occurs frequently in genomic epidemiology studies ranging from high throughput microarray experiments to high-density genome scans. Furthermore, the complexity of common diseases motivates more comprehensive studies involving gene-gene interactions, gene-environment, and gene-risk factor interactions. As an example, we consider hypertension. It is well-known that hypertension is the consequence of interactions among many genetic and environmental factors. As part of the Genetic Epidemiology Network of Atherosclerosis (GENOA) study, a large number of single nucleotide polymorphisms (SNPs) were typed in individuals with and without hypertension. Some SNPs were found to be associated with hypertension-related traits when considered univariately (Barkley *et al.*, 2004). In our paper, we extend this research to examine the impact of SNPs, risk factors, and their SNP-risk factors interactions on age of hypertension onset in the non-Hispanic white population. The 57 SNPs from 15 genes on chromosome 2 and 9 risk factors in 342 unrelated individuals were selected from the GENOA study. Therefore, the potential predictors in this analysis consist of 123 main effects (114 dummy variables + 9 risk factors) and 1149 SNP-risk factor interaction terms. Traditional variable selection procedures cannot identify susceptible SNP-risk factor interactions in this situation.

The identification of the predictors in the "large p small n" (West, 2000) setting

2

has motivated a considerable amount of statistical research in recent years. Univariate (individual gene) selection procedures are fairly common, such as those in Rosenwald *et al.* (2002) and Beer *et al.* (2002). Such procedures must deal with the inherent multiple testing issue, which may be avoided by reducing the dimension of the data. For example, West (2000) applied the singular-value decomposition to the design matrix to reduce the dimension of the problem in the Bayesian framework.

A Bayesian procedure for variable selection, the Stochastic Search Variable Selection (SSVS), was proposed by George and McCullough (1993, 1997). It entails the specification of a hierarchical latent mixture prior and uses the posterior probabilities to identify the more promising models. Lee *et al.* (2003) applied the SSVS method to the problem of gene selection with microarray data. Similarly, Sha *et al.* (2004) utilized the multivariate Bayesian variable selection method of Brown *et al.* (1998a, 1998b) to the classifications of microarray data. These methods all focus on the issue of selecting main effects. While some recent scientific research has focused on the importance of the interactions in genetic studies (Longmate 2001; Culverhouse *et al.* , 2002; Devlin *et al.* , 2003), identification of interactions has not been studied much in the statistic literature.

This paper addresses the issue of variable selection, primarily that of interactions selection, when the number of variables greatly exceeds the sample size. We propose a new method called Bayesian Selection of Interactions (BSI). It extends the framework in SSVS and introduces novel priors for interactions. Section 2 specifies the model and describes the prior distributions. In Section 3, we discuss how to implement it using Gibbs sampling (Gelfand and Smith, 1990). Identification of candidate sets of variables is discussed in Section 4. The performance of BSI is assessed using simulation studies in Section 5. In Section 6, BSI is applied to the previously described hypertension study. We conclude with some discussion in Section 7.

3

2   Bayesian variable selection framework

2.1   Normal mixture model

We start with a linear regression model with normal errors:

$$Y|\beta, \sigma \sim N_n(X\beta, \sigma^2 I_n),$$

where Y is an n-vector of response, N_n an n-dimensional multivariate normal distribution, X an $n \times p$ design matrix of predictors, β a p-vector of regression coefficients, σ^2 a scalar, and I_n an $n \times n$ identity matrix. The central concept of the SSVS method is to introduce γ , a p-vector of binary latent variable with components $\gamma_1, \ldots \gamma_p$. It represents the importance of the corresponding regressors. For $j = 1, \ldots p$, if $\gamma_j = 1$ then X_j is included in the model. If $\gamma_j = 0$ then X_j is excluded from the model. A mixture normal prior for the coefficients β is conditioned on γ:

$$\beta_j | \gamma_j \sim (1 - \gamma_j) N_1(0, \tau_j^2) + \gamma_j N_1(0, c_j^2 \tau_j^2) \tag{1}$$

In a matrix form, (1) can be written as: $\beta|\gamma \sim N_p(0, D_\gamma^2)$, where D_γ is a $p \times p$ diagonal matrix with the jth diagonal element equal to $(1-\gamma_j)\tau_j + \gamma_j c_j \tau_j$. For simplicity, the prior distribution assumes the components of β are *a priori* independent. Then the variable selection problem is to make inferences regarding γ.

The choice of appropriate values of τ_j and c_j is important. The recommendations for choosing τ_j and c_j are outlined by George and McCulloch (1997). The key point is to introduce a "practical significance" δ . If $|\beta_j| < \delta_j$ then $\gamma_j = 0$ and X_j would be excluded. That is, δ_j is the intersection point of the densities of $N(0, \tau_j^2)$ and $N(0, c_j^2 \tau_j^2)$. A value of c_j determines the magnitude of the difference between the two mixture normal distributions. Empirical evidence (George and McCulloch, 1997) has shown that c_j between 10 and 100 leads to an efficient computation. For a given δ_j and c_j , τ_j is

4

then obtained as $\tau_j \equiv \left[2c_j^2 \log(c_j)/(c_j^2 - 1)\right]^{-1/2} \delta_j$. Note that the non-conjugate prior distribution (1) of β does not depend on the unknown value of σ, so τ_j does not depend on σ. When a "practical significance" δ_j cannot be meaningfully specified, one might consider:

$$\beta_j | \gamma_j \sim (1 - \gamma_j) I_0 + \gamma_j N_1(0, \tau_j^2),$$

where I_0 is a point mass at 0. For this choice, $\delta_j \equiv 0$ corresponding to the preference that any $\beta_j \neq 0$ be included in the model. This is the setup considered by Geweke (1996), and Brown *et al.* (1998a, 1998b).

Similar to ridge regression, τ_j and c_j will determine the degree of shrinkage for the coefficients β through the posterior distributions. This shrinkage effect allows BSI to handle the "large p small n" problem.

A conjugate Inverse Gamma prior is chosen for the residual variance σ^2:

$$\sigma^2 \sim IG(\nu/2, \nu\lambda/2).$$

This prior can be thought of as an additional likelihood from a dataset with sample size ν and sample variance λ. In the absence of prior information, we suggest using an uninformative prior by taking ν and λ close to zero.

2.2 Prior specification of γ in BSI

2.2.1 Prior choice for γ without interactions

We expect that only a small portion of candidates are associated with the outcome in most high-throughput studies. Based on this prior knowledge, larger weights should be assigned to more parsimonious models. This can be implemented via either of the two different prior distributions for γ, which we call the Beta-Bernoulli and Beta-Binomial prior.

5

The Beta-Bernoulli prior is defined as follows:

$$\gamma_j | \pi_j \quad \sim \quad Bernoulli(\pi_j),$$
$$\pi_j \quad \sim \quad Beta(a,b).$$

This prior allows each γ_j to have a different prior probability π_j, and γ_i and γ_j are *a priori* independent for $i \neq j$. If all the π_j are same ($\pi_j \equiv \pi$ for $j = 1, \ldots, p$), we call this the Beta-Binomial prior. This specification assumes *a priori* marginal dependency between γ_i and γ_j. The covariance of γ_i and γ_j is equal to the variance of the Beta distribution.

To favor parsimonious models, hyperparameters (a, b) of the Beta distribution can be chosen to force π to be small. For example, by specifying the mean and mode for the Beta distribution and solving for (a, b) we obtain

$$a = \frac{\text{mean} - 2 \times \text{mode} \times \text{mean}}{\text{mean-mode}}, \; b = \frac{(1 - 2 \times \text{mode}) \times (1 - \text{mean})}{\text{mean-mode}}$$

Intuitively, the mode is the most likely proportion of all potential predictors to be selected in the model. The mean corresponds to the expected proportion of all potential predictors to be included in the model. The difference of mean and mode controls the variation in the prior distribution. The smaller the difference, the more informative is the prior distribution. For example, if our initial guess that the most likely number of selected predictors is 10 for a dataset with 1000 potential predictors, we specify the mode as 10/1000. If we specify the mean as 12/1000 (20% larger than the mode), then the values of the parameters (a, b) are (5.88,484.12). The sensitivities with respect to the choice of parameters (a, b) are assessed in the simulation studies described in Section 5.

