Unconstrained Bayesian Model Selection on Inverse Correlation Matrices With Application to Sparse Networks

Nitai D. Mukhopadhyay
Department of Biostatistics
Virginia Commonwealth University Richmond VA 23298

Sarat C. Dass
Department of Statistics & Probability
Michigan State University East Lansing MI

November 11, 2010

Abstract

Bayesian statistical inference for an inverse correlation matrix is challenging due to non-linear constraints placed on the matrix elements. The aim of this paper is to present a new parametrization for the inverse correlation matrix, in terms of the Cholesky decomposition, that is able to model these constraints explicitly. As a result, the associated computational schemes for inference based on Markov Chain Monte Carlo sampling are greatly simplified and expedited. The Cholesky decomposition is also utilized in the development of a class of hierarchical correlation selection priors that allow for varying levels of network sparsity. An explicit expression is obtained for the volume of the elicited priors. The Bayesian model selection methodology is developed using a Reversible Jump algorithm and is applied to a dataset consisting of gene expressions to infer network associations.

1 Introduction and Motivation

A majority of the work related to Bayesian inference on graphical models have assumed the multivariate Gaussian as the preferred joint distribution on nodes. This assumption, though mathematically tractable, severely limits the applicability of such models since the marginal distribution at each node is forced to be normal. Recent work thus has focused on joint distributions elicited in terms of the Gaussian copula (see, for example, Pitt et al. (2006)) which uses the inverse of a correlation matrix to model network associations. While gaining modeling flexibility, several constraints are placed on the inference methodology due to the use of a correlation, instead of a covariance, matrix. The entries of a inverse correlation matrix are constrained in a manner so that the diagonal elements of its inverse (a correlation matrix necessarily) are equal to unity. Further, the correlation matrix (and its inverse) are also required to be positive definite. Previous work has accounted for these restrictions by updating entries of the correlation matrix one by one, each time calculating an interval of admissible values so that the resulting correlation matrix is positive definite (see, for example, Pitt et al. (2006), Wong et al. (2003) and Barnard et al. (2000)).

Our objective was to perform unconstrained model selection on the space of sparse graphical networks models. A typical scenario we encountered is inferring network associations for a dataset of gene expressions consisting of n samples of each of p genes, with $p >> n$. Typically, again, in these cases, most network associations are negligible with only a few significant ones, thus giving rise to sparsity in the entries of the inverse correlation matrix. The Bayesian analysis of graphical models entails sampling from the class of all inverse correlation matrices, and in high dimensional problems, this estimation procedure should exploit the sparsity of the network associations. The term by term updating scheme of Pitt et al. (2006), Wong et al. (2003) and Barnard et al. (2000), for example, can be slow to converge.

This paper provides an alternative parametrization of the inverse correlation matrix which is able to free up the constrains placed on its elements. We are able to do this by exploiting several useful properties of the Cholesky decomposition of the inverse correlation matrix, W. Our approach explores network characteristics via L, the lower triangular matrix of the Cholesky decomposition. Although W and L have a one-to-one correspondence, zero entries of W do not in general correspond to zero entries of L. Thus, the nature of network sparsity and the space of all models we consider are the ones characterized by the zero and non-zero elements of L. Nevertheless, this is a fairly general representation of sparsity;

for example, when the inverse correlation matrix W is banded, it follows that L is banded as well. Generally speaking, L may contain a larger number of non-zero terms compared to the corresponding W but not significantly larger.

The Cholesky decomposition additionally allows us to develop a class of prior distributions on the space of all inverse correlation matrices that models sparsity and gives an explicit formula for the volume; note that Wong et al. (2003) had to assume a block diagonal structure on the inverse correlation matrices to obtain an explicit volume formula. The Bayesian inferential procedure utilizes a Reversible Jump Markov Chain Monte Carlo (RJMCMC) algorithm to jump between model spaces of varying dimensions. Due to freeing up of constraints, the RJMCMC scheme has significantly lower computation time and encourages faster mixing.

The rest of the paper is organized as follows. Section 2 discusses the multivariate Gaussian distribution with the inverse correlation matrix as the parameter of interest. The new parametrization of the inverse correlation matrix in terms of its Cholesky decomposition is presented here. Section 3 develops a class of prior distributions on the space of all inverse correlation matrices with an explicit formula for the volume derived. Section 4 develops the RJMCMC algorithm for Bayesian inference. Experiments with simulated and real data are presented in Section 5 together with results on the sensitivity of the analysis on model specifications.

2 The Distributional Model

Let $\mathbf{X} = (X_1, X_2, \cdots, X_p)^T$ denote a p-variate random vector taking values in $\mathbb{R}^p$ distributed as multivariate Gaussian with $E(X_i) = 0$ and $V(X_i) = 1$ for all $i = 1, 2, \cdots, p$. The joint density of $\mathbf{X}$ is

$$\phi_p(x_1, x_2, \cdots, x_p \,|\, R) = \frac{1}{(2\pi)^{p/2}(\det(R))^{1/2}} \exp\left\{ -\frac{1}{2}\, \mathbf{x}^T\, R^{-1}\, \mathbf{x} \right\} \qquad (1)$$

where $\mathbf{x} = (x_1, x_2, \cdots, x_p)^T \in \mathbb{R}^p$; in (1), R denotes a symmetric positive definite matrix and $\det(R)$ denotes the determinant of R. The matrix R is the correlation matrix; the (i, j)-th entry of R, r_{ij}, represents the correlation between X_i and X_j for $i \neq j$. We denote by $\mathcal{C}_p$ to be the space of all $p \times p$ correlation matrices.

