NIH Public Access
Author Manuscript
Comput Stat Data Anal. Author manuscript; available in PMC 2012 October 1.

Published in final edited form as:
Comput Stat Data Anal. 2011 October 1; 55(10): 2807–2818. doi:10.1016/j.csda.2011.04.019.

An Efficient Stochastic Search for Bayesian Variable Selection with High-Dimensional Correlated Predictors

Deukwoo Kwon[1,*], **Maria Teresa Landi**[1], **Marina Vannucci**[2], **Haleem J. Issaq**[3], **DaRue Prieto**[3], and **Ruth M. Pfeiffer**[1]

[1] Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland 20852, U.S.A

[2] Department of Statistics, Rice University, Houston, Texas 77251, U.S.A

[3] Lab of Proteomics and Analytical Technologies, SAIC-Frederick, Inc., Frederick, Maryland 21702, U.S.A

Abstract

We present a Bayesian variable selection method for the setting in which the number of independent variables or predictors in a particular dataset is much larger than the available sample size. While most existing methods allow some degree of correlations among predictors but do not consider these correlations for variable selection, our method accounts for correlations among the predictors in variable selection. Our correlation-based stochastic search (CBS) method, the hybrid-CBS algorithm, extends a popular search algorithm for high-dimensional data, the stochastic search variable selection (SSVS) method. Similar to SSVS, we search the space of all possible models using variable addition, deletion or swap moves. However, our moves through the model space are designed to accommodate correlations among the variables. We describe our approach for continuous, binary, ordinal, and count outcome data. The impact of choices of prior distributions and hyper-parameters is assessed in simulation studies. We also examined performance of variable selection and prediction as the correlation structure of the predictors varies. We found that the hybrid-CBS resulted in lower prediction errors and better identified the true outcome associated predictors than SSVS when predictors were moderately to highly correlated. We illustrate the method on data from a proteomic profiling study of melanoma, a skin cancer.

Keywords

Correlated predictors; correlation-based search; proteomic data

1 Introduction

In genomic experiments and other molecular studies one frequently encounters very high-dimensional data. Microarrays simultaneously monitor expression levels for several thousands of genes. Expression levels of genes that are co-regulated or in the same pathway are often correlated. Proteomic profiling studies in serum using mass spectrometry (MS)

[*]kwonde@mail.nih.gov.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

instruments measure size and charge of proteins, peptides and protein fragments and result in up to 15,000 intensity levels at pre-specified mass-to-charge ratio (*m*/*z*) values for each sample. Measurements on protein fragments and peptides arising from the same parent protein also tend to be highly correlated. Even after an initial pre-screening step to reduce dimensionality, investigators face large numbers of molecular measurements, often larger than the number of available samples, and many or most of these variables do not provide any information about the outcome measure. One key problem, termed *feature* or *variable selection*, in high-dimensional settings is thus to identify the optimal set among all the possible predictors.

In the Bayesian paradigm, variable selection can be formulated as a model selection problem. When the number of variables, *p*, is small compared with the available sample size, *n*, approaches based on the Bayes factor work well, since one can compute posterior model probabilities for all possible 2^p models (Hoeting et al., 1999; Ibrahim and Chen, 1999; Chen, Shao, and Ibrahim, 2000; Carlin and Chib, 1995). However, such computations are not feasible when *p* is very large. For the setting of large *p*, stochastic search variable selection (SSVS) methods that search over the model space have been suggested by George and McCulloch (1993 and 1997) and Brown et al. (1998a and 1998b). Related approaches for the large-*p* setting are Occam's window and Markov chain Monte Carlo model composition (MC^3) methods for Bayesian model averaging (Madigan and York, 1995; Hoeting et al., 1999), reversible jump Markov chain Monte Carlo methods (Green, 1995), and the shotgun stochastic search method and its extensions (e.g., Hans et al., 2007). Improved MCMC schemes have been proposed for a faster exploration of the posterior space, such as the evolutionary Monte Carlo schemes combined with parallel tempering of Bottolo and Richardson (2010).

Standard implementations of the approaches mentioned above however, do not account for correlations between the predictors. This can result in the inclusion of highly correlated variables into the model, at the cost of ignoring others that may improve the predictive performance of a model. We therefore propose a correlation-based search (CBS) algorithm, the hybrid-CBS algorithm, an extension of SSVS, to address the problem of variable selection with highly correlated predictors. Our algorithm extends SSVS by incorporating correlation information among the predictors in the search through the model space. The rest of the paper is organized as follows: In Section 2.1 we describe the Bayesian framework for the linear regression model and briefly review the SSVS method for the linear model in Section 2.2. We then show how to incorporate correlation information among the predictor variables in the model search and present the hybrid-CBS algorithm in Section 2.3. We use data augmentation and data transformations to adapt the algorithm for binary, ordinal, and count outcomes in Section 2.4. In Section 3 we assess the performance of the hybrid-CBS search on simulated data for various correlation settings, study the impact of the choices of prior distributions and hyperparameters and compare it to the performance of SSVS. We illustrate our approach on data from proteomic profiles of samples from a melanoma case-control study in Section 4. We close with a discussion in Section 5.

2 Methods

2.1 Bayesian Framework

For ease of exposition, we first present the Bayesian framework that is the basis of SSVS and the hybrid-CBS for a univariate linear model. Let $\mathbf{X} = (X_1, \ldots, X_p)'$ denote the predictor values and *Z* the continuous outcome variable. Without loss of generality we assume the columns of the matrix **X** and *Z* are centered. As *p* is large, we assume that only a small subset of the predictors $\mathbf{X}^* = (X_1^*, \ldots, X_{p^*}^*)'$, with $p^* \ll p$, relates to the outcome through

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$Z=\alpha+\mathbf{X}^{*'}\beta+\varepsilon,\ \ \varepsilon \sim N(0,\sigma^2). \tag{1}$$

To aid the identification of the relevant predictors, $\mathbf{X}^*$, we define a latent binary vector $\gamma = (\gamma_1, \ldots, \gamma_p)$ that characterizes a specific submodel,

$$Z=\alpha_\gamma+\mathbf{X}_\gamma'\beta_\gamma+\varepsilon, \tag{2}$$

by letting $\gamma_j = 1$ if the j-th predictor variable X_j is included in model (2) and zero otherwise. Our goal is to approximate model (1) by model (2), or equivalently $\mathbf{X}^*$ by a set of predictors $\mathbf{X}_{\hat{\gamma}}$, for which $\hat{\boldsymbol{\gamma}}$ has a large posterior probability.