6

2.2.2 Prior choice for γ with interactions

In some studies, the interactions are of interest as well. We adopt the convention that a model containing an interaction term should also contain the corresponding main effects (Neter *et al.*, 1996), since the interaction represents a deviation from an additive model. The previously specified prior distribution for γ can be adjusted to manifest this convention.

Without loss of generality, suppose that there are three main effects A, B and C. Let the corresponding latent variables for the main effects and the pair-wise interactions be $(\gamma_A, \gamma_B, \gamma_C)$ and $(\gamma_{AB}, \gamma_{AC}, \gamma_{BC})$, respectively. The joint prior can be factored by assuming that $(\gamma_{AB}, \gamma_{AC}, \gamma_{BC})$ are independent given π, and $(\gamma_A, \gamma_B, \gamma_C)$ are independent given all the interactions and π:

$$
\begin{aligned}
&p(\gamma_A, \gamma_B, \gamma_C, \gamma_{AB}, \gamma_{AC}, \gamma_{BC}|\pi) \\
=\ & p(\gamma_{AB}, \gamma_{AC}, \gamma_{BC}|\pi)p(\gamma_A, \gamma_B, \gamma_C|\gamma_{AB}, \gamma_{AC}, \gamma_{BC}, \pi) \\
=\ & p(\gamma_{AB}|\pi)p(\gamma_{AC}|\pi)p(\gamma_{BC}|\pi)\cdot \\
& p(\gamma_A|\gamma_{AB}, \gamma_{AC}, \gamma_{BC}, \pi)p(\gamma_B|\gamma_{AB}, \gamma_{AC}, \gamma_{BC}, \pi)p(\gamma_C|\gamma_{AB}, \gamma_{AC}, \gamma_{BC}, \pi).
\end{aligned}
$$

This can be further simplified by assuming that the importance of the main effect only depends on the importance of its interaction terms. That is γ_A depends on $(\gamma_{AB}, \gamma_{AC}, \pi)$ but not γ_{BC}. Thus, the joint prior is

$$
\begin{aligned}
&p(\gamma_A, \gamma_B, \gamma_C, \gamma_{AB}, \gamma_{AC}, \gamma_{BC}|\pi) \\
=\ & p(\gamma_{AB}|\pi)p(\gamma_{AC}|\pi)p(\gamma_{BC}|\pi) \\
& \cdot (\gamma_A|\gamma_{AB}, \gamma_{AC}, \pi)p(\gamma_B|\gamma_{AB}, \gamma_{BC}, \pi)p(\gamma_C|\gamma_{AC}, \gamma_{BC}, \pi).
\end{aligned} \tag{2}
$$

These two assumptions correspond to the commonly used principles for variable selection. The dependence between the main effect and the interactions is defined by the

components of joint prior (2). For example:

$$p(\gamma_A = 1|\gamma_{AB}, \gamma_{AC}, \pi) = \begin{cases} 1 & \text{if } \max(\gamma_{AB}, \gamma_{AC}) = 1 \\ \pi & \text{if } \max(\gamma_{AB}, \gamma_{AC}) = 0 \end{cases}$$

$$p(\gamma_{AB} = 1) = p(\gamma_{AC} = 1) = \pi \tag{3}$$

$$\pi \sim Beta(a, b)$$

We call this the Conditional-Beta-Binomial prior distribution. If π is different for each main and interaction term, we call this the Conditional-Beta-Bernoulli prior. This prior specification reduces the model space greatly. Note that the unconditional prior is a special case of the conditional prior when there is no interaction term.

3   Gibbs sampling algorithm

When the number of predictors p is large, it is impractical to exhaustively evaluate the posterior probabilities for 2^p models. Indeed, most of the models have a very small probability and will appear very rarely. These models can be ignored. We use a Gibbs sampling algorithm to estimate the posterior distribution as used in SSVS focuing on the models that have high probabilities.

The Gibbs sampler simulates a Markov chain $\beta^{(1)}, \sigma^{(1)}, \gamma^{(1)}, \pi^{(1)}, \beta^{(2)}, \sigma^{(2)}, \gamma^{(2)}, \pi^{(2)}, \ldots$ from the full conditionals which converge to the joint posterior $p(\beta, \sigma, \gamma, \pi | Y)$ (Gelfand and smith, 1990). It can be shown that the full conditional distributions for β and σ^2 are normal and inverse Gamma, respectively. Their density functions are:

$$\begin{aligned} p(\beta|\sigma^2, \gamma, \pi, Y) &= p(\beta|\sigma^2, \gamma, Y) \\ &= N_p((X^T X + \sigma^2 D_\gamma^{-2})^{-1} X^T Y, \sigma^2 (X^T X + \sigma^2 D_\gamma^{-2})^{-1}) \\ p(\sigma^2|\beta, \gamma, \pi, Y) &= p(\sigma^2|\beta, Y) \\ &= IG\left(\frac{n+\nu}{2}, \frac{|Y - X\beta|^2 + \nu\lambda}{2}\right). \end{aligned}$$

Using Bayes rule, the full conditional distribution of γ using conditional-Beta-Binomial prior is Bernoulli with probability

$$
\begin{aligned}
p(\gamma_j = 1|\beta, \sigma^2, \gamma_{-j}, Y, \pi) &= p(\gamma_j = 1|\beta, \gamma_{-j}, \pi) \\
&= \frac{p(\gamma_j = 1, \beta, \gamma_{-j}, \pi)}{p(\gamma_j = 1, \beta, \gamma_{-j}, \pi) + p(\gamma_j = 0, \beta, \gamma_{-j}, \pi)}
\end{aligned}
$$

where γ_{-j} is a vector of the elements of γ except for the jth element . Define a p-vector of indicator variables $\xi \equiv (\xi_1, \ldots, \xi_p)$ as $\xi_j = 0$ if an interaction, and $\xi_j = 1$ if a main effect, $j = 1, \ldots, p$. Let Ω_j denote the set that includes all the interactions related to the jth main effect, and thus the full conditional distribution of γ is as follows (see Appendix for the derivation):

$$
\begin{aligned}
&p(\gamma_j = 1|\beta, \sigma^2, \gamma_{-j}, Y, \pi) \\
&= \begin{cases} \frac{p(\beta_j|\gamma_j=1)\pi}{p(\beta_j|\gamma_j=1)\pi + p(\beta_j|\gamma_j=0)(1-\pi)} & \text{when } \xi_j = 0 \text{ or } \xi_j = 1 \text{ and } \sum_{\gamma_{j'} \in \Omega_j} \gamma_{j'} = 0 \\ 1 & \text{when } \xi_j = 1 \text{ and } \sum_{\gamma_{j'} \in \Omega_j} \gamma_{j'} > 0. \end{cases}
\end{aligned}
$$

Note that the final equation is a simple form given the independent prior for β in (1). The full conditional distribution for γ using Conditional-Beta-Bernoulli prior is:

$$
\begin{aligned}
&p(\gamma_j = 1|\beta, \sigma^2, \gamma_{-j}, Y, \pi_j) \\
&= \begin{cases} \frac{p(\beta_j|\gamma_j=1)\pi_j}{p(\beta_j|\gamma_j=1)\pi_j + p(\beta_j|\gamma_j=0)(1-\pi_j)} & \text{when } \xi_j = 0 \text{ or } \xi_j = 1 \text{ and } \sum_{\gamma_{j'} \in \Omega_j} \gamma_{j'} = 0 \\ 1 & \text{when } \xi_j = 1 \text{ and } \sum_{\gamma_{j'} \in \Omega_j} \gamma_{j'} > 0. \end{cases}
\end{aligned}
$$

The derivation is similar and thus omitted.