Entries of the inverse correlation matrix, $W \equiv R^{-1} \equiv ((w_{ij}))$ reflects the extent of conditional dependence between a pair of component of $\mathbf{X}$, that is,

$$w_{ij} = 0 \Leftrightarrow X_i \text{ and } X_j \text{ are conditionally independent given the rest of the } X_k\text{'s, } k \neq \{i, j\}. \qquad (2)$$

Although W has altogether $p(p+1)/2$ distinct elements, not all of these entries are free to vary. The p diagonal elements w_{ii}, $i = 1, 2, \cdots, p$, and the $p(p-1)/2$ off-diagonal elements w_{ij}, $i > j$ satisfy p non-linear constraints: If $W = ((w_{ij}))$ is inverted to get W^{-1}, then, the diagonal entries of this inverse should be unity. Until now, inference on W has been difficult due to the presence of these non-linear constraints; see, for example, Pitt et al. (2006) and Wong et al. (2003) which report the difficulties involved and possible solutions when dealing with these constraints. The goal of this paper is to provide a new solution to the problem of inference on inverse correlation matrices. Our first step is to model the relation between the diagonal and off diagonal elements of W in an explicit manner. Let $W = LL^T$ be the Cholesky decomposition of W where L is a lower triangular matrix. The partitions of W and L are represented as

$$W = \begin{pmatrix} \star & \star & \star \\ \star & w_{jj} & \mathbf{w}_j^T \\ \star & \mathbf{w}_j & W_{jj} \end{pmatrix} \text{ and } L = \begin{pmatrix} \star & 0 & 0 \\ \star & l_{jj} & 0 \\ \star & \mathbf{l}_j & L_{jj} \end{pmatrix}, \tag{3}$$

where the $\star$s are some arbitrary elements of the corresponding matrices. The following proposition explicitly models the constraints imposed on the diagonal elements of W in terms of the vector $\mathbf{l}_j \equiv (l_{j+1j}, l_{j+2j}, \cdots, l_{pj})^T$ for $j = 1, 2, \cdots, p-1$, which consists of the free (unconstrained) elements $l_{ij}, j < i$.

Theorem 2.1 *Let $W = LL^T$ be the Cholesky decomposition of W as given by (3). Then,*

$$l_{jj}^2 = 1 + \mathbf{l}_j^T (L_{jj} L'_{jj})^{-1} \mathbf{l}_j. \tag{4}$$

We refer the reader to a proof in the Appendix. For $j = 1$, we have $w_{i1} = l_{i1} l_{11}$ and so, $w_{i1} = 0 \Leftrightarrow l_{i1} = 0$. For the remaining columns $j \geq 2$, $i_{ij} = 0$ corresponds to w_{ij} being equal to some pre-specified value (but not necessarily zero). There are several attractive properties of the above parametrization of W that we shall use later: (1) For each $j = 1, 2, \cdots, p$, we have an explicit expression for the diagonal element l_{jj} in terms of $l_{ij}, i > j$ and L_{jj}, thus removing the implicit constraint imposed on W; (2) The above proposition allows us to treat the l_{ij}s for $i > j$ as the free (i.e., unconstrained) parameters with each $l_{ij} \in R$, allowing proposal distributions for $\mathbf{l}_j$ be elicited conveniently, and (3) the elements of L_{jj} involve $l_{kk'}$s for indices $k > k' > j$ only, and not any of the $l_{kk'}$'s for $k, k' \leq j$. In particular, in the expression (4) for l_{jj}, L_{jj} does not depend on $\mathbf{l}_j$ or any of the other $\mathbf{l}_k$ for $k < j$.

We also partition $R = W^{-1}$ as

$$R = \begin{pmatrix} \star & \star & \star \\ \star & 1 & \mathbf{r}_j^T \\ \star & \mathbf{r}_j & R_{jj} \end{pmatrix}, \tag{5}$$

and note that $(L_{jj}L_{jj}^T)^{-1} = R_{jj}$ for $j = 2, 3, \cdots, p-1$.

Suppose our dataset consists of n iid observations of $\mathbf{X}$ with pdf in (1) given by $\mathbf{D} = (\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n)$ where $R \in \mathcal{C}_p$ is the unknown correlation matrix with $R^{-1} = W$. Denote S to be the sample covariance matrix defined as $S = \sum_{i=1}^n X_i X_i^T$. The likelihood for W based on the n observations is

$$\ell(\mathbf{D} \,|\, W) = e^{-\frac{1}{2}tr(R^{-1}S)}|R|^{-n/2} = e^{-\frac{1}{2}tr(WS)}|W|^{n/2}. \tag{6}$$

Using the Cholesky decomposition of W in (3), the likelihood can be written as a product of component likelihoods

$$\ell(\mathbf{D}\,|W) = \prod_{j=1}^{p} \ell_j(\mathbf{l}_j \,|\, L_{jj}) \tag{7}$$

where

$$\ell_j(\mathbf{l}_j \,|\, L_{jj}) = (l_{jj}^2)^{n/2} e^{-\frac{1}{2}(s_{jj}l_{jj}^2 + 2l_{jj}\mathbf{s}_j^T\mathbf{l}_j + \mathbf{l}_j^T S_{jj}\mathbf{l}_j)} \tag{8}$$

for $j = 1, 2, \cdots, p$. Also, we give S the representation

$$S = \begin{pmatrix} \star & \star & \star \\ \star & s_{jj} & \mathbf{s}_j^T \\ \star & \mathbf{s}_j & S_{jj} \end{pmatrix}. \tag{9}$$

The Bayesian inference methodology requires the development of a suitable class of prior on $\mathcal{C}_p$ for model selection. We develop the prior on $\mathcal{C}_p$ in a hierarchical way, similar in spirit to the prior in Wong et al. (2003) and Pitt et al. (2006), which gives positive probabilities on combinations of off-diagonal elements of L being identically zero. The details are presented in the subsequent section.