A typical choice for a prior distribution on σ^2 is an Inverse Gamma distribution, $\mathcal{IG}(\nu/2, \nu\lambda/2)$, with shape and scale parameters $\nu/2$ and $\nu\lambda/2$, respectively. Given σ^2, the prior distribution for α is a normal distribution, $N(\alpha_0, h\sigma^2)$, with hyperparameters h and α_0. Given σ^2 and γ, a conjugate prior for $\boldsymbol{\beta}_\gamma$ is $N(\boldsymbol{\beta}_0, \sigma^2 H_\gamma)$. Common choices for H_γ are cI_{p_γ} and $c(X'X)^-$, independent prior and Zellner's g-prior, respectively. Various choices for the hyperparameters H_γ, h, ν, λ, α_0, and $\boldsymbol{\beta}_0$ are discussed in Section 3. A commonly adopted prior for γ assumes independent Bernoulli distributions,

$$p(\gamma)=w^{p_\gamma}(1-w)^{(p-p_\gamma)}, \tag{3}$$

where p_γ denotes the number of selected variables ($p_\gamma=\sum_{i=1}^{p}\gamma_i$) and w the ratio of the expected number of variables selected into the model to the total number of variables (George and McCulloch, 1997; Brown et al., 1998a, 1998b). In most biomarker discovery studies with a large number p of variables we expect a relatively small number of predictors to be associated with outcome. We thus let w be a small number, for example, $10/p$.

The choice of conjugate priors allows the calculation of the marginal posterior model probability, $p(\gamma \mid \mathbf{X}, Z)$, by integrating out the nuisance parameters α, β, and σ^2. This marginal posterior probability distribution is

$$p(\gamma|\mathbf{X}, \boldsymbol{Z}) \propto g(\gamma)= |\mathrm{I}_n+\mathbf{X}_\gamma\mathbf{H}_\gamma\mathbf{X}'_\gamma|^{-1/2}|\mathbf{Q}_\gamma|^{-(\nu+n)/2}p(\gamma), \tag{4}$$

where $\mathbf{Q}_\gamma=\nu\lambda+\boldsymbol{Z}'(\mathrm{I}_n-\mathbf{X}_\gamma\mathbf{K}_\gamma^{-1}\mathbf{X}'_\gamma)\boldsymbol{Z}$ and $\mathbf{K}_\gamma=\mathbf{X}'_\gamma\mathbf{X}_\gamma+\mathbf{H}_\gamma^{-1}$.

2.2 Stochastic Search Variable Selection (SSVS)

For small p, the best model, defined by the corresponding vector $\boldsymbol{\gamma}$ with the largest posterior probability $p(\boldsymbol{\gamma}|\mathbf{X}, \boldsymbol{Z})$, can be selected by computing all 2^p possible models. While George and McCulloch (1993) used Gibbs sampling which works well for moderate p, Brown et al. (1998b) sample $\boldsymbol{\gamma}$ from $g(\boldsymbol{\gamma})$ in (4) by a Metropolis algorithm and show that one can find good models by randomly exploring only a small fraction of the whole 2^p model space for large p. At any given iteration, their search method generates $\boldsymbol{\gamma}'$ from the current $\boldsymbol{\gamma}$ by either adding or deleting, with probability φ, a randomly chosen predictor from the model, or, with probability $1-\varphi$, swapping two predictors, by randomly and independently selecting a 0

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

and 1 in $\boldsymbol{\gamma}$ and exchanging them. The default choice of φ is 1/2. This leads to the proposal distribution

$$q(\gamma'|\gamma)=\begin{cases}\frac{\varphi}{p} & \text{if } |p_\gamma - p_{\gamma'}|=1\\ \frac{1-\varphi}{p_\gamma(p-p_\gamma)} & \text{if } |p_\gamma - p_{\gamma'}|=0.\end{cases}$$

Variable selection using $\boldsymbol{\gamma}$ can be based on two different criteria. One approach is to choose variables based on the vector $\boldsymbol{\gamma}$ with the highest posterior probability, $p(\boldsymbol{\gamma}|\mathbf{X}, \boldsymbol{Z})$, among all visited models. Alternatively, one can select the predictor X_i if the corresponding posterior probability of γ_i, $p(\gamma_i|\mathbf{X}, \mathbf{Z})$ exceeds a given threshold τ, for example $\tau = 0.5$ (Barbieri and Berger, 2004). We compare both approaches in the Simulation Section.

2.3 The Hybrid-Correlation-Based Search (hybrid-CBS)

Similarly to SSVS, our correlation-based search (CBS) method also searches the model space using three moves, 'addition', 'deletion', and 'swap'. However, while SSVS applies the moves to randomly chosen predictors, CBS uses correlation information among the variables to select predictors. This modification of SSVS is motivated as follows: if we wish to add a predictor to the current model, i.e. we choose the addition move, a predictor that has little correlation with variables already included in the model is preferable to one that is highly correlated with current model predictors. Similarly, when we choose the deletion move, the predictive performance of the model will not be impacted strongly if a variable that is highly correlated with another one in the model is removed. Thus in the CBS method the components of γ are no longer independent Bernoulli variables and we modify the prior for γ accordingly and use

$$p(\gamma) \propto \begin{pmatrix} p \\ p_\gamma \end{pmatrix}^{-1} \frac{1}{p_\gamma}, \tag{5}$$

where the components of γ are exchangeable but not independent (Chipman et al., 2001).

Next, we describe the implementation of the moves through the model space. Let Υ_X denote the correlation matrix of the predictors $\mathbf{X}$ with entries $(\Upsilon_X)_{ij} = \rho_{ij}$. Let $\mathcal{I}_\gamma = \{i: \gamma_i = 1, i = 1, \ldots, p\}$ denote the set of indices of the variables included in the current model characterized by the vector γ, and let $\mathcal{E}_\gamma = \{i: \gamma_i = 0, i = 1, \ldots, p\}$ denote the set of indices of variables not included in the current model. If the 'addition' move is selected we first randomly choose an index i' in $\mathcal{I}_\gamma$. We then find the index j' satisfying $\{j \in \mathcal{E}_\gamma: |\rho_{i',j'}| = \min |\rho_{i',j}|\}$ and add the corresponding predictor $x_{j'}$ to the model. Similarly, for the 'deletion' move, we first randomly choose an index i' in $\mathcal{I}_\gamma$. We then find the index j' satisfying $\{j \in \mathcal{I}_\gamma, j \neq i': |\rho_{i',j'}| = \max |\rho_{i',j}|\}$ and exclude the $x_{j'}$ from the next model. The swap move simply combines addition and deletion moves. The proposal distribution $q(\boldsymbol{\gamma}'|\boldsymbol{\gamma})$ for our search method is

$$q(\gamma'|\gamma)=\begin{cases}\frac{\varphi}{2p_\gamma} & \text{if } |p_\gamma - p_{\gamma'}|=1\\ \frac{1-\varphi}{p_\gamma} & \text{if } p_\gamma - p_{\gamma'}=0.\end{cases} \tag{6}$$

Since (6) is not symmetric we use a Metropolis-Hastings algorithm instead of a Metropolis algorithm that is used in SSVS. To ensure irreducibility of the resulting chain, a requirement for convergence, we combine CBS with SSVS into a hybrid-CBS, that is we randomly

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

choose a CBS move with probability 0.9 and an SSVS move with probability 0.1. While we want most of the moves to be based on CBS for computational efficiency, the choice of 0.9 is somewhat arbitrary. However, we found that results were not strongly impacted by the choice of the mixing proportion, which we further address in the Simulation Section. One technical difficulty in the setting of large p and small n is that the sample correlation matrix Υ_X can become unstable. To avoid this problem, we use a shrinkage estimator of the correlation matrix, $\Upsilon_\lambda = (1 - \lambda)\,\Upsilon_X + \lambda I$, based on an approach by Schäfer and Strimmer (2005) that allows one to compute the optimal shrinkage parameter λ in closed form. As a consequence of the shrinkage, Υ_λ, will be close to the identity matrix when p is much larger than n.