The full conditional distribution for π using Conditional-Beta-Binomial prior is:

$$
\pi|\beta, \sigma^2, \gamma, Y \sim Beta(a + \sum_{j=1}^{p} \gamma_j - q, \ p - \sum_{j=1}^{p} \gamma_j + b),
$$

where $q \equiv \sum_{j=1}^{p} I_{\{\sum_{\gamma_{j'} \in \Omega_j} \gamma_{j'} > 0, \ \xi_j = 1\}}$ is the number of main effects that are forced to be in the model due to the interactions (See the Appendix for the derivation). Note that the first scale parameter of this Beta distribution is reduced by a quantity of q, since

9

some main effects do not provide any information on π as in (3). Similarly, the full conditional distribution for π_j using Conditional-Beta-Bernoulli is

$$\pi_j | \beta, \sigma^2, \gamma, Y \sim Beta\left(a + \gamma_j - I_{\{\sum_{\gamma_{j'} \in \Omega_j} \gamma_{j'} > 0,\ \xi_j = 1\}},\ 1 - \gamma_j + b\right)$$

These conditionals are standard distributions which can be simulated by routine methods. When p is large, the number of possible models 2^p in the model space often far exceeds the number of iterations of the Gibbs sampler. The convergence rate and criteria of the MCMC algorithm become a critical issue. The correlation structure of a design matrix can have a great impact on the convergence rate of MCMC methods (Geweke, 1996). Making the explanatory variables orthogonal to each other can strongly improve convergence and mixing (Clyde *et al.* 1996, Gelfand *et al.* 1995). Therefore, we standardize each variable in the data set before we start the variable selection procedure.

There is much literature regarding convergence criteria of Markov Chain Monte Carlo algorithms. A good review can be found in Mengersen *et al.* (1999). We use Rubin's R (Gelman *et al.*, 1995), between and within sequence variance, to monitor the convergence for each scalar estimate, such as σ and $\sum_{j=1}^{p} \gamma_j$. Ten independent sequences were simulated simultaneously with starting points drawn from the overdispersed distributions. For example, for the initial value of β, ridge regression estimates were used, given different shrinkage values. The shrinkage values are between zero and the smallest eigenvalue of the matrix $X^T X$. Here X is the design matrix.

4   Selection of predictors

Selecting the predictors under the joint posterior distribution is commonly used in a Bayesian framework. However, when the number of predictors greatly exceeds the sample size, the prior has a substantial impact on the posterior distribution. A joint

10
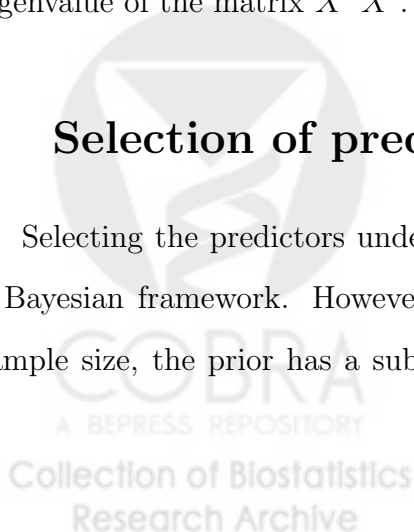
prior distribution $p(\gamma)$ which favors the models with few or no predictors leads the posterior to favor such models. This property is observed in a simulation study as well (data not shown). Furthermore, the model space is extremely large when p is large. Most models will only appear once in the stochastic process.

A simple illustration of the prior probability of γ through a simple example with 2 main effects and 1 interaction is shown in Table 1. Under the unconditional prior, the model space consists of 8 possible combinations. As stated in Section 2, Table 1 shows that the model space is smaller when we impose the constraint due to the interactions. We let $w = a(a + b)^{-1}$ be the mean of the Beta distribution. The smaller the w, the bigger the joint probability $p(\gamma)$ of the null model. The null model is defined as no variable selected. If we set $w = 0.01$, then the prior odds ratio of the null model to a one variable selected model is $(1 - w)/w = 99$, to a two variables selected model is $(1 - w)^2/w^2 = 9801$. This weighting scheme leads to the more parsimonious models. While the joint prior probability varies according to the number of regressors in the model, the marginal prior probabilities are invariant. Under the unconditional prior, the $p(\gamma_j)$ are the same. Under the conditional prior, the the $p(\gamma_j)$ of the main effects are higher but still identical to each other, and that of the interactions are same as before. Extending this result, the marginal prior probability of a main effect selected is $1 - (1 - w)^{\nu+1}$ and that of an interaction term is w, where ν is the number of interactions that related to the main effect in question. As ν increases, the marginal prior probability for that main effect increases. If only two-way pair-wise interactions are considered, then $\nu = p - 1$.

We select the important predictors based on marginal posterior probabilities $P(\gamma_j|Y)$, which is estimated by the occurrence of selection of a variable across the iterations of the Markov Chain. The resulting list of predictors is referred to as an importance list. However, a model with all the important predictors selected by the marginal posterior

11
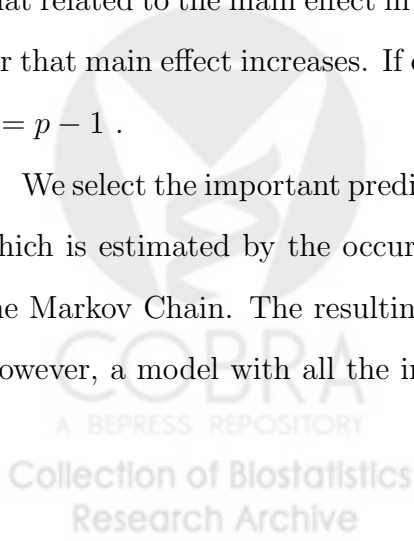
probabilities may never appear in the iterations of Gibbs sampling. For prediction purposes, it is desirable to select under the joint posterior probability. In this case, we propose a two-stage selection method. Stage I is for the variable selection. We perform BSI on all the potential predictors and rank the predictors according to the estimated $P(\gamma_j|Y)$. The top k interactions including the main effects are selected based on this quantity. (Section 5 discuss the guidelines for choosing k). Stage II is for one model selection or model averaging. BSI is conducted again on the selected terms, but we rank the models according to the estimated $P(\gamma|Y)$. If one model is preferred in prediction, we select the model with the highest posterior probability. If model averaging is preferred, we keep all the visited distinctive models or a subset of models with high posterior probabilities.

5   Simulation study

In this section, we evaluate the BSI through simulation studies. First, we compare the performance of the Beta-Bernoulli prior versus the Beta-Binomial prior when interactions are not considered. Second, we compare the performance of the Conditional-Beta-Binomial versus the Beta-Binomial when interactions are considered. For each comparison, the sensitivity of the hyperparameters is also studied.

Some factors that may affect the performance of the BSI are considered during the study design, such as the possible sample circumstances, the method's underlying assumptions, and the choice of parameters in the hyperprior. For simplicity, we restricted our attention to four factors: the correlation coefficient ρ among the potential predictors, the prior distributions for γ , the hyperparameters of the Beta distribution, and the cutoff value k in Stage I.

We constructed a sample of size $n = 20$ on $p = 50$ potential predictors by gen-

12

erating $z, z_1, \ldots, z_{50}$ i.i.d. $\sim N_{20}(0, I_{20})$ and setting $X_j = z_j + z \left(\rho(1-\rho)^{-1} \right)^{1/2}$ for $j = 1, \ldots, 50$. This produced a pair-wise correlation of ρ among $X_1, \ldots, X_p$, and $\rho = (0.1, 0.3, 0.5, 0.7, 0.9)$. We drew $Y \sim N_{20}(X_1 + 1.5X_2 + 3X_3, \sigma^2 I_{20})$ where $\sigma = 1$. For each value of ρ, we fixed the simulation replications at 250. Two prior distributions were used for γ: the Beta-Bernoulli prior and the Beta-Binomial prior. For the hyperprior Beta distribution, two sets of the (mode, mean) were used: (0.1, 0.12) and (0.2, 0.22). At Stage I, we used different cutoff values $k = (2, 4, 6, 8, 10, 12, 14)$ to obtain subsets resulting from the rank of the $P(\gamma_j | Y)$.