3 Inverse Correlation Selection Priors

We closely follow the exposition in Pitt et al. (2006) and Wong et al. (2003) to develop the framework of inverse correlation selection priors for W. To elicit

the prior on a constrained version of $R^{-1} = W = ((w_{ij}))$, we first develop some additional notation. The prior is developed in terms of the number of zero and non-zero entries of L, the lower triangular matrix in the Cholesky decomposition of W. Let the binary random variable $J_{ij} = 1$ if $l_{ij} \neq 0$ and $J_{ij} = 0$ if $l_{ij} = 0$ for $j < i$ and $i = 2, 3, \cdots, p$. Let $\mathbf{J} = \{\, J_{ij}\, j < i,\, i = 2, 3, \cdots, p \,\}$ denote the collection of all the J_{ij}. The random variable $N(J)$ will denote the total number of elements in $\mathbf{J}$ that are 1 out of the maximum possible number $H = p(p-1)/2$. Let $J_{\cdot j}$ denote the collection of all $\{\, J_{ij},\, i = j+1, j+2, \cdots, p \,\}$ for each $j = 1, 2, \cdots, p-1$. For the j-th column, let I_j denote the collection of indices (i, j) in $J_{\cdot j}$ such that $J_{ij} = 1$ and I_{j+} denote the collection of all indices $\{\, (i,k) \,:\, (i,k) \in J_{\cdot k},\, i > k > j \,\}$ such that $J_{ik} = 1$. The collection of elements of $\mathbf{l}_j$ for the j-th column that are non-zero is denoted by $\mathbf{l}_{I_j}$. Also, let $\mathbf{r}_{I_j}$ and $\mathbf{r}_{I_{j+}}$ denote, respectively, the collection of all r_{ij} with $(i, j) \in I_j$ and $(i, j) \in I_{j+}$. The sets $\mathbf{r}_{I_j}$ and $\mathbf{r}_{I_{j+}}$, respectively, denotes all free (unconstrained) parameters corresponding to $\mathbf{J}$ in column j and in columns $j+1, j+2, \cdots, p$, for each $j = 1, 2, \cdots, p-1$.

We denote the prior distribution on R for a configuration $\mathbf{J}$ by $g(R \,|\, \mathbf{J}\,)$ given by

$$g(R \,|\, \mathbf{J}) \propto \prod_{j=1}^{p-1} (\det(R_{\{I_j, I_j\}}))^{-1/2}; \tag{10}$$

in (10), $R_{\{I_j, I_j\}}$ is the submatrix of R_{jj} consisting of the $(I_j - j)$-th rows and columns of R_{jj}.

Our hierarchical prior specification for W is as follows:

$$\pi_0(R \,|\, \mathbf{J}) = V(\mathbf{J})^{-1}\, dr_{\{\mathbf{J}=1\}}\, I\{r_{\{\mathbf{J}=0\}}\}\, g(\, R \,|\, \mathbf{J}\,), \tag{11}$$

$$\pi_0\{\, \mathbf{J} \,|\, N(\mathbf{J}) = h \,\} = \binom{H}{h}^{-1}, \quad \text{and} \tag{12}$$

$$\pi_0\{\, N(\mathbf{J}) = h \,|\, \psi \,\} = \binom{H}{h} \psi^h\, (1-\psi)^{H-h}, \quad \text{and} \tag{13}$$

$$\pi_0(\psi) = Uniform(0, 1) \tag{14}$$

where

$$V(\mathbf{J}) = \int_{R \in \mathcal{C}_p} g(R \,|\, \mathbf{J})\, I\{r_{\{\mathbf{J}=0\}}\}\, dr_{\{\mathbf{J}=1\}} \tag{15}$$

is the normalizing constant for g, and $0 \leq \psi \leq 1$ is the probability that $J_{ij} = 1$. There are some major differences between the prior elicitation in (11-14) above with that of Wong et al. (2003) and Pitt et al. (2006). First, the prior for R in

(11) is defined in terms of the entries l_{ij} that are non-zero. Given the position of the non-zero l_{ij}s, $N(J)$, the r_{ij} entries at those positions can be taken to be the free parameters for R. The Lebesgue measure $dR_{\mathbf{J}=1}$ in equations (11) and (15) is induced on these r_{ij} elements. The remaining r_{ij} entries are a function of $r_{\mathbf{J}=1}$, but not necessarily zero. Wong et al. (2003), and subsequently Pitt et al. (2006), defined the prior directly in terms of the indices of r_{ij}s that are exactly zero and non-zero, and therefore, is a different approach from the case here. Second, the prior distribution on J given $N(J) = h$ is taken to be uniform on the space of all configurations J satisfying $N(J) = h$. This is different from the prior specification in Wong et al. (2003) who take this prior to depend on the average volume $\bar{V}(h)$ over all such combinations of J. Wong's approach avoids the need to compute $V(J)$ during each update of the Gibbs sampler but requires the computation of an average volumes $\bar{V}(h)$ based on non-linear regression and Monte carlo sampling. In the present case, we take the prior on J to be uniform, which means that we will be required to compute the volume $V(J)$ at each iteration of the Gibbs sampler. However, the expression for the normalizing constant $V(J)$ can be obtained analytically. We present

Theorem 3.1 *Let* **J** *correspond to a configuration in $\mathcal{C}_p$. Then, the volume*

$$V(\mathbf{J}) = \prod_{j=1}^{p-1} V(J_{\cdot j}) = 2^{-(p-1)} \prod_{j=1}^{p-1} \left(\mathcal{B}(\alpha_j, \beta_j)\, V_0(\mathcal{S}^{n_j})\right) \tag{16}$$

with $\alpha_j = n_j/2$, $\beta_j = 1 + n_j/2$ and n_j is cardinality of I_j (i.e., the number of non-zero entries in $\mathbf{l}_j$); in (16), $\mathcal{B}(\alpha, \beta)$ is the Beta function given by $\mathcal{B}(\alpha, \beta) = \int_0^1 x^{\alpha-1}\,(1-x)^{\beta-1}\,dx$, and $V_0(\mathcal{S}^m)$ is the volume of the unit m-dimensional sphere in $\mathbb{R}^m = \frac{2\pi^{m/2}}{m\Gamma(m/2)}$.