2.4 Extensions to Binary, Ordinal and Count Outcome Data

Our hybrid-CBS algorithm can be extended to other types of outcomes Y, including binary, ordinal responses, and count data. While some of these model formulations have been described previously for SSVS, we believe a concise summary will be helpful to practitioners.

Binary Outcomes—We now treat Z, linearly associated to the predictors $\mathbf{X}$ via model (1), as a latent variable and relate it to the binary outcome Y through $Y = I(Z \geq 0)$, leading to the probit model $P(Y = 1) = 1 - \Phi(\mathbf{X}'\boldsymbol{\beta})$ where Φ denotes the probability distribution for the standard normal distribution (Albert and Chib, 1993). For the identifiability of this model we set $\sigma^2 = 1$ in (1). Bayesian methods for variable selection in this framework have been proposed for example, by Sha et al. (2003) and Holmes and Held (2006). As Z is not observed, the appropriate posterior distribution after integrating out all other nuisance parameters is $p(\mathbf{Z}, \gamma\,|\mathbf{X}, \mathbf{Y})$. Based on a Metropolis-Hastings algorithm we iteratively first sample γ conditional on $\mathbf{Z}$ and then sample $\mathbf{Z}$ from the marginal posterior distribution $p(\mathbf{Z}|\gamma, \mathbf{X}, \boldsymbol{Y}) \sim \mathcal{TN}(\alpha_0 1_n + \boldsymbol{X}_\gamma'\beta_\gamma, \sigma^2\mathbf{P}_\gamma)$, where $\mathbf{P}_\gamma = \mathbf{I}_n + h\mathbf{1}\mathbf{1}' + \mathbf{X}_\gamma\mathbf{H}_\gamma\mathbf{X}'_\gamma$ and $\mathcal{TN}$ is a multivariate truncated normal distribution with truncation at zero.

Ordinal Outcomes—The binary probit model can be extended by to accommodate outcomes Y that take one of the ordered values $\{0, \ldots, J - 1\}$ (Kwon et al., 2007). The relationship between Y and the latent variable Z in (1) with $\sigma^2 = 1$, is given by $Y = jI(\delta_j < Z \leq \delta_{j+1})$, $j = 0, \ldots, J - 1$, leading to $P(Y \leq j) = \Phi(\delta_{j+1} - \mathbf{X}'\boldsymbol{\beta})$. The cutoff parameters δ_j are estimated under the constraint $-\infty = \delta_0 < \delta_1 < \delta_2 < \ldots < \delta_{J-1} < \delta_J = \infty$ and $\delta_1 = 0$. The marginal posterior distribution of δ_j for the Metropolis-Hastings algorithm is $\pi(\delta_j|\gamma, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{Y}, \delta_{(-j)})$

$\sim \mathcal{U}(\max(\max\{Z: Y = j - 1\}, \delta_{j-1}), \min(\min\{Z: Y = j\}, \delta_{j+1}))$, where $\delta_{(-j)}$ indicates the vector δ excluding the j-th component. The marginal posterior distribution of $\mathbf{Z}$ is now a truncated normal distribution that depends on δ.

Count Data—We transform the Poisson distributed outcome Y, where Y is the $n \times 1$ vector of counts, to obtain approximately normally distributed data that directly fit into the linear setting (1). Using a Taylor series expansion with two terms, we linearize $E[\boldsymbol{Y}^{1/2}|\boldsymbol{X}, \boldsymbol{\beta}]$ around the point $\log(\overline{Y}^{1/2})$ (Clyde and DeSimone-Sasinowska, 1997) to obtain $\boldsymbol{W}$, which has an approximately normal distribution

$$\boldsymbol{W} = 2(\overline{Y})^{-1/2}\left[\boldsymbol{Y}^{1/2} - \overline{Y}^{1/2}\left(1 - \frac{1}{2}\log\overline{Y}\right)\mathbf{1}_n\right] \sim N(\mathbf{X}'\beta, (1/\overline{Y})\boldsymbol{I}_n). \tag{7}$$

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

This transformation works well for relatively large counts as seen in our melanoma example (Clyde and DeSimone-Sasinowska, 1997). An alternative procedure appropriate for small counts using data augmentation was proposed by Frühwirth and Wagner (2005).

3 Simulation Study

We assessed the performance of the hybrid-CBS algorithm for continuous, binary, and count responses and compared it with the performance of the SSVS method. For each simulation, we used $n = 100$ observations and $p = 1000$ predictors $\mathbf{X}$ generated from a multivariate normal distribution with mean zero, variance one and correlation matrix Υ_X. The number of predictors truly associated with the outcome variable was $p^* = 10$. We assumed that the predictors were ordered so that the first ten, $X_1, \ldots, X_{10}$, were associated with the outcome, while $X_{11}, \ldots, X_p$ were not.

For the linear and the Poisson models we used 500, 000 iterations in the Metropolis-Hastings algorithm, but for the binary case we used 100, 000 iterations for the Metropolis-Hastings algorithm with 5, 000 burn-in iterations for computational efficiency. For all models we used hyperparameters $\alpha_0 = 0$, $\boldsymbol{\beta}_0 = \mathbf{0}$, $\nu = 3$, $\lambda = 1$ and $h = 10^6$ to induce vague priors on α and σ. We set $w = 10/p$ in equation (3) to induce parsimonious models.

3.1 Simulation Scenarios

We compare the hybrid-CBS with SSVS for several simulation scenarios, labeled S1 through S7. We consider two different values for the effect sizes for the outcome associated predictors: (1) either 0.8 or −0.8 and (2) either 1 or −1. The correlation matrix Υ_X of all scenarios has a block structure. The first block, denoted by $\Upsilon_{X,11}$ a 10×10 matrix, corresponds to the correlations of the outcome associated predictors, the second block, $\Upsilon_{X,12}$ a 990×10 matrix, contains the correlations between the outcome associated and the noisy predictors, and the third block, $\Upsilon_{X,22}$, is a 990×990 matrix of correlations between the 990 noisy predictors.

The entries of all three blocks are described below for the various scenarios and were chosen to capture different strengths of correlations among the groups of predictors.