The true and false positive rates of identifying the true predictors for various choices of k were calculated. Receiver Operating Characteristic (ROC) curves were constructed by plotting the true positive rate (sensitivity) against the false positive rate (1- specificity) for the different possible cut off values. The area under the curve measures how well the method separates the true predictors and noise. In Figure 1(a), we provide the ROC curves of sensitivity analysis of prior distribution: Beta-Bernoulli vs. Beta-Binomial. The sensitivity analysis of the choice of the hyperparameters (mode, mean) for Beta distribution is shown in Figure 1(b).

Both figures show the choice of the prior distribution or the parameters in hyperprior (even when we double the value of the mode) is not sensitive in our method. Furthermore, an analysis of variance (ANOVA) test of four factors shows that the prior distribution and parameters in hyperprior are not significant factors in terms of the true positive rates (See Table 2).

As we expected, the higher the correlation among the potential predictors, the lower the true positive rates. If high collinearity exists in the data, the selection of predictors will be affected by the 'dilution' effect (George, 1999). This happens because the correlated predictors change the relative marginal posterior probability allocation to those predictors. The sensitivity is also highly associated with cutoff value k. The higher the

13
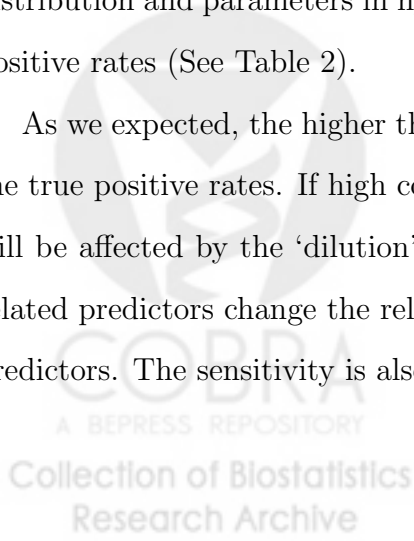
cutoff value k, the higher the true positive rate. The simulations presented in this paper have the ratio of p/n equal to about 2.5. Higher ratios are studied for the power of BSI. According to the empirical study we suggest a cutoff range for k, $3\times$ mode $\times p \leq k < n$. The mode is the initial guess of the most likely proportion of all potential predictors to be selected in the model. If the ratio of p/n is larger than 10, select k close to n.

We proceeded to evaluate the proposed conditional prior for variable selection with interaction terms. A sample of size $n = 20$ on 10 main effects and 45 pair-wise interaction terms was generated. We induced a pair-wise correlation of $\rho = (0.1, 0.3, 0.5, 0.7, 0.9)$ among the main effects in the same manner as in the previous sample. We drew $Y \sim N_{20}(X_1 + X_2 + X_3 + X_4 + X_1 X_2 + X_1 X_4, \sigma^2 I_{20})$, where $\sigma = 1$. Again, for each value of ρ, we fixed the simulation replications at 250. Two prior distributions were used for γ: Conditional-Beta-Binomial and Unconditional-Beta-Binomial. For the beta distribution, the (mode,mean) were set at (0.12, 0.14). We ran the Stage I variable selection. For the main effects (Figure 2(a)) and the interaction terms (Figure 2(b)), the proposed conditional priors always perform better than the unconditional prior in terms of the true positive rate. The inclusion of the main effects whenever a related interaction term is selected helps to identify the interaction term itself.

Different from the previous result for the main effects only simulation, an interesting fact that we observed here is that high multicollinearity among the main effects helps in the identification of the true predictors. For example, when the main effects A and B are highly correlated, in part due to the surrogate effect of B, A and B both help identifying the important interaction term, say AC. Therefore, there is a higher probability of selecting AC. We also observed this phenomena in other simulation studies.

14
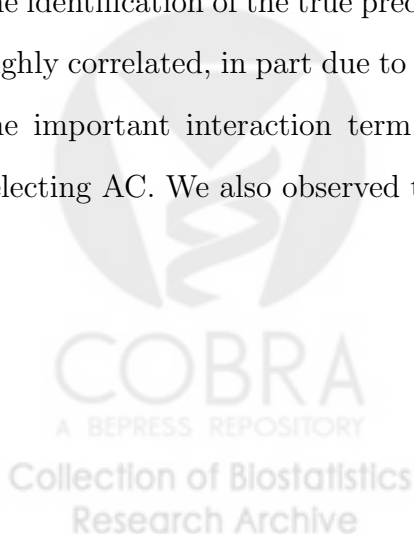
6 Analysis of Hypertension Data

A detailed description of the GENOA sampling design and strategy are presented elsewhere (O'Meara *et al.*, submitted). The data consists of 342 unrelated, hypertensive non-Hispanic white individuals from the Mayo Clinic field center of the GENOA study who had complete SNP and risk factor data. The 57 SNPs (Table 3) were from 15 genes on chromosome 2 where there had been previous linkage evidence of genes for hypertension (Province *et al.*, 2003). The hypertension risk factors included in our study are age (yrs), weight (kg), height (cm), BMI (kg/m2), family history of hypertension, fasting plasma cholesterol (mg/dl), fasting plasma triglycerides (mg/dl), systolic blood pressure (mmHg), and diastolic blood pressure (mmHg). Even though the BMI is correlated to weight and height, we still include all of them and rely on the BSI method to select the important effects and build a prediction model.

In this section, we illustrate the application of proposed method to the selection of predictors for hypertension age of onset. Since each SNP is trichotomous by the nature of the genotype, we code each of them into 2 dummy variables. Adopting the convention of "all include" or "all exclude" (Neter *et al.*, 1996), we assign a single latent variable γ_j to a group of dummy variables. That is a many-to-one mapping (Chipman, 1996) from the vector of regression coefficients to the vector of latent variables. This strategy greatly reduces the model space in this data. Furthermore, Chipman (1996) discussed that when a large group of variables is not important in terms of prediction, the chances are large that one variable will appear important because of random variation. By only considering the importance of the group instead of each variable in that group, the chances of making such an error are reduced.

Only the SNP-risk factor interactions were considered in our hypertension data. SNP-SNP interactions were not included because rarely do we have all combinations of

15

genotypes from the data. For example, a 3 by 3 frequency table for any SNP pair often shows that we have zero count in some allele combinations. This scenario is common due to linkage disequilibrium (LD) and the rare relative allele frequencies.

To obtain the prior knowledge of the hyperparameters for the Beta distribution, we used classical two-stage stepwise selection to identify the possible number of predictors. Briefly, in stage one, we used stepwise selection to identify the main effects. In stage two, we used the stepwise selection again on all the selected main effects and their SNP-risk factor interactions, forcing the main effects to be included in the model. The selected predictors showed in Table 5. Among 579 possible predictors, we have 7 predictors chosen by the stepwise selection method. The ratio is about 0.012.

The two-stage stepwise method was chosen for the study purely for its simplicity in implementation. However, it has disadvantages besides the well-known problem of the F statistics. First, if the number of main effects to start with already exceeds the number of samples, we cannot obtain a reasonable model without specifying a maximum number of terms included. The selection of a maximum number of terms could be subjective without cross-validation, which is computationally intensive. Second, SNPs without strong association with the outcome will be eliminated in the first stage. Therefore, some SNPs which may contribute through the SNP-risk factor interaction will not be considered in the second stage.

We next performed Stage I of the selection procedure. The Conditional-Beta-Bionomial prior was chosen for the latent variable γ. The mode and mean of the Beta prior were set at (0.012, 0.018) according to the previous discussion. The trace of σ^2 and $\sum_{j=1}^{579} \gamma_j$ from 10 independent chains are shown in Figure 3(a) and Figure 3(b), respectively.