The reader is referred to a proof in the Appendix. There is a strong motivation for choosing g as in (11). The prior g, after an appropriate transformation, has the tail behavior of multivariate t with one degree of freedom. Thus, in the univariate case (with $n_j = 1$), this tail behavior is like Cauchy. It is well known that densities with Cauchy tail-like behavior has been proposed by many researchers as the appropriate default prior for the univariate variable; see, for example, Gelman. To demonstrate this tail behavior, we transform the variables $(\mathbf{r}_{I_j}, \mathbf{r}_{I_{j+}}) \rightarrow (\mathbf{l}_{I_j}, \mathbf{l}_{I_{j+}})$ using the Jacobian in (25), and derive the density for $\mathbf{l}_{I_j}$ (conditional on $\mathbf{l}_{I_{j+}}$) as

$$h_j(\mathbf{l}_{I_j} \,|\, \mathbf{l}_{I_{j+}}) = \frac{2}{\mathcal{B}(\alpha_j, \beta_j) \times V_0(\mathcal{S}^{n_j})} \times \frac{(\det(R_{\{I_j, I_j\}}))^{1/2}}{(1 + \mathbf{l}_{I_j}^T R_{\{I_j, I_j\}} \mathbf{l}_{I_j})^{1+(n_j/2)}} \tag{17}$$

for $j = 1, 2, \cdots, p$. Another transformation $\mathbf{y} = R^{1/2}_{\{I_j, I_j\}} \mathbf{l}_{I_j}$ gives the multivariate t density with one degree of freedom for $\mathbf{y}$. In terms of L, the prior on the non-zero entries of L, $\{\, \mathbf{l}_{I_j},\, j = 1, 2, \cdots, p \,\}$ is

$$\pi_0(L \,|\, \mathbf{J}) = \prod_{j=1}^{p} h_j(\mathbf{l}_{I_j} \,|\, \mathbf{l}_{I_{j+}}) \tag{18}$$

where h_j is as (17).

4 Bayesian Inference

Inference is obtained based on the posterior distribution of L. The unknown parameters are (1) the values and (2) positions of non-zero entries of L, and (3) ψ. The posterior of $(L, \mathbf{J}, \psi)$, up to a proportionality constant, is given by

$$\pi(\mathbf{J},\, L,\, \psi \,|\, \mathbf{D}\,) \propto \ell(\mathbf{D} \,|\, L)\, \pi_0(L \,|\, \mathbf{J})\, \pi_0\{\, \mathbf{J} \,|\, N(\mathbf{J})\,\}\, \pi_0\{\, N(\mathbf{J}) \,|\, \psi\,\}\, \pi_0(\psi)$$

from equations (6), (11-14) and (18). The update of $(\mathbf{J}, L, \psi)$ to a new state $(\mathbf{J}^*, L^*, \psi^*)$ may change the number of unconstrained entries of L, and therefore, can be viewed as an updating scheme that moves between parameter spaces of varying dimensions. We develop a posterior sampling procedure based on the Reversible Jump Markov Chain Monte Carlo (RJMCMC) algorithm of Green (1995) and Green and Richardson (1997). Fix ψ and a column j. We consider three updating steps. The first two moves are reversible moves types for $(\mathbf{J}, L)$ for fixed ψ. These updating moves are

• **Remove Zero:** In this step, $(\mathbf{J}, L) \rightarrow (\mathbf{J}^*, L^*)$ by increasing the number of non-zero entries in $J_{\cdot j}$ by one (hence $n_j^* = n_j + 1$). One of the zero entries in $\mathbf{l}_{I_j}$ is selected at random and converted to non-zero. All other entries of L remain the same.

• **Add Zero:** In this step, $(\mathbf{J}, L) \rightarrow (\mathbf{J}^*, L^*)$ by decreasing the number of non-zero entries in $J_{\cdot j}$ by one (hence $n_j^* = n_j - 1$). One of the non-zero entries in $\mathbf{l}_{I_j}$ is selected at random and converted to zero. All other entries of L remain the same.

• **Update ψ:** In this step, no reversible moves are needed. The update of ψ can be carried out using a standard Gibbs step. Given $\mathbf{J}$ and L, the posterior density of ψ,

$$\pi(\psi \,|\, \mathbf{J},\, L,\, \mathcal{D}\,) \sim Beta(\alpha_\psi, \beta_\psi)$$

with $\alpha_\psi = N(\mathbf{J}) + 1$ and $\beta_\psi = H - N(\mathbf{J}) + 1$.

The proposal densities corresponding to the **Remove Zero** and **Add Zero** steps are given as follows: If $\mathbf{J} \rightarrow \mathbf{J}^*$ for a **Remove Zero** move, the proposal density is

$$q_R(\mathbf{J}, \mathbf{J}^*) = r_R(\mathbf{J}, \mathbf{J}^*) \times \frac{1}{(p - j - n_j)} \times q(l_{i^* j}), \tag{19}$$

where r_R is the probability of selecting this move type, $p-j-n_j$ is the number of available zeros for conversion to non-zero and $q(l_{i^* j})$ is the density of the proposal distribution for the selected non-zero position i^* in l_{I_j}. If the transition $\mathbf{J} \rightarrow \mathbf{J}^*$ represents an **Add Zero** move, the associated proposal density is

$$q_A(\mathbf{J}, \mathbf{J}^*) = r_A(\mathbf{J}, \mathbf{J}^*) \times \frac{1}{n_j} \tag{20}$$

where r_A is the probability of selecting this move type; the new number of non-zero elements is $n_j^* = n_j - 1$ where n_j is the number of available non-zeros in $\mathbf{J}$ for conversion to zero. The acceptance probabilities corresponding to the **Remove Zero** move is given by

$$\alpha_R((\mathbf{J}, L), (\mathbf{J}^*, L^*)) = \min \left\{ 1, \frac{\pi((\mathbf{J}^*,\, L^*,\, \psi\,) \,|\, \mathbf{D}\,) q_A(\mathbf{J}^*, \mathbf{J})}{\pi((\mathbf{J},\, L,\, \psi\,) \,|\, \mathbf{D}\,) q_R(\mathbf{J}, \mathbf{J}^*)} \right\}; \tag{21}$$

while the acceptance probabilities for the **Add Zero** step is

$$\alpha_A((\mathbf{J}, L), (\mathbf{J}^*, L^*)) = \min \left\{ 1, \frac{\pi((\mathbf{J}^*,\, L^*,\, \psi\,) \,|\, \mathbf{D}\,) q_R(\mathbf{J}^*, \mathbf{J})}{\pi((\mathbf{J},\, L,\, \psi\,) \,|\, \mathbf{D}\,) q_A(\mathbf{J}, \mathbf{J}^*)} \right\}. \tag{22}$$