For scenario (S1) for the linear model, the entries of all block matrices are constant. The correlations between the outcome associated predictors were very high; $\Upsilon_{X,11}$ had the entries $\rho_{ji} = \rho_{ij} = 0.8$, $i \neq j$. The entries of $\Upsilon_{X,22}$ were $\rho_{ji} = \rho_{ij} = 0.4$ for $i \neq j$, and $\Upsilon_{X,12}$ hat entries $\rho_{ij} = 0.25$, $i \neq j$. We let $\sigma = 1.5$, and $\beta_i = 0.8$ for $i = 1, \ldots, 5$, $\beta_i = -0.8$ for $i = 6, \ldots, 10$ and $\beta_i = 0$ for $i \geq 11$ in model (1). For scenario (S2), Υ_X was the same as in (S1), but with larger effect sizes, $\beta_i = 1$ for $i = 1, \ldots, 5$, , $\beta_i = -1$ for $i = 6, \ldots, 10$ and $\beta_i = 0$ for $i \geq 11$. For scenario (S3), $\beta_i = 1$ for $i = 1, \ldots, 5$, $\beta_i = -1$ for $i = 6, \ldots, 10$ and $\beta_i = 0$ for $i \geq 11$ and $\Upsilon_{X,11}$ had constant off diagonal entries $\rho = 0.8$, the correlations in $\Upsilon_{X,22}$ were all equal to $\rho = 0.4$ and the outcome associated and noisy predictors were uncorrelated, that is $\Upsilon_{X,12}$ had constant entries $\rho = 0$, $i \neq j$. For scenarios (S4)–(S7) we let $\sigma = 1.5$, and $\beta_i = 1.0$ for $i = 1, \ldots, 5$, $\beta_i = -1.0$ for $i = 6, \ldots, 10$ and $\beta_i = 0$ for $i \geq 11$ in model (1) and only varied Υ_X.

For scenario (S4), $\Upsilon_{X,11}$ had the entries $\rho_{ji} = \rho_{ij} = 0.8 - |i - j - 1|0.02$ for $i \neq j$, and the entries of $\Upsilon_{X,22}$ and $\Upsilon_{X,12}$ were constant with $\rho_{ji} = \rho_{ij} = 0.2$ and $\rho_{ij} = 0.1$, respectively. Scenario (S5) had the same Υ_X as (S4) but with uncorrelated outcome associated predictors, i.e. $\Upsilon_{X,11}$ was the identity matrix. For scenario (S6), $\Upsilon_{X,11}$ and $\Upsilon_{X,22}$ were the same as for scenario (S4), but the between group correlation was stronger, with entries $\rho_{ij} = 0.3$ for $\Upsilon_{X,12}$. For scenario (S7), Υ_X was the same as for (S6) but with $\Upsilon_{X,11}$ replaced by the identity matrix.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

For the binary model the latent linear variable Z in equation (1) was simulated from the three correlation structures as described above for the linear case. The values of nonzero regression coefficients were $\beta_i = 0.8$ for $i = 1, \ldots, 5$ and $\beta_i = -0.8$ for $i = 6, \ldots, 10$ in model (1) for simulation S1, corresponding to odds ratios 2.3 and 0.43, and $\beta_i = 1$ for $i = 1, \ldots, 5$, $\beta_i = -1$ for $i = 6, \ldots, 10$ for simulations (S2) and (S3), corresponding to odds ratios 3 and 0.33 respectively.

For count data, we generated $y_i \sim Poisson(\mu_i)$, with $\mu_i = A\lambda_i = A\exp(\boldsymbol{x}_i'\beta)$, $i = 1, \ldots, n$, with $A = 15$. We simulated predictors using the same three correlation structures as for (S1), (S2) and (S3) in the linear case. In simulation S1, $\beta_i = 0.2$ for $i = 1, \ldots, 5$, $\beta_i = -0.2$ for $i = 6, \ldots, 10$ and $\beta_i = 0$ for $i \geq 11$. In simulation (S2) and (S3) we used $\beta_i = 0.275$ for $i = 1, \ldots, 5$, $\beta_i = -0.275$ for $i = 6, \ldots, 10$ and $\beta_i = 0$ for $i \geq 11$.

3.2 Results

3.2.1 Variable Selection—First, we assessed the ability of the hybrid-CBS and SSVS to identify true outcome associated predictors for the different simulation scenarios (S1), (S2) and (S3) using the independent prior, $H_\gamma = cI_{p_\gamma}$ with $c = 1$. Figure 1 summaries results for SSVS (striped bars) and hybrid-CBS (solid bars) for continuous outcomes (row A), count outcomes (row B), and binary outcomes (row C). Column I of Figure 1 shows the marginal posterior probabilities of inclusion for the 10 true predictors averaged over 50 simulated datasets. Column II depicts the average number of predictors that were declared associated with outcome, based on their marginal probability of inclusion, $P(\gamma_i = 1|Data) > \tau$. For continuous and count data we let $\tau = 0.5$ and for binary outcomes $\tau = 0.2$. Columns III and IV of Figure 1 present the mean number of true positive (TP) and false positive (FP) predictors, respectively, corresponding to the selection made on the basis of $P(\gamma_i = 1|Data) > \tau$ with the same τ as in column II.

The hybrid-CBS had much larger marginal posterior probabilities of inclusion than SSVS for all simulation settings and outcomes (Figure 1) and selected six or more of the ten true predictors for all simulations for continuous and count outcomes and four or more for binary outcomes. It had relatively small number of FPs for continuous and count outcomes. For binary outcomes SSVS had large number of FPs for all three simulation settings. However, the number of TPs was much higher for hybrid-CBS than for SSVS, even for binary outcomes.

Figure 2 for the linear model illustrates the impact of various correlation structures on the performance of the algorithms. The hybrid-CBS had much larger marginal posterior probabilities of inclusion and more true positives than SSVS for scenarios (S4), (S6), and (S7), although hybrid-CBS gave more false positives. For scenario (S5), where all correlations were quite low, SSVS performed very well, it had a higher TP rate then the hybrid-CBS, but interestingly also the highest FP rate among all scenarios studied.

We also assessed the impact of the choice of covariance matrix, H_γ, in the prior distribution of β on the performance of the algorithms. We studied three choices: (1)$H_\gamma = cI_{p_\gamma}$, which results in an independent prior, (2) $H_\gamma = c(\boldsymbol{X}_\gamma'\boldsymbol{X}_\gamma)^-$, which yields the g-prior (Zellner, 1986), both common choices in the literature on Bayesian variable selection and (3) a shrinkage version of the g-prior, where $X'X$ was replaced by $(n-1)\,\hat{\Upsilon}_{\hat{\lambda}}$ based on Schäfer and Strimmer (2005). The constant c plays an important role. The larger the value of c is, the fewer variables are selected due to regularization. We let $c = 1$, 3, and 5 for the independent prior. In order to get a similar amount of regularization, we used $c = 75$, 225, and 370 for the g-prior and $c = 3$, 8, and 14 for the shrinkage g-prior.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Figure 3 summarizes our findings for scenario (S2) for continuous outcomes. The rows again correspond to linear, count and binary outcomes and the columns to the probability of inclusion for the 10 true predictors (column I), the average number of predictors that were declared associated with outcome (column II), and the mean number of TPs (column IV) and FPs predictors (column IV). The independent prior with $c = 1$ yielded the best results in terms of the number of TPs and number of true predictors selected, while still resulting in a low number of false positive selections. Again, hybrid-CBS showed better performance than SSVS for all settings in Figure 2. For all three choices of priors for $\boldsymbol{\beta}$ the number of predictors selected decreased as the value of c increased, due to the regularization.