Figure 4 shows the marginal posterior distributions of γ for 579 regressors. The spike in the graph corresponds to age, which makes sense. The age would be expected to be highly correlated to the age of onset of hypertension. We ranked the $P(\gamma_j|Y)$ at

16

Stage I. Table 4 shows the top 20 main effects and the top 5 interaction terms associated with hypertension. It is plausible that multiple hypertension genes are located in the region examined. Thus, the present findings are suggestive and should be validated with follow-up studies.

To find the "best" prediction model, we proceeded to the Stage II selection using the main effects and the interactions in Table 4. The mode and mean of the Beta prior at Stage II were set at (0.45, 0.50), which defines a uniform distribution for the probability of $\gamma_j = 1$, where $j = 1, \ldots, 25$. Table 5 shows the top 3 ranked models selected by Stage II. Our preliminary analysis identified a "best" model with 2 risk factors: age and strong family history; two SNPs: AUP1_cv11727841 and SLC20A1_CV9546434; and one interaction term: strong family history $\times$ AUP1_cv11727841. Its estimated coefficients and p-value are showed in Table6. Age, SLC20A1_CV9546434, AUP1_cv11727841, and strong family history $\times$ AUP1_cv11727841 reached the statistical significance for age of hypertension onset.

SLC20A1 is a sodium-dependent phosphate transporter with expression across tissues, including the kidney. The protein encoded by AUP1 has been shown to bind to the conserved membrane-proximal sequence of the cytoplasmic tail of integrin alpha(IIb) subunits. These subunits play a crucial role in the integrin alpha(IIb) beta(3) inside-out signaling ($http://www.ncbi.nlm.nih.gov$). Future studies should attempt to replicate these findings in other populations. The adjusted R-square of this model is 0.3161, which is similar to 0.3191 from a model selected by stepwise selection method (See Table5). We notice that BSI method achieves comparable predictive accuracy through a more parsimonious model.

17
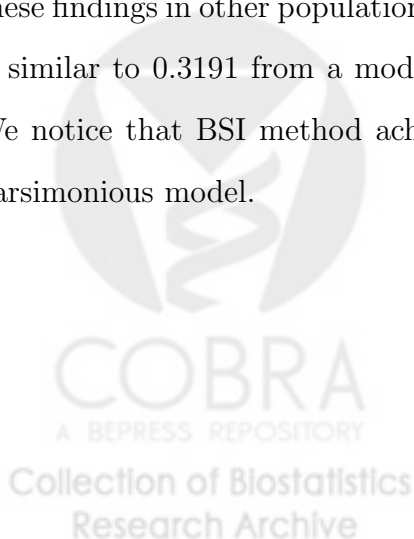
7 Discussion

BSI is quite general and can be applied in many variable selection or model building problems, especially the challenging problems such as limited sample size and related predictors. The same idea of constructing the prior for interactions can be extended to exploring the functional terms of predictors. For example, the lower order polynomials are included if a higher order one is selected. BSI also can be applied in a variety of scientific disciplines. For instance, it can be applied in the epistasis (the interaction between two or more genes to control a single phenotype) in genetic study, the interaction effects of the biomarkers in pharmacogenomic study, and the gene selections in high-throughput microarray analysis.

To force the inclusion of the main effects when the interaction term is selected, we imposed a constraint on γ. An alternative method is to impose constraints on β. The general idea is that if a predictor is important, the magnitude of its coefficient is large. Consider the same example with 3 main effects in section 2.2.2. Mathematically this is accomplished by modifying the equation (1) to:

$$p(\beta_A|\gamma_j) = (1 - \gamma_A)(1 - \gamma_{AB})(1 - \gamma_{AC})N(0, \tau_A^2)$$
$$+ \left[1 - (1 - \gamma_A)(1 - \gamma_{AB})(1 - \gamma_{AC})\right] N(0, c_A^2 \tau_A^2)$$

Since our primary interest is in the posterior distribution of γ, we chose imposing constraints directly on γ due to its computational efficiency using MCMC.

To select a "best" model, we proposed a two-stage selection. If a subset of important predictors is desired instead of a predictive model, the importance list provided by stage I selection is sufficient.

Although variable selection in the linear regression model is discussed in this paper, we expect that the method can be extended to handle the generalized linear model (McCullagh and Nelder, 1989) or proportional hazard model (Cox, 1972). For example,

18

using the idea of data augmentation(Tanner and Wong, 1987), the probit regression model for binary outcomes has an underlying normal regression structure on latent continuous data. The method can also be generalized to multinomial response models with more than 2 categories (Albert and Chib, 1993). This data augmentation can also be used in proportional hazard model for censored survival times. We are currently exploring this area.

Appendix

.1 Full conditional distribution of γ using conditional-Beta-Binomial prior

Using Bayes' rule, the full conditional distribution of γ is Bernoulli

$$
\begin{aligned}
p(\gamma_j = 1|\beta, \sigma^2, \gamma_{-j}, Y, \pi) &= p(\gamma_j = 1|\beta, \gamma_{-j}, \pi) \\
&= \frac{p(\gamma_j = 1, \beta, \gamma_{-j}, \pi)}{p(\gamma_j = 1, \beta, \gamma_{-j}, \pi) + p(\gamma_j = 0, \beta, \gamma_{-j}, \pi)},
\end{aligned}
$$

where γ_{-j} is a vector of the elements of γ except for the jth element γ_j. Let $U \equiv p(\gamma_j = 1, \beta, \gamma_{-j}, \pi)$, we have

$$
U = p(\gamma_j = 1, \beta, \gamma_{-j}, \pi)
$$

19

$$= p(\beta|\gamma_j = 1, \gamma_{-j}, \pi)p(\gamma_j = 1, \gamma_{-j}|\pi)p(\pi)$$

$$= p(\beta_j|\gamma_j = 1)p(\beta_{-j}|\gamma_{-j})p(\gamma_j = 1|\gamma_{-j}, \pi)p(\gamma_{-j}|\pi)p(\pi).$$

Let $V \equiv p(\gamma_j = 0, \beta, \gamma_{-j}, \pi)$, we have

$$V = p(\beta_j|\gamma_j = 0)p(\beta_{-j}|\gamma_{-j})p(\gamma_j = 0|\gamma_{-j}, \pi)p(\gamma_{-j}|\pi).$$

Define a p-vector of indicator variables $\xi \equiv (\xi_1, \ldots, \xi_p)$ as $\xi_j = 0$ if an interaction, and $\xi_j = 1$ if a main effect, $j = 1, \ldots, p$. Let Ω_j be the set that includes all the interactions related to the jth main effect. Substituting U and V back into the full conditional distribution and simplifying the equation, we have

$$p(\gamma_j = 1|\beta, \sigma^2, \gamma_{-j}, Y, \pi) = \frac{U}{U+V}$$
$$= \frac{p(\beta_j|\gamma_j = 1)p(\gamma_j = 1|\gamma_{-j}, \pi)}{p(\beta_j|\gamma_j = 1)p(\gamma_j = 1|\gamma_{-j}, \pi) + p(\beta_j|\gamma_j = 0)p(\gamma_j = 0|\gamma_{-j}, \pi)}$$
$$= \frac{p(\beta_j|\gamma_j = 1)\pi^{1-I_{\{\sum_{\gamma_{j'} \in \Omega_j} \gamma_{j'} > 0, \ \xi_j = 1\}}}}{p(\beta_j|\gamma_j = 1)\pi^{1-I_{\{\sum_{\gamma_{j'} \in \Omega_j} \gamma_{j'} > 0, \ \xi_j = 1\}}} + p(\beta_j|\gamma_j = 0)\pi^{1-I_{\{\sum_{\gamma_{j'} \in \Omega_j} \gamma_{j'} > 0, \ \xi_j = 1\}}}}$$
$$= \begin{cases} \frac{p(\beta_j|\gamma_j = 1)\pi}{p(\beta_j|\gamma_j = 1)\pi + p(\beta_j|\gamma_j = 0)(1 - \pi)} & \text{when } \xi_j = 0 \text{ or } \xi_j = 1 \text{ and } \sum_{\gamma_{j'} \in \Omega_j} \gamma_{j'} = 0 \\ 1 & \text{when } \xi_j = 1 \text{ and } \sum_{\gamma_{j'} \in \Omega_j} \gamma_{j'} > 0 \end{cases}$$

.2 Full conditional distribution for π using conditional-Beta-Binomial prior

The derivation of the full conditional distribution for π follows from the following fact. If I_A and I_B are two binary indicator variables, x is any real number, and $I_B = 1$ when $I_A = 1$, then $\left(1 - x^{1-I_A}\right)^{1-I_B} = (1 - x)^{(1-I_A)(1-I_B)}$. The proof of this fact is straightforward and is omitted.