The expression of the acceptance probabilities of general RJMCMC schemes involves a Jacobian that corresponds to the transformation relating the random variables generated using the proposal distribution (either using q_A or q_R) with the new proposed state. However, in our case, the Jacobian is 1 since we directly sample $l_{i*,j}$, an element of L. We also consider another step **Unchanged Zero** where $\mathbf{J}$ remains fixed and only the entries of L are updated based on a proposal distribution q_1. The acceptance probability for the **Unchanged Zero** move for the j column of L is

$$\alpha_U((\mathbf{J},\, L,\, \psi\,), (\mathbf{J},\, L^*,\, \psi\,)) = \min \left\{ 1, \frac{\pi((\mathbf{J},\, L^*,\, \psi\,) \,|\, \mathbf{D}\,) q_1(l_{I_j})}{\pi((\mathbf{J},\, L,\, \psi \,|\, \mathbf{D}\,) q_1(l_{I_j}^*)} \right\} \tag{23}$$

where l_{I_j} and $l_{I_j}^*$ represent the current and proposed values for the j-th column of L. Note that since $\mathbf{J}$ remains unchanged , both l_{I_j} and $l_{I_j}^*$ are of the same dimension n_j. The probability of selecting the **Unchanged Zero** move type is denoted

by $r_U(\mathbf{J}, \mathbf{J})$ but does not appear in the expression for the acceptance probability due to cancelation from the numerator and denominator.

To run the RJMCMC, we obtain an initial estimate of W based on the inverse of the sample covariance matrix S. The Cholesky decomposition of $W = LL^T$ is then obtained. In order to update the entries of L, the RJMCMC performs a cycle starting from column $j = p - 1$, then $j = p - 2$, and so on, until $j = 1$ of L. At step j, one of the **Add**, **Remove** and **Unchanged Zero** moves are selected with their corresponding r_0 probabilities. If either **Add Zero** or **Remove Zero** are selected, the chain is updated from state $(\mathbf{J}, W) \rightarrow (\mathbf{J}^*, W^*)$ according to the acceptance probabilities (22) and (21). If **Unchanged Zero** is selected, the chain is updated based on the acceptance probability (23) based on the proposal q_1. Running through the indices $j = p - 1, p - 2, \cdots, 1$ and finally updating ψ based on its conditional posterior density completes one iteration of the RJMCMC. The RJMCMC is then run through a large number of iterations and checked for convergence before posterior samples are obtained for inference.

Suitable choices for the proposal densities q and q_1 are challenging to obtain for the following reason: Note that for each fixed j, the posterior is a complicated function of $\mathbf{l}_j$; for example, $\mathbf{l}_j$ is present in the conditional posterior density $\pi(\mathbf{l}_k \mid L_{kk}, \mathbf{D})$ for all $k < j$ in a very complicated way, and cannot be factored out or approximated easily. To develop efficient sampling procedures, the proposal densities should be as close as possible to $\pi(\mathbf{J}, W, \psi \mid \mathbf{D})$ (viewed as a function of $\mathbf{l}_j$ only) to avoid low acceptance probabilities and slow mixing.

We develop proposal densities for the **Remove Zero** and **Unchanged Zero** steps based on the initial estimates of L, $\hat{l}_{ij}$ for $i > j$ and $i, j = 1, 2, \cdots, p$ obtained from the empirical correlation matrix $\hat{R}$. The candidate non-zero entry l_{i^*j} is sampled from a normal density with mean $\hat{l}_{i^*j}$ and standard deviation σ_0. We choose a small value of σ_0 in our experiments in Section 5. For the **Unchanged Zero** move, q_1 is chosen to be the multivariate normal distribution with mean $\hat{\mathbf{l}}_{I_j}$ and covariance matrix $\sigma_0^2 \, I_{n_j}$.

5 Experimental Results

5.1 Simulation

A commonly used covariance structure is the band structure where going down each column, only a few entries closest to the diagonal are non-zero and rest are zero. For our simulation, we generated observations from a multivariate normal

distribution on R^5 with zero mean and a banded correlation matrix R. The matrices R, W and L are given as follows:

$$R = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -0.28 & 0 & 0 \\ 0 & -0.28 & 1 & -0.2 & 0 \\ 0 & 0.06 & -0.2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, W = R^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1.09 & 0.31 & 0 & 0 \\ 0 & 0.31 & 1.13 & 0.2 & 0 \\ 0 & 0 & 0.2 & 1.04 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\text{and } L = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1.04 & 0 & 0 & 0 \\ 0 & 0.03 & 1.02 & 0 & 0 \\ 0 & 0 & 0.2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

A total of $n = 2,000$ realizations are generated which constitute the observed data. The RJMCMC is started from several different initial values of R, W and L. The variance of the proposal density σ_0^2 is taken to be the sample variance calculated based on the off-diagonal entries of $\hat{L} = ((\hat{l}_{ij}))$.

The RJMCMC chain was run upto 20,000 iterations and convergence of the simulation was checked and established using "Potential Scale Reduction Factor" as described in Brooks and Gelman (1998). The marginal distribution of each l_{ij} is a two component mixture with one component giving point mass at zero and the other forming a smooth density based on the non-zero realizations. Figure 1 gives the density plots using a Gaussian kernel for the non-zero entries as well as the true value of each l_{ij}. The number on top right corner of each panel is the proportion of times that l_{ij} was chosen to be zero. A large number indicates that the posterior puts high probability on the value 0 which is indicated by the grey background of the corresponding panel. Clearly from Figure 1 the sparsity of L and W matrix has been estimated correctly by the sampling scheme. The recovered sparsity structure of R matrix is also similar to the true structure (not shown in figure).