3.2.2 Sensitivity Analysis for Variable Selection

Variable Selection Criterion: In the previous section we selected variables based on the marginal posterior inclusion probabilities, $p(\gamma_i = 1|X, Z) > \tau$. Alternatively one could use the highest joint posterior model probability, $\max\{p(\gamma|X, Z)\}$ to select predictors associated with outcome. We compared both approaches with respect to the number of predictors declared important and the numbers of TPs and FPs. Both approaches gave very similar results (Table 1) and selected on average the same predictors.

Choice of Mixing Proportion: For the hybrid-CBS we propose to use a mixing proportion of 90% for the CBS moves. However, this choice is somewhat arbitrary. We therefore assessed the sensitivity of the algorithm to other choices of proportion, namely 95%, 80% and 50% and found that they yielded very similar result (Table 2). The average of the posterior inclusion probabilities for the true predictors was around 0.75 for all four choices of mixing proportion (second column in Table 2). The average number of selected predictors using the criterion $p(\gamma_i = 1|X, Z) > 0.5$ was around 8.5 (third column); the average number of TPs was 7.3 (fourth column); and average number of FPs was 1.2 (fifth column).

Choice of Hyperparameter w: We added to assess the impact of the magnitude of w used in the prior distribution for γ given in (3) on both SSVS and hybrid-CBS (Table 3). Not surprisingly, SSVS was very sensitive to the choice of w, while hybrid-CBS was not, as w only affects the SSVS component of the algorithm. For continuous outcomes in the S2 setting, selecting predictors based on the marginal inclusion probabilities SSVS selected around 2 predictors with 1.5 TPs, and 0.7 FPs for $w = 5/p$ and $w = 10/p$. However, for $w = 20/p$, SSVS selected 10.5 predictors, with 0.5 TPs and 10 FPs. The hybrid-CBS selected approximately seven predictors, with six TPs, and 1.3 FPs for all choices of w, $w = 5/p$, $10/p$, and $w = 20/p$.

3.2.3 Prediction—To study the prediction error of the algorithms, we applied the model selected by the hybrid-CBS or SSVS algorithm based on a training set with $n = 100$ observations and $p = 1000$ predictors to 50 independent test datasets with $N = 1000$ observations each, generated under the same scenario as the training set.

Figure 4 shows the mean squared prediction error (MSPE) for the test set for SSVS and hybrid-CBS with three different prior settings for the linear model for scenario (S1). Letting Z denote the observed continuous outcome and $\hat{Z}$ the corresponding fitted value based on the model selected in the training set, $\text{MSPE} = 1/N \sum_{i=1}^{N} \left(Z_i - \widehat{Z}_i\right)^2$. To reduce the impact of the hyperparameter c in the prediction, we set $c = 1$, 75, and 3 for the independent prior, g-prior, and shrinkage g-prior, respectively. The hybrid-CBS also had a lower MSPE than SSVS for all choices of prior distributions for $\boldsymbol{\beta}$, with the independent prior resulting in the smallest MSPE overall.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

For comparison purposes we also computed the MSPE using Bayesian Model Averaging (BMA) based on the 10 best models. This approach also resulted in lower MSPE for the hybrid-CBS than for SSVS (Table 4).

For count outcomes, the hybrid-CBS also had a smaller MSPE than SSVS with $c = 1$ and the independent prior (see Figure 5, a). For binary outcomes we computed the Brier score, $\left(\sum_{i=1}^{N}\left(Y_i - \Phi(X'_{\hat{\gamma}}\widehat{\beta}_{\hat{\gamma}})\right)^2\right)/N$, where Y denotes the binary outcome in the test dataset, Φ stands for the standard normal distribution function, $X_{\hat{\gamma}}$ is the predictor matrix of a test dataset with columns based on the predictors identified in the training set, and $\beta_{\hat{\gamma}}$ are the least square estimates of regression coefficients corresponding to $X_{\hat{\gamma}}$. We also computed the misclassification rate, $\left(\sum_{i=1}^{N}\mathbf{I}_{(Y_i \neq \widehat{Y}_i)}\right)/N$ where $\mathbf{I}$ denotes an indicator function. The Brier scores are shown in Figure 5(b). With hyperparameter $c = 1$ and the independent prior, the hybrid-CBS method had a lower Brier score on average than SSVS. The average misclassification rate of the hybrid-CBS was 35%.

Figure 6 shows the MSE scenarios (S4)–(S7) with $c = 1$ and independent prior for $\boldsymbol{\beta}$. Except for scenario (S5), that was based on very low correlations, hybrid-CBS resulted in a substantially lower MSPEs than SSVS.

4 Data example

We illustrate our method on data from a proteomic profiling study on melanoma skin cancer conducted in Northeastern Italy (Landi et al., 2005). The study aim was to identify predictors for 1) tumor aggressiveness, measured by melanoma thickness (continuous outcome); 2) the number of nevi (or moles) (count outcome), a risk factor for melanoma; and 3) case-control status (binary outcome). Mass spectrometry (MS) measurements were obtained from 173 individuals diagnosed with melanoma (cases, $Y = 1$), and 178 healthy individuals (controls, $Y = 0$) with the Protein Biology System 2 (PBS II) SELDI-TOF mass spectrometer (Bio-Rad Laboratories, Hercules, CA). The resulting data are mass spectra, that are patterns representing the distribution of ions by mass-to-charge ratio (m/z).

Before any statistical analyses, we preprocessed the 351 mass spectra with the following steps: denoising, baseline subtraction, normalization, peak detection, and peak alignment. The spectra were denoised using an algorithm by Kwon et al. (2008). For baseline subtraction and peak detection we used the *PROcess* package in *R*. For all analyses, the predictors were the intensities at the *m*/*z* values corresponding to 113 peaks that were present in at least ten percent of all spectra. The 113 peaks were highly correlated; 65.8% of the pairwise empirical correlations were larger than 0.5, and 22.2% larger than 0.75.

We applied the hybrid-CBS and SSVS methods with the independent prior for the regression parameters $\boldsymbol{\beta}$ with $c = 1$ for all three responses. This value of c ensured the identification of a sufficient number of predictors. The threshold for the marginal posterior probability of inclusion was 0.2. We used the shrinkage version of correlation matrix of the predictors for the hybrid-CBS moves.

First, we aimed to identify predictors associated with melanoma thickness. This analysis was restricted to the 145 melanoma cases on whom melanoma thickness measurements were available. After a log-transformation, the thickness measurements were normally distributed. Based on a linear model (1), six *m*/*z* values (5,565.405, 5,828.874, 11,186.37, 11,754.4, 17,114.41 and 18,904.11) had marginal posterior probabilities of inclusion 0.99, 0.99, 0.99, 0.38, 0.34, and 0.26, respectively. While three of the six *m*/*z* values (5,565.405, 5,828.87 and

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

11,186.37) were also identified by SVSS with marginal posterior probabilities of inclusion 0.27, 0.60, and 0.21, respectively, the inclusion probabilities were much lower.