20
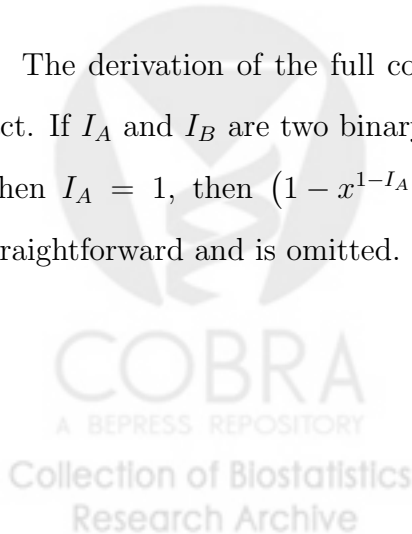
The full conditional distribution for π is

$$p(\pi|\beta, \sigma^2, \gamma, Y) = p(\pi|\gamma)$$

$$\propto p(\gamma|\pi)p(\pi)$$

$$\propto \prod_{j=1}^{p} \left(\pi^{1-I_{\{\sum_{\gamma_{j'} \in \Omega_j} \gamma_{j'} > 0, \ \xi_j = 1\}}} \right)^{\gamma_j} \left(1 - \pi^{1-I_{\{\sum_{\gamma_{j'} \in \Omega_j} \gamma_{j'} > 0, \ \xi_j = 1\}}} \right)^{1-\gamma_j} \left(\pi^{a-1}(1-\pi)^{b-1} \right)$$

$$= \prod_{j=1}^{p} \pi^{\left(1-I_{\{\sum_{\gamma_{j'} \in \Omega_j} \gamma_{j'} > 0, \ \xi_j = 1\}} \right)\gamma_j} (1-\pi)^{\left(1-I_{\{\sum_{\gamma_{j'} \in \Omega_j} \gamma_{j'} > 0, \ \xi_j = 1\}} \right)(1-\gamma_j)} \left(\pi^{a-1}(1-\pi)^{b-1} \right)$$

$$= \pi^{\sum_{j=1}^{p} \gamma_j - \sum_{j=1}^{p} I_{\{\sum_{\gamma_{j'} \in \Omega_j} \gamma_{j'} > 0, \ \xi_j = 1\}}} (1-\pi)^{p - \sum_{j=1}^{p} \gamma_j} \left(\pi^{a-1}(1-\pi)^{b-1} \right)$$

$$= Beta(a + \sum_{j=1}^{p} \gamma_j - q, p - \sum_{j=1}^{p} \gamma_j + b),$$

where $q \equiv \sum_{j=1}^{p} I_{\{\sum_{\gamma_{j'} \in \Omega_j} \gamma_{j'} > 0, \ \xi_j = 1\}}$ is the number of main effects that are forced to be in the model due to the interactions.

References

[1] Albert, J.H., and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88**, 669-679.

[2] Barkley, R.A., Chakravarti, A., Cooper, R.S., Ellison, R.C., Hunt, S., Province, M.A., Turner, S.T., Weder, A. B., and Boerwinkle, E. (2004). Positional identification of hypertension susceptibility genes on chromosome 2. *Hypertension* **43**, 477-482.

[3] Beer, D.G., Kardia, S.L., Huang, C.C., Giordano, T.J., Levin, A.M., Misek, D.E., Lin, L., Chen, G., Gharib, T.G., Thomas, D.G., Lizyness, M.L., Kuick, R., Hayasaka, S., Taylor, J.M., Iannettoni, M.D., Orringer, M.B., and Hanash, S. (2002). Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nature Medicine* **8**, 816-824.

21
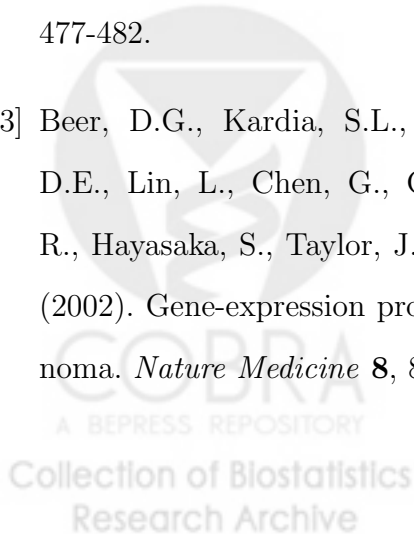
[4] Brown, P.J., Vannucci, M., and Fearn, T. (1998a). Multivariate Bayesian variable selection and prediction. *Journal of the Royal Statistical Society, Series B* **60**, 627-641.

[5] Brown, P.J., Vannucci, M., and Fearn, T. (1998b). Bayesian wavelength selection in multicomponent analysis. *J. of Chemometrics* **12**, 173-182.

[6] Chipman, H. (1996). Bayesian variable selection with related predictors. *The Canadian Journal of Statistics* **24**, 17-36.

[7] Clyde, M., DeSimone, H., and Parmigiani, G. (1996). Prediction via orthogonalized model mixing. *Journal of the American Statistical Association* **91**, 1197-1208.

[8] Cox, D.R. (1972). Regression Models and Life Tables. *Journal of the Royal Statistical Society, Series B* **34**, 187-220.

[9] Culverhouse, R., Suarez, B.K., Lin, J., and Reich, T. (2002). A perspective on epistasis: limits of models displaying no main effect. *Am J Hum Genet* **70**, 461-471.

[10] Devlin, B., Roeder, K., and Wasserman L. (2003). Analysis of multilocus models of association. *Genetic Epidemiology* **25**, 36-47.

[11] Gelfand, A.E., Sahu, S., and Carlin, B.P. (1995). Efficient parameterization for normal linear mixed effects models. *Biometrika* **82**, 479-488.

[12] Gelfand, A.E., and Smith, A.F.M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398-409.

[13] Gelman, A. B., Carlin J.S., Stern H. S., and Rubin D. B. (1995). *Bayesian Data Analysis*. Chapman & Hall.

22

[14] George, E.I., and McCulloch, R.E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association* **88**, 881-889.

[15] George, E.I., and McCulloch, R.E. (1997). Approaches for Bayesian variable selection. *Statistica Sinica* **7**, 339-374.

[16] George, E.I. (1999). Discussion of Bayesian model averaging and model search strategies by M.A. Clyde. *Bayesian Statistics 6* (J.M. Bernardo, J.O. Berger, A.P. David and A.F.M. Smith, eds.) Oxford University Press.

[17] Geweke, J. (1996). Variable selection and model comparison in regression. *Bayesian Statistics* **5**, 609-620.

[18] Lee, K.E., Sha, N., Dougherty, E.R., Vannucci, M., and Mallick, B.K. (2003). Gene selection: a Bayesian variable selection approach. *Bioinformatics* **19**, 90-97.

[19] Longmate, J.A. (2001). Complexity and power in case-control association studies. *Am J Hum Genet* **68**, 1229-1237.

[20] McCullagh, P., and Nelder J.A. (1989). *Generalized Linear Models.* London: Chapman and Hall, 2nd edition.

[21] Mengersen, K.L., Robert, C.P., and Guihenneuc-Jouyaux, C. (1999). MCMC Convergence Diagnostics: A Review. *Bayesian Statistics* **6**, 415-440.

[22] Neter, J., Kutner, M.H., Nachtsheim, C.J., and Wasserman, W. (1996). *Applied linear statistical models.* The McGraw-Hill Companies, Inc.