5.2 Real Data

Our method proposes a way to estimate covariance or correlation matrix through sampling. This is applicable in a bigger context to answer biological questions about pathway association. As a biological hypothesis, pathway association refers to association of a group of genetic markers that are functionally related with some

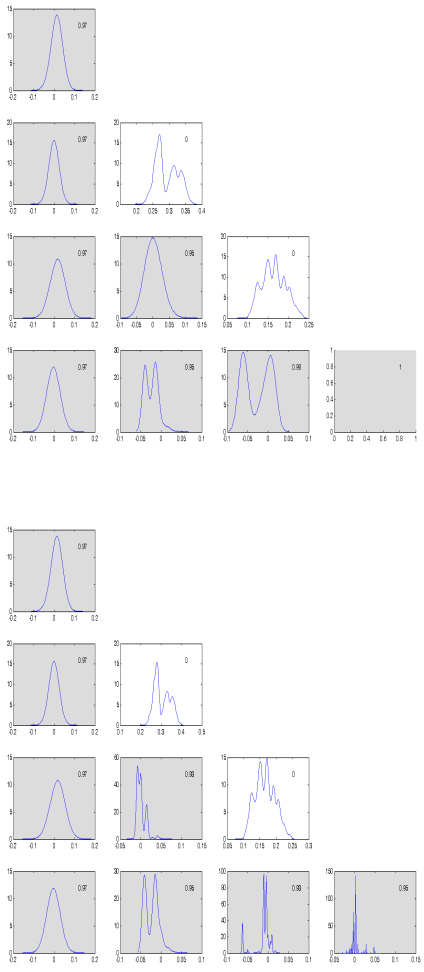

Figure 1: Location-wise Distribution of nonzero values of MCMC sample of lower triangular part of L and W matrix. Grey background indicates posteior mode at 0, and the point probability at 0 is displayed at the top right corner.

phenotype(disease). However, statistical attempts to make joint inference on a set of genes are often inadequate in capturing the possible complexity in the association pattern. In most cases, a combined analysis of all the genes in a study, or a subset of them that are biologically connected, may not be feasible due to small sample sizes and a large number of genes. Biological nature of gene interaction

or dependence may be difficult to capture in the statistical parametrization. Also, quite often the genes in a study is only a subset of all the different types of markers that construct the genetic pathways and the overall dependence pattern of them may not translate in a simplistic way to the network of the subnet.

One recent example of pathway analysis is provided in Hendriksen et al. (2006) where the Androgen pathway has been linked with progression of prostate cancer. The data used in Hendriksen et al. (2006) is available in Gene Expression Omnibus (GSE4084). A similar study, Singh et al. (2008), on Androgen receptor related genes concluded similar association from a different cohort. The significant genes reported in Singh et al. (2008) are classified as over- and under-expressed in cancer specimens. We took the data from Hendriksen et al. (2006) with the over-expressed genes from Singh et al. (2008). Data missingness reduced this set by two more genes. Eventually, $p = 10$ genes in $n = 12$ specimens were available with genes denoted by `AKAP9`, `GAGEB1`, `MET`, `MYLK`, `MYO3A`, `NR2F1`, `NRXN3`, `PRLR`, `TCF4` and `TNS`. We applied our covariance estimation algorithm on this dataset to find possible sparsity structures.

Reported structural data from KEGG database has not been useful as many of the above genes are not part of the Androgen receptor pathway shown there. We will use common wisdom about natural networks in defining our prior. Natural networks in various context shows a degree distribution with polynomial tail with $P(degree = n) \propto 1/n^p$. Recent studies of natural networks suggests $p \approx 3$ (add reference) and some suggestions on growth models suggest even larger values of p. For our purpose, we take $p = 3.5$ to have a prior value of both mean and variance of the degree distribution. For $p = 3.5$, mean degree is $\frac{\zeta(2.5)}{\zeta(3.5)} = 1.19$ and variance is $\frac{\zeta(1.5)}{\zeta(3.5)} - (\frac{\zeta(2.5)}{\zeta(3.5)})^2 = 0.901$. In our setup, total degree $N(J) = \frac{1}{2}\sum node\ degrees$, hence $E(N(J)) = 5 \times E(degree)$, assuming degree distribution of each node is iid. And $V(N(J)) = 25 \times V(degree)$. We incorporate these information into the prior $\psi \sim Beta(\alpha, \beta)$ by choosing $\alpha = 0.9596$ and $\beta = 3.0724$.

Figures 2, 3 and 4 shows the estimated structure of the inverse correlation matrix of the 10 genes along with the L matrix for chains 1, 2 and 3 respectively. Convergence of the simulation has been checked using "Potential Scale Reduction Factor" as before; figure 5 shows the mixture of sequence and within sequence variance estimates against the simulation index. Clearly the simulation has converged by 10,000 iterations.

The locations in figures 2- 4 with grey background shows the location having 0.5 or higher point probability at zero. Although the sparsity estimated seems to

be somewhat lower than observed sparsity of natural networks, the threshold for sparsity can be made more stringent or the prior can be made more skewed to enforce more sparsity provided we have some prior knowledge about the network dependence pattern.

6 discussion

Bayesian estimation and model selection with a correlation matrix is challenging due to the presence of non-linear constraints. We present an approach in this paper that explicitly models the non-linear constraints in terms of the lower triangular matrix of the Cholesky decomposition. Our algorithm reduced the domain restriction on the Cholesky factor to only the diagonal elements being non-stochastic, and all the off-diagonals being free parameters. Thus almost any default prior on them would work. Also, the updating scheme updates one column at a time, and hence the computation time is $O(p)$ where p is the dimension of the correlation matrix. This a significant improvement over the earlier attempts to default bayesian analysis of correlation matrix.

A Bayesian analysis to infer pathway association will go through a model selection exercise with an association parameter and the covariance matrix allowed to have stochastic zeros. Posterior simulation will include steps to sample from the posterior of the association parameter and then another step to sample from the posterior of the covariance matrix conditional on the current value of the association parameter. We intend to demonstrate in the stated applications that our method provides an algorithm to simulate from the covariance matrix posterior that allows for structural zeros.

7 references

References

Barnard, J., McCulloch, R., and Meng, X. (2000). Modeling Covariance Matrices in Terms of Standard Deviations and Correlations with Application to Shrinkage. *Statistica Sinica*, 10:1281–1311.

Brooks, S. and Gelman, A. (1998). General methods for monitoring convergence

of iterative simulations. *Journal of Computational and Graphical Statistics*, 7:434–455.

Hendriksen, P., Dits, N., Kokame, K., Veldhoven, A., van Weerden, W., Bangma, C., Trapman, J., and Jenster, G. (2006). Evolution of the androgen receptor pathway during progression of prostate cancer. *Cancer Research*, 66(10):5012–5020.