Second we analyzed case-control data, based on all 173 cases and 175 controls. The hybrid-CBS method identified ten (m/z) values (2,758.60, 5,955.27, 6,079.15, 6,876.81, 11,186.37, 11,590.23, 11,954.82,17,114.41, 17,843.58, and 33,488.04) with marginal posterior probabilities of inclusion 0.96, 0.99, 0.99, 0.99, 0.96, 1.00, 0.98, 0.99, 0.99 and 0.99, respectively. SSVS identified only one peak (11,590.23) with a marginal posterior probability of inclusion equal to one.

Finally, as having a large number of nevi is a strong risk factor for melanoma, we identified protein measurements to predict the number of nevi, based on a Poisson model in the controls. The hybrid-CBS identified six peaks, corresponding to m/z values (6,950.327, 9,746.156, 13,135.19, 12,635.65, 17,843.58, and 18,336.08) with marginal posterior probabilities of inclusion 0.67, 0.71, 0.69, 0.31, 0.98 and 0.76, respectively. The two peaks (17,843.58 and 18,336.08) were also chosen by SSVS, with the somewhat smaller posterior probabilities 0.21 and 0.25, respectively.

To evaluate the predictive performance of the selected models we used leave-one-out cross validation prediction appropriate for the Bayesian setting (Sha et al., 2004). We selected predictors of which marginal posterior probabilities of inclusion, $P(\gamma_i = 1|X, Z)$, are greater than 0.5. The MSPEs for melanoma thickness were 0.7433 (SSVS) and 0.7368 (hybrid-CBS); and for nevi count the MSPEs were 1.4357 (SSVS) and 1.39 (hybrid-CBS). For case-control status, SSVS had a Brier score of 0.23 and a 39% misclassification rate; and hybrid-CBS had a Brier score of 0.22 and a 37% misclassification rate. Thus, the predictive performance of models selected by hybrid-CBS was better than those for SSVS.

5 Discussion

In this paper we proposed a correlation-based search algorithm, the hybrid-CBS, that extends a popular Bayesian search algorithm for high-dimensional data, the stochastic search variable selection (SSVS) method, to accommodate the setting of correlated high-dimensional predictors. Similar to SSVS, we search the model space using variable addition, deletion, or swap moves. However, our moves are driven by the correlations seen in the data. To ensure irreducibility in the Markov chain that is the basis for our search method we combine the purely correlation-based search with SSVS into a hybrid algorithm. We present details on the implementation of the hybrid-CBS for binary, ordinal and count data. Modifications of SSVS for survival outcome and multiple categorical outcomes are described in Sha et al. (2004 and 2006) and can also easily be extended to the hybrid-CBS algorithm.

We assessed performance of our new algorithm compared to SSVS on simulated data. The hybrid-CBS performed better than SSVS in terms of selecting true outcome associated predictors, and had lower prediction errors, when predictors were highly correlated for continuous, binary, and count response data. In the simulation study we also investigated the sensitivity of both methods, SSVS and hybrid-CBS, to the choice of some of the model parameters, in particular the choice of prior for the regression parameters $\boldsymbol{\beta}$ that relate the predictors to the outcome, either directly or through a latent variable. We studied the g-prior and shrinkage g-prior as well as the independent prior. Both the g-prior and independent prior are related to shrinkage; the former is equivalent to proportional shrinkage and the independent prior corresponds to absolute shrinkage. When the predictors have very different scales, the g-prior is recommended because of its automatic scaling feature (Chipman et al., 2001; Bottolo and Richardson, 2010). Absolute shrinkage is closely related

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

to ridge regression (Hoerl and Kennard, 1970). While the *g*-prior is widely used and has been studied extensively in Bayesian variable selection for low-dimensional settings (Cui and George, 2008; Liang et al., 2008), its performance in high-dimensional situations is not well understood. We found the *g*-prior did not lead to adequate performance in variable selection in most cases. As can be seen from Figure 2, substantial correlations between the true predictor group and noisy variables resulted in poor performance of the *g*-prior. When we used the shrinkage version of *g*-prior with small *c*, the performance of both SSVS and hybrid-CBS improved slightly. Of course, for high dimensions the shrinkage estimate of the correlation matrix will be close to the identity matrix. However, the performance of the hybrid-CBS with the shrinkage *g*-prior was still inferior compared with that of the hybrid-CBS with the independent prior. We therefore recommend using the hybrid-CBS with the independent prior due to the computational ease.

The hybrid-CBS does not substantially increase the computational burden. For the continuous outcomes the computation times for the hybrid-CBS were maybe ~ 10%–15% higher than those for SSVS. With 3.0Ghz quadcore CPU, it took around 101 second for SSVS and 114 seconds for hybrid-CBS on average for 50 simulated datasets with 200,000 iterations. Finally, the proposed correlation driven search can be easily adapted to other variable selection methods based on stochastic searches, for example the method proposed by Holmes and Held (2006) for binary and multinomial outcomes, and the method of Casella and Moreno (2006) for continuous outcomes.

Acknowledgments

We thank the Editor, Associate Editor and two referees for constructive suggestions that led to a significant improvement of the paper. Kwon, Landi, and Pfeiffer are supported by the Intramural Research Program of the NIH National Cancer Intitute. Vannucci is partially supported by NIH grant R01-HG0033190-05 and NSF grant DMS1007871.

References

Albert J, Chib S. Bayesian analysis of binary and polychotomous response data. J Amer Statist Assoc. 1993; 88:669–679.

Barbieri M, Berger J. Optimal predictive model selection. Ann Statist. 2004; 32:870–897.

Bottolo L, Richardson S. Evolutionary stochastic search for Bayesian model exploration. Bayesian Analysis. 2010; 5:583–618.

Brown P, Vannucci M, Fearn T. Multivariate Bayesian variable selection and prediction. J Roy Statist Soc Ser B. 1998a; 60:627–641.

Brown P, Vannucci M, Fearn T. Bayesian wavelength selection in multicomponent analysis. J Chemometrics. 1998b; 12:173–182.

Carlin B, Chib S. Bayesian model choice via Markov chain Monte Carlo methods. J Roy Statist Soc Ser B. 1995; 57:473–484.

Casella G, Moreno E. Objective Bayesian variable selection. J Amer Statist Assoc. 2006; 101:157–167.

Chen, M.; Shao, Q.; Ibrahim, J. Monte Carlo methods in Bayesian computation. New York: Springer-Verlag; 2000.

Chipman, H.; George, E.; McCulloch, R. The practical implementation of Bayesian model selection. In: Lahiri, P., editor. IMS Lecture Notes - Monograph Series. New York: Cambridge University Press; 2001. p. 65-116.

Clyde, M.; DeSimone-Sasinowska, H. Technical report 97-06. Birmingham, Alabama: Duke University; 1997. Accounting for model uncertainty in Poisson regression models: Particulate matter and mortality.