[23] O'Meara, J., Kardia, S. L., Armon, J.J., and Turner, S.T. Prevalence, treatment and control of dyslipidemia in a hypertensive population. *Archives of Internal Med.* Submitted.

23

[24] Province, M.A., Kardia, S.L.R., Ranade, K., Rao, D.C., Thiel, B.A., Cooper, R.S., Risch, N., Turner, S.T., Cox, D.R., Hunt, S.C., Weder A.B., and Boerwinkle, E. (2003). A Meta-Analysis of Genome-Wide Linkage Scans for Hypertension: The National Heart, Lung and Blood Institute Family Blood Pressure Program. *The American Journal of Hypertension* **16**, 144-147.

[25] Rosenwald, A., Wright, G., Chan, W.C., Connors, J.M., Campo, E., Fisher, R.I., Gascoyne, R.D., Muller-Hermelink, H.K., Smeland, E.B., Giltnane, J.M., Hurt, E.M., Zhao, H., Averett, L., Yang, L., Wilson, W.H., Jaffe, E.S., Simon, R., Klausner, R.D., Powell, J., Duffey, P.L., Longo, D.L., Greiner, T.C., Weisenburger, D.D., Sanger, W.G., Dave, B.J., Lynch, J.C., Vose, J., Armitage, J.O., Montserrat, E., Lopez-Guillermo, A., Grogan, T.M., Miller, T.P., LeBlanc, M., Ott, G., Kvaloy, S., Delabie, J., Holte, H., Krajci, P., Stokke, T., Staudt, L.M.;Lymphoma/Leukemia Molecular Profiling Project (2002). The use of molecular profiling to predict survival after chemotherapy for diffuse large B cell lymphoma. *The New England Journal of Medicine* **346**, 1937-1947.

[26] Sha, N., Vannucci, M., Tadesse, M.G., Brown, P.J., Dragoni. I., Davies, N., Roberts, T.C., Contestabile, A., Salmon, M., Buckley, C., and Falciani, F. (2004). Bayesian variable selection in multinomial probit models to identify molecular signatures of disease stage. *Biometrics* **60**, 812-819.

[27] Tanner, M.A. and Wong, W.H. (1987) The calculation of posterior distributions by data augmentation (with discussion), *Journal of the American Statistical Association* **82**, 528-550.

[28] West, M. (2003) Bayesian factor regression models in the "Large p, Small n" paradigm. *Bayesian Statistics* **7**, 723-732.
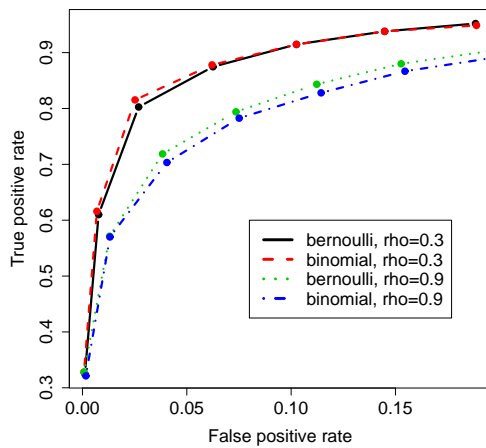
(a) Beta-Bernoulli vs. Beta-Binomial
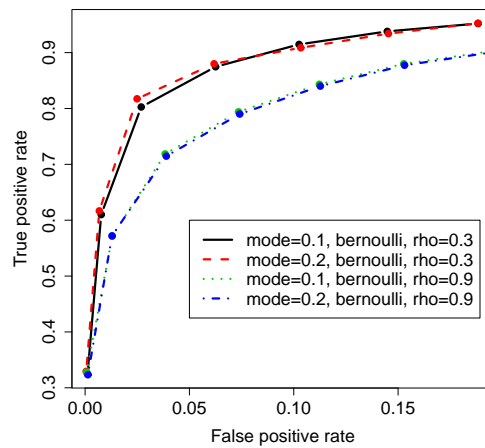
(b) Beta prior(mode,mean): (0.1, 0.12) vs. (0.2, 0.22)

Figure 1. ROC curves of sensitivity analysis with $n = 20$, $p = 50$, replicates=250, and number of true predictors=2. Cutoff points at: 2,4,6,8,10,12,14
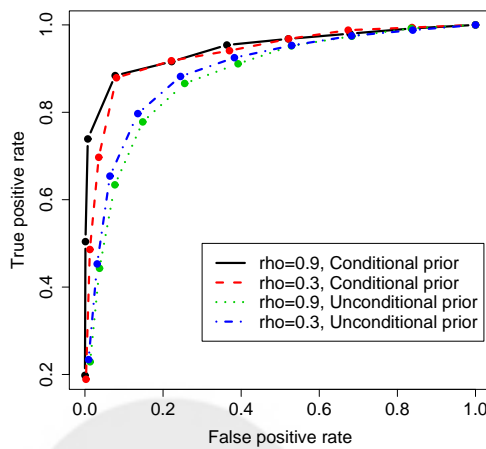
(a) main effects: cutoff points at: $1, \ldots, 10$. Number of true predictors $= 4$.
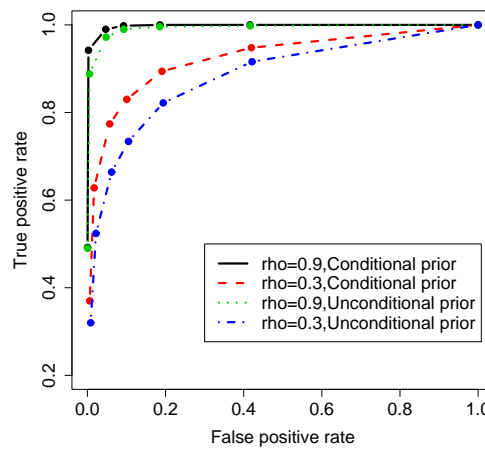
(b) Interactions: cutoff points at: 1, 2, 4, 6, 10, 20, 45. Number of true predictors $= 2$.

Figure 2. ROC curves for comparing Conditional vs. Unconditional prior with (mode, mean)=(0.12, 0.14), $n = 20$, main effects=10, interactions=45, and replicates=250.

25

(a) Trace of σ^2

(b) Trace of $\sum_{j=1}^{579} \gamma_j$

Figure 3. Trace from 10 chains at selection Stage I.

Figure 4. The marginal posterior distributions of γ for 579 regressors at stage I. The first 66 are the main effects. The rest are the interaction terms.