Pitt, M., Chan, D., and Kohn, R. (2006). Efficient Bayesian inference for Gaussian copula regression models. *Biometrika*, 93(3):537–554.

Singh, A., Bafna, S., Chaudhary, K., Venkatraman, G., Smith, L., Eudy, J., Johansson, S., Lin, M., and Batra, S. (2008). Genome-wide expression profiling reveals transcriptomic variation and perturbed gene networks in androgen-dependent and androgen-independent prostate cancer cells. *Cancer Letters*, 259:28–38.

Wong, F., Carter, C., and Kohn, R. (2003). Efficient estimation of covariance selection models. *Biometrika*, 90(4):809–830.

Appendix

We provide the proofs of the Theorems 2.1 and 3.1 in this section. The proof of Theorem 2.1 proceeds by partitioning W and L as

$$W = \begin{pmatrix} B_{1j} & \mathbf{b}_j & B_{2j} \\ \mathbf{b}_j^T & w_{jj} & \mathbf{w}_j^T \\ B_{2j}^T & \mathbf{w}_j & W_{jj} \end{pmatrix} = \begin{pmatrix} M_{j1} & 0 & 0 \\ M_j & l_{jj} & 0 \\ M_{j2} & \mathbf{l}_j & L_{jj} \end{pmatrix} \begin{pmatrix} M_{j1}^T & M_j^T & M_{j2}^T \\ 0 & l_{jj} & \mathbf{l}_j^T \\ 0 & 0 & L_{jj}^T \end{pmatrix}.$$

Comparing the blocks we have $B_{1j} = M_{j1}M_{j1}^T$, $B_{2j} = M_{j1}M_{j2}^T$, $W_{jj} = M_{j2}M_{j2}^T + \mathbf{l}_j\mathbf{l}_j^T + L_{jj}L_{jj}^T$, $w_{jj} = l_{jj}^2 + M_jM_j^T$, $\mathbf{w}_j = M_{j2}M_j^T + l_{jj}\mathbf{l}_j$ and $\mathbf{b}_j = M_{j1}M_j^T$. The diagonal entries of the inverse of W are 1. These restrictions translate to

$$w_{jj} - \begin{pmatrix} \mathbf{b}_j^T & \mathbf{w}_j^T \end{pmatrix} \begin{pmatrix} B_{1j} & B_{2j} \\ B_{2j}^T & W_{jj} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{b}_j \\ \mathbf{w}_j \end{pmatrix} = 1. \tag{24}$$

Using the linear algebra result on deriving a matrix inverse, we have

$$\begin{pmatrix} B_{1j} & B_{2j} \\ B_{2j}^T & W_{jj} \end{pmatrix}^{-1} = \begin{pmatrix} B_{1j}^{-1} + FE^{-1}F^T & -FE^{-1} \\ -E^{-1}F^T & E^{-1} \end{pmatrix}$$

where $E = W_{jj} - B_{2j}^T B_{1j}^{-1} B_{2j} = \mathbf{l}_j \mathbf{l}_j^T + L_{jj} L_{jj}^T$ and $F = B_{1j}^{-1} B_{2j} = (M_{j2} M_{j1}^{-1})^T$. After some algebra, the second term on LHS of (24) can be simplified in the following way:

$$w_{jj} - \left(\mathbf{b}_j^T B_{1j}^{-1} \mathbf{b}_j + \mathbf{b}_j^T F E^{-1} F^T \mathbf{b}_j - 2 \mathbf{b}_j^T F E^{-1} \mathbf{w}_j + \mathbf{w}_j^T E^{-1} \mathbf{w}_j \right) = l_{jj} - l_{jj}^2 \mathbf{l}_j^T E^{-1} \mathbf{l}_j.$$

Equating the last expression above to unity, we get $1 = l_{jj}^2 - l_{jj}^2 \mathbf{l}_j^T E^{-1} \mathbf{l}_j$ or $l_{jj}^2 = 1/(1 - \mathbf{l}_j^T (\mathbf{l}_j \mathbf{l}_j^T + L_{jj} L_{jj}^T)^{-1} \mathbf{l}_j)$. Now use the result

$$(\mathbf{l}_j \mathbf{l}_j^T + L_{jj} L_{jj}^T)^{-1} = (L_{jj} L_{jj}^T)^{-1} - \frac{(L_{jj} L_{jj}^T)^{-1} \mathbf{l}_j \mathbf{l}_j^T (L_{jj} L_{jj}^T)^{-1}}{1 + \mathbf{l}_j^T (L_{jj} L_{jj}^T)^{-1} \mathbf{l}_j}$$

to simplify the last expression for l_{jj}^2 to $l_{jj}^2 = 1 + \mathbf{l}_j^T (L_{jj} L_{jj}^T)^{-1} \mathbf{l}_j$. □

Let $\mathbf{J}$ denote a configuration in $\mathcal{C}_p$. We state the following lemma which will be required for the proof of Theorem 3.1.

Lemma **7.1** *The Jacobian of the transformation from* $(r_{I_j}, r_{I_{j+}}) \rightarrow (l_{I_j}, r_{I_{j+}})$ *is given by*

$$\Delta_j = \frac{det(R_{\{I_j, I_j\}})}{l_{jj}^{2+n_j}} \tag{25}$$

where $R_{\{I_j, I_j\}}$ *is the submatrix of* R_{jj} *consisting of the* $(I_j - j)$*-th rows and columns of* R_{jj}*, and* n_j *is the number of elements in* I_j*.*