Cui W, George E. Empirical Bayes vs. fully Bayes variable selection. J Statist Plann Inference. 2008; 138:888–900.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Frühwirth-Schnatter S, Wagner H. Data augmentation and Gibbs sampling for regression models for small counts. Student. 2005; 5:207–220.

George E, McCulloch R. Variable selection via Gibbs sampling. J Amer Statist Assoc. 1993; 88:881–889.

George E, McCulloch R. Approaches for Bayesian variable selection. Statist Sinica. 1997; 7:339–373.

Green P. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika. 1995; 82:711–732.

Hans C, Dobra A, West M. Shotgun stochastic search for "large *p*" regression. J Amer Statist Assoc. 2007; 102:507–516.

Hoerl A, Kennard R. Ridge Regression: Applications to Nonorthogonal Problems. Tecnometrics. 1970; 12:69–82.

Hoeting J, Madigan D, Raftery A, Volinsky C. Bayesian model averaging: a tutorial. Statist Sci. 1999; 14:382–417.

Holmes C, Held L. Bayesian auxiliary variable models for binary and multinomial regression. Bayesian Analysis. 2006; 1:145–168.

Ibrahim, J.; Chen, M. Bayesian methods for variable selection in the Cox model Generalized linear models: A Bayesian perspective. Dey, D.; Ghosh, D.; Mallick, B., editors. New York: Marcel Dekker; 1999. p. 287-311.

Kwon D, Tadesse M, Sha N, Pfeiffer R, Vannucci M. Identifying biomarkers from mass spectrometry data with ordinal outcomes. Cancer Informatics. 2007; 3:19–28. [PubMed: 19455232]

Kwon D, Vannucci M, Song J, Jeong J, Pfeiffer R. A novel wavelet-based thresholding method for preprocessing mass spectrometry data that account for heterogeneous noise. Proteomics. 2008; 8:3019–3029. [PubMed: 18615428]

Landi M, Kanetsky P, Tsang S, Gold B, et al. MC1R, ASIP, and DNA repair in sporadic and familial melanoma in a Mediterranean population. J National Cancer Institute. 2005; 98:998–1007.

Liang F, Paulo R, Molina G, Clyde M, Berger J. Mixture of *g*-priors for Bayes variable selection. J Amer Statist Assoc. 2008; 103:410–423.

Madigan D, York J. Bayesian graphical models for discrete data. International Statist Rev. 1995; 63:215–232.

Schäfer J, Strimmer K. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. Statist Appl Genomics and Molecular Biology. 2005; 4:1–32.

Sha N, Tadesse M, Vannucci M. Bayesian variable selection for the analysis of microarray data with censored outcome. Bioinformatics. 2006; 22:2262–2268. [PubMed: 16845144]

Sha N, Vannucci M, Brown P, Trower M, Amphlett G. Gene selection in arthritis classification with large-scale microarray expression profiles. Comparative and Functional Genomics. 2003; 4:171–181. [PubMed: 18629129]

Sha N, Vannucci M, Tadesse M, Brown P, et al. Bayesian variable selection in multinomial probit models to identify molecular signatures of disease stage. Biometrics. 2004; 60:812–819. [PubMed: 15339306]

Zellner, A. On assessing prior distributions and Bayesian regression analysis with *g*-prior distributions. In: Goel, P.; Zellner, A., editors. Studies in Bayesian Econoetrics and Statistics. New York: Elsevier; 1986. p. 233-243.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

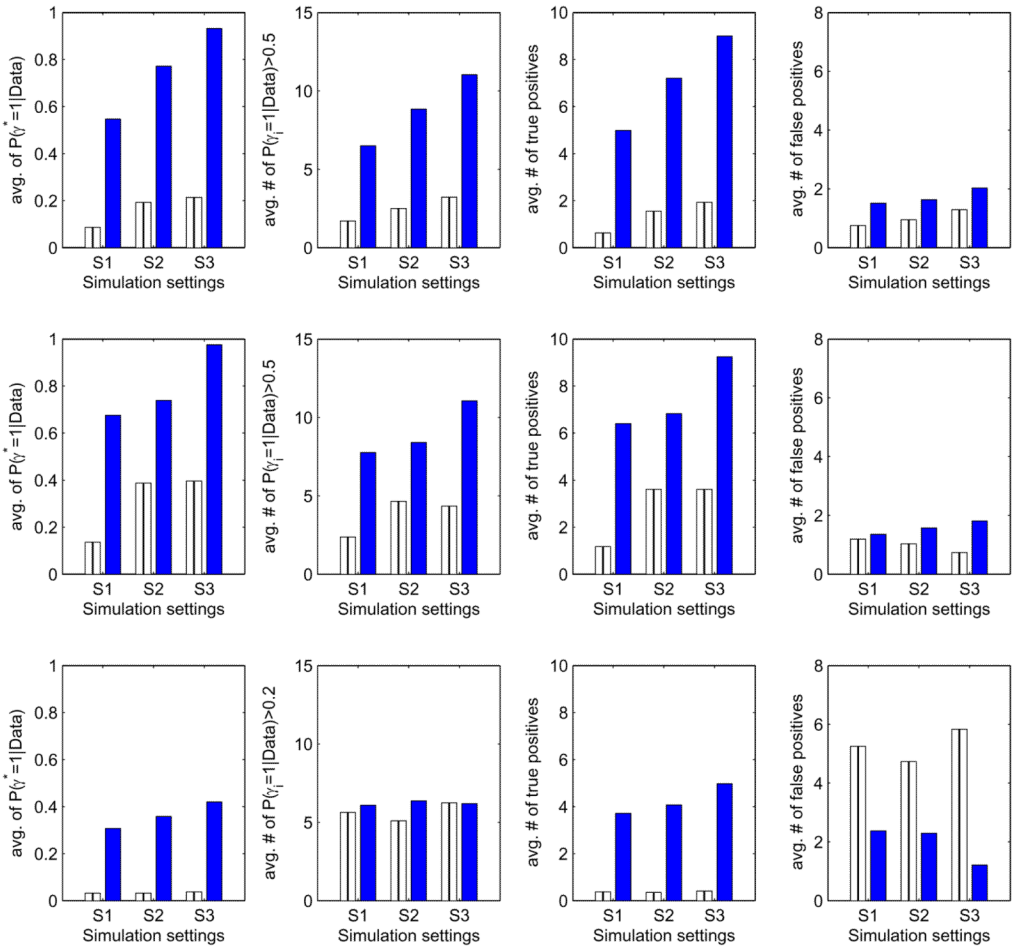


Figure 1.

Simulation Result: Plots for continuous (row A), count (row B), and binary (row C) outcomes. Striped bars represent SSVS and solid bars represent hybrid-CBS. S1, S2, and S3 denote Simulation 1, Simulation 2, and Simulation 3, respectively.