26

Table 1. The joint and marginal prior probability for γ
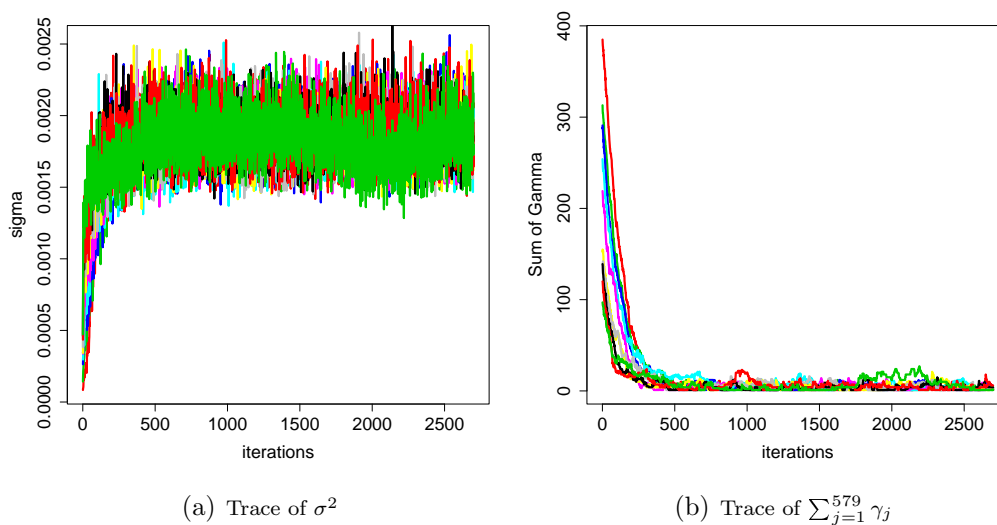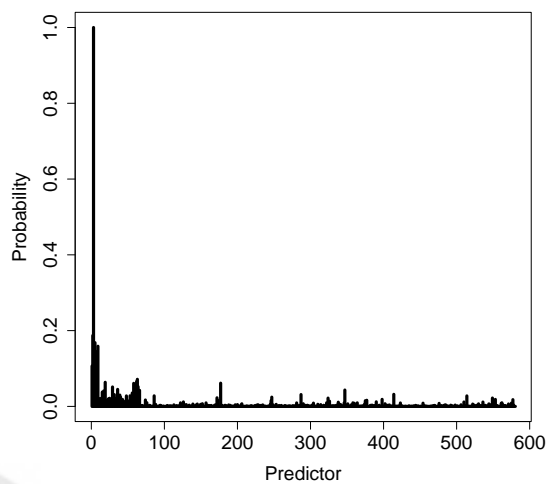
		γ_A	γ_B	γ_{AB}	Joint prior probability	
					Beta-Binomial	Conditional-Beta-Binomial
		0	0	0	$(1-w)^3$	$(1-w)^3$
		1	0	0	$w(1-w)^2$	$w(1-w)^2$
		0	1	0	$w(1-w)^2$	$w(1-w)^2$
		0	0	1	$w(1-w)^2$	0
		1	1	0	$w^2(1-w)$	$w^2(1-w)$
		1	0	1	$w^2(1-w)$	0
		0	1	1	$w^2(1-w)$	0
		1	1	1	w^3	w
Marginal	Beta-Binomial	w	w	w		
Prior	Conditional-Beta-binomial	$1-(1-w)^2$	$1-(1-w)^2$	w		
Probability						

w is the mean $a(a+b)^{-1}$ of the hyperprior Beta distribution

Table 2. ANOVA test of sensitivity
Est. is the true positive rate

Effect	Level	Mean (se)	p-value
Rho	0.1	0.773 (0.213)	
	0.3	0.776 (0.216)	
	0.5	0.771 (0.213)	
	0.7	0.756 (0.209)	
	0.9	0.716 (0.194)	< 0.001
Prior	bernoulli	0.759 (0.208)	
	binomial	0.758 (0.208)	0.793
Hyperprameter	(0.1, 0.12)	0.757 (0.208)	
	(0.2, 0.22)	0.76 (0.208)	0.173
Cutoff	2	0.327 (0.003)	
	4	0.602 (0.017)	
	6	0.782 (0.038)	
	8	0.852 (0.034)	
	10	0.892 (0.029)	
	12	0.918 (0.023)	
	14	0.935 (0.020)	< 0.001

Table 3. SNP look-up Table

Gene Name	Gene description	Locuslink ID	SNP
ALMS1	Alstrom syndrome 1	7840	cv11541174, cv167406, cv221633, cv8884674, cv9517435
AUP1	ancient ubiquitous protein 1	550	cv11727841
C2ORF6	chromosome 2 open reading frame 6	407654	cv1137486
DQX1	DEAQ box polypeptide 1 (RNA-dependent ATPase)	165545	cv2630842
MGC10955	hypothetical protein MGC10955	84762	cv1137462, cv1137467, cv1137468
MTHFD2	methylene tetrahydrofolate dehydrogenase (NAD+ dependent), methenyltetrahydrofolate cyclohydrolase	10797	cv339864, cv9517519
SLC4A5	solute carrier family 4, sodium bicarbonate cotransporter, member 5	57835	rs4853018, CV1137521, CV1137522, CV1137528, CV1137534,CV1137535, CV1137538, CV1137542, CV11727981, CV197439
PRSS25	protease, serine, 25	27429	cv2294084
SCN7A	sodium channel, voltage-gated, type VII, alpha	6332	cv2161217, cv356958
SLC4A10	solute carrier family 4, sodium bicarbonate transporter-like, member 10	57282	cv1300937, cv1300974
WBP1	WW domain binding protein 1	23559	cv2630862
ADD2	adducin 2 (beta)	119	CV11184, CV11464892, CV2634640, CV2634654, CV2634655, CV2805331, CV2805332, CV2805339, CV9987151
ADRA2B	adrenergic, alpha-2B-, receptor	151	hx2173, hx2767, hx4699
ATP6B1	ATPase, H+ transporting, lysosomal 56/58kDa, V1 subunit B, isoform 1 (Renal tubular acidosis with deafness)	525	CV2308451, CV2673636, CV2673637, CV2673638, CV2673639, rs1024764, rs1024765, rs2239484, rs2239487, rs2266917
SLC20A1	solute carrier family 20 (phosphate transporter), member 1	6574	CV11888001, CV248672, CV478858, CV9546434, CV9546587, rs1061254

Table 4. The partial importance list from stage I for hypertension data

Variable Name	Rank			Marginal Posterior Probability
	All	Main Effect	Interaction	
age	1	1	.	1
BMI	2	2	.	0.19
systolic BP	3	3	.	0.17
height	4	4	.	0.16
diastolic BP	5	5	.	0.15
triglyceride	6	6	.	0.14
cholesterol	7	7	.	0.11
weight	8	8	.	0.11
strong family history	9	9	.	0.07
SLC20A1_CV478858	10	10	.	0.07
SLC20A1_CV248672	11	11	.	0.07
MGC10955_cv1137467	12	12	.	0.06
BMI $\times$ SLC20A1_CV478858	13	.	1	0.06
NBC4_CV1137542	14	13	.	0.06
NBC4_CV11727981	15	14	.	0.06
SLC20A1_CV9546434	16	15	.	0.05
WBP1_cv2630862	17	16	.	0.05
ADD2_CV2805332	18	17	.	0.04
diastolic BP $\times$ SLC20A1_CV248672	19	.	2	0.04
SLC20A1_rs1061254	20	18	.	0.04
DQX1_cv2630842	21	19	.	0.04
AUP1_cv11727841	22	20	.	0.04
strong family history $\times$ AUP1_cv11727841	25	.	3	0.03
triglyceride $\times$ NBC4_CV11727981	28	.	4	0.03
cholesterol$\times$ WBP1_cv2630862	31	.	5	0.03

29

Table 5. The adjusted R-squares for selected models

Model		Regressors			Adj. Rsq.
		Covariates	SNPs	Interactions	
Two Stage Bayessian Variable selection	∗	Age, strong family history	AUP1_cv11727841, SLC20A1_CV9546434	strong family history × AUP1_cv11727841	0.3161
	∗∗	Age, strong family history	AUP1_cv11727841,	strong family history × AUP1_cv11727841	0.2898
	∗∗∗	Age, strong family history, cholesterol	AUP1_cv11727841, WBP1_cv2630862	strong family history × AUP1_cv11727841, cholesterol × WBP1_cv2630862	0.3022
Two stage stepwise selection		Age, strong family history	ALMS1_cv167406, MGC10955_cv1137467, ADD2_CV2805339, SLC20A1_CV9546434	Age × SLC20A1_CV9546434	0.3191

∗ The model with the highest joint posterior probability in Stage II.
∗∗ The model with the second highest joint posterior probability in Stage II.
∗ ∗ ∗ The model with the second highest joint posterior probability in Stage II.

Table 6. Parameter Estimates

Variable	Est. (se)	p-value
intercept	-5.32 (5.50)	0.3342
age	0.59 (0.05)	< .0001
strong family history	18.95 (9.96)	0.0580
AUP1_cv11727841		
AA	9.69 (3.84)	0.0120
AC	6.88 (3.95)	0.0823
CC	reference	-
SLC20A1_CV9546434		
AA	9.96 (2.84)	0.0005
AG	7.69 (2.93)	0.0091
GG	reference	-
strong family history × AUP1_cv11727841		
AA	-23.49 (10.04)	0.0199
AC	-16.72 (10.15)	0.1005
CC	reference	-

30

31