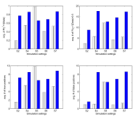


Figure 2.
Simulation Result: Plots for continuous outcomes. S2, S4, S5, S6, and S7 denote Simulation 2, Simulation 4, Simulation 5, Simulation 6 and Simulation 7, respectively.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

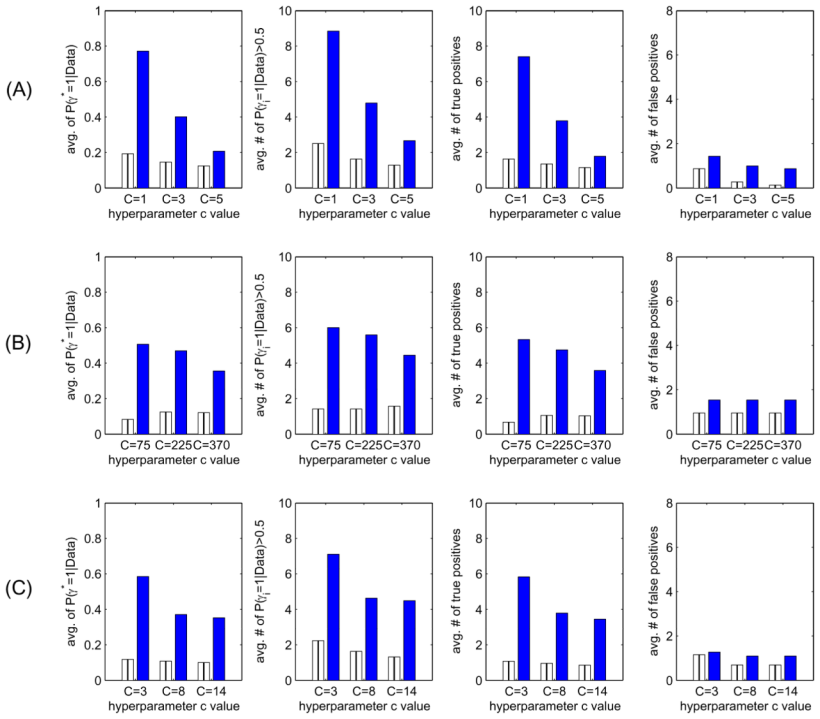


Figure 3.

Comparison of hyperparameter *c* and priors of $\boldsymbol{\beta}$: Plots for an independent prior (row A), *g*-prior (row B), and shrinkage *g*-pior (row C). Striped bars represent SSVS and solid bars represent hybrid-CBS.

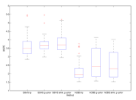
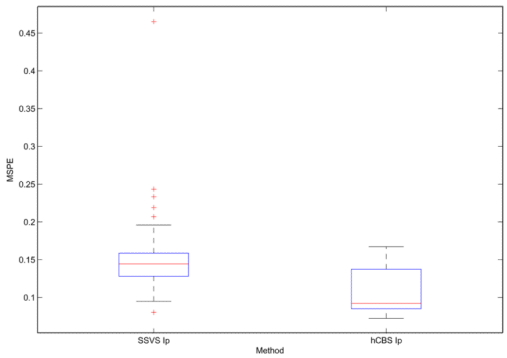


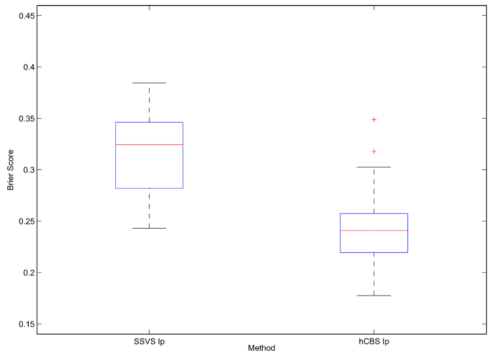
Figure 4.
Comparison of Mean Squared Prediction Errors for SSVS and hybrid-CBS: Plots for continuous outcome. hybrid-CBS indicates hybrid-CBS. Ip and shrk. g-prior denote an independent prior and shrinkage *g*-pior, respectively.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript



(a)



(b)

Figure 5.
(a) Comparison of Mean Squared Prediction Errors for SSVS and hybrid-CBS: Plots for count outcome; (b) Comparison of Brier Scores for SSVS and hybrid-CBS: Plots for binary outcome.

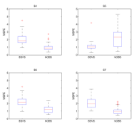


Figure 6.
Comparison of Mean Squared Prediction Errors for SSVS and hybrid-CBS: Plots for continuous outcome. hybrid-CBS indicates hybrid-CBS. Ip denotes an independent prior.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 1

Comparison of average number of selected predictors, number of TPs, and number of FPs based on highest joint posterior model probability and marginal inclusion probabilities given in parentheses.

Scenario	Avg. No. of selected predictors		No. of TPs		No. of FPs	
	SSVS	hybrid-CBS	SSVS	hybrid-CBS	SSVS	hybrid-CBS
S1	2.44(1.72)	6.72(6.52)	0.96(0.64)	5.42(5)	1.48(0.76)	1.3(1.52)
S2	3.12(2.52)	9.46(8.86)	1.92(1.56)	7.7(7.22)	1.2(0.96)	1.76(1.64)
S3	3.52(3.64)	11.86(11.06)	2.1(1.94)	9.4(9.02)	1.42(1.3)	2.46(2.04)

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 2

Sensitivity analysis for different proportions of CBS moves.

% of CBS	Avg. of $p(\gamma_i^*=1\|\mathbf{X},\mathbf{Z})$	Avg. No. of predictors with $p(\gamma_i = 1\|X, Z) > 0.5$	No. of TPs	No. of FPs
90%	**0.77**	**8.86**	**7.22**	**1.64**
95%	0.75	8.44	7.34	1.10
80%	0.73	8.3	7.24	1.06
50%	0.76	8.7	7.54	1.16

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 3

Sensitivity analysis for different w values.

w	Avg. of $p(\gamma_i^*=1\|\mathbf{X},\mathbf{Z})$		Avg. No. of $p(\gamma_i = 1\|X, Z) > 0.5$		No. of TPs		No. of FPs	
	SSVS	hybrid-CBS	SSVS	hybrid-CBS	SSVS	hybrid-CBS	SSVS	hybrid-CBS
5/p	0.13	0.59	1.54	6.98	1.2	5.8	0.34	1.18
10/p	0.19	0.78	2.8	8.7	1.8	7	1	1.7
20/p	0.27	0.64	10.64	7.26	0.54	6.24	9.92	1.02

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 4

Mean squared prediction errors (MSPEs) for SSVS and hybrid-CBS (standard deviations (SDs) in parenthesis).

Method	Independent prior		*g*-prior		Shrinked *g*-prior	
	SSVS	hybrid-CBS	SSVS	hybrid-CBS	SSVS	hybrid-CBS
Incl. prob	3.75(1.05)	2.4(1.57)	3.73(0.43)	2.60(0.83)	3.81(0.5)	2.57(0.78)
BMA	2.12(0.48)	1.76(0.63)	3.12(0.48)	2.91(0.49)	3.28(0.48)	2.56(0.66)