The proof of Theorem 3.1 now proceeds as follows. We have $l_{jj}^2 = 1 + l_j^T (L_{jj} L_{jj}^T)^{-1} l_j = 1 + l_{I_j}^T R_{\{I_j, I_j\}} l_{I_j}$ where $R_{\{I_j, I_j\}}$ are the $I_j - j$ rows and columns of R_{jj}. Now fix $j = 1$. To evaluate the integral in (15), note that

$$\int_{\mathcal{C}_p} \prod_{j=1}^{p-1} (\det(R_{\{I_j, I_j\}}))^{-1/2} \, I\{r_{\{\mathbf{J}=0\}}\} \, dr_{\{\mathbf{J}=1\}} = \int_{\mathcal{C}_p} \prod_{j=1}^{p-1} (\det(R_{\{I_j, I_j\}}))^{-1/2} \, I\{r_{\{\mathbf{J}=0\}}\} \, dr_{I_{01}} \, dr_{I_{11}}$$

$$= \int_{\mathcal{C}_{p-1}} \left(\int_{l_{I_j}} \det(R_{\{I_j, I_j\}}))^{-1/2} \Delta_1 \, dl_{I_{01}} \right) \prod_{j=2}^{p-1} (\det(R_{\{I_j, I_j\}}))^{-1/2} \, I\{r_{I_{11}^c} = 0\} \, dr_{I_{11}},$$

using Lemma 7.1. The inner integral with respect to l_{I_1} is

$$\int_{l_{I_1}} \frac{(\det(R_{\{I_1, I_1\}}))^{1/2}}{l_{I_1}^{2+n_1}} \, dl_{I_1} = \int_{l_{I_1}} \frac{(\det(R_{\{I_1, I_1\}}))^{1/2}}{(1 + \mathbf{l}_{I_1}^T R_{\{I_1, I_1\}} \mathbf{l}_{I_1})^{1+(n_1/2)}} \, dl_{I_1}$$

$$= \int_{\mathbf{y}} (1 + \mathbf{y}^T \mathbf{y})^{-(1+(n_1/2))} \, d\mathbf{y}$$

where $\mathbf{y} = R^{1/2}_{\{I_1,\, I_1\}}\, l_{I_1}$ is an n_1-dimensional vector. Using polar transformation, the integral with respect to $\mathbf{y}$ can be simplified to

$$\int_{s>0} (1+s^2)^{-(1+n_1/2)}\, s^{n_1-2}\, ds \;\times V_0(\mathcal{S}^{n_1}) = 2^{-1}\cdot \mathcal{B}(\alpha_1,\beta_1)\times V_0(\mathcal{S}^{n_1})$$

using the substitution $u = s^2/(1+s^2)$, where α_1, β_1 and $V_0(\mathcal{S}^{n_1})$ are as defined in Theorem 3.1. Now, we repeat the above procedure for $j = 2, 3, \cdots, p-1$ for the outer integral with respect to $r_{I_{1+}}$. □

We now give the proof of Lemma 7.1. It is easy to check by matrix multiplication that

$$L = \begin{pmatrix} \star & 0 & 0 \\ \star & l_{jj} & 0 \\ \star & \mathbf{l}_j & L_{jj} \end{pmatrix}, \text{ and } L^{-1} = \begin{pmatrix} \star & 0 & 0 \\ \star & 1/l_{jj} & 0 \\ \star & -\frac{1}{l_{jj}} L_{jj}^{-1}\mathbf{l}_j & L_{jj}^{-1} \end{pmatrix}. \tag{26}$$

Hence,

$$R = \begin{pmatrix} \star & \star & \star \\ \star & 1 & r_j^T \\ \star & r_j & R_{jj} \end{pmatrix} = (L^{-1})^T L^{-1} = \begin{pmatrix} \star & \star & \star \\ \star & \frac{1}{l_{jj}^2} + \frac{\mathbf{l}_j^T (L_{jj}^{-1})^T L_{jj}^{-1}\mathbf{l}_j}{l_{jj}^2} & -\frac{\mathbf{l}_j^T (L_{jj}^{-1})^T L_{jj}^{-1}}{l_{jj}} \\ \star & -\frac{1}{l_{jj}}(L_{jj}^{-1})^T L_{jj}^{-1}\mathbf{l}_j & (L_{jj}^{-1})^T L_{jj}^{-1} \end{pmatrix}.$$

Equating the diagonal entries provide a second proof of Theorem 2.1. From the off-diagonal entries, we have $R_{jj} = (L_{jj}^{-1})^T L_{jj}^{-1}$ and $\mathbf{r}_j = -\frac{1}{l_{jj}} R_{jj}\mathbf{l}_j$. Using the notation from section 3, we have $r_{I_j} = -\frac{1}{l_{jj}} R_{\{I_j,\, I_j\}}\mathbf{l}_{I_j}$. Choose the indices i and k such that $(i,j) \in I_j$ and $(k,j) \in I_j$. Then,

$$\frac{\partial\, r_{ij}}{\partial\, l_{kj}} \;=\; -\frac{r_{kj}}{l_{jj}} + \frac{1}{l_{jj}^2} \sum_{(m,j)\in I_j} r_{mj}\, l_{mj}\, \frac{\partial\, l_{jj}}{\partial\, l_{kj}} = -\frac{r_{kj}}{l_{jj}} + \frac{1}{l_{jj}^2} \sum_{(m,j)\in I_j} r_{mj}\, l_{mj}\, \frac{1}{l_{jj}} \mathbf{r}_{I_j}^T \mathbf{l}_{I_j} \tag{27}$$

Note that the Jacobian of the transformation $(\mathbf{r}_{I_j}, \mathbf{r}_{I_{j+}}) \to (\mathbf{l}_{I_j}, \mathbf{r}_{I_{j+}})$ is given by

$$J = det \begin{bmatrix} \frac{\partial\, \mathbf{r}_{I_j}}{\partial\, \mathbf{l}_{I_j}} & \frac{\partial\, \mathbf{r}_{I_{j+}}}{\partial\, \mathbf{l}_{I_j}} \\ \frac{\partial\, \mathbf{r}_{I_j}}{\delta \mathbf{r}_{I_{j+}}} & \frac{\partial\, \mathbf{r}_{I_{j+}}}{\partial\, \mathbf{r}_{I_{j+}}} \end{bmatrix} = det \begin{bmatrix} \frac{\partial\, \mathbf{r}_{I_j}}{\partial\, \mathbf{l}_{I_j}} & \frac{\partial\, \mathbf{r}_{I_{j+}}}{\partial\, \mathbf{l}_{I_j}} \\ 0 & I \end{bmatrix} = det \left[\frac{\partial\, \mathbf{r}_{I_j}}{\partial\, \mathbf{l}_{I_j}} \right],$$

the 0 in the off-diagonal position being due to the fact that the vector $\mathbf{r}_{I_j}$ is not constrained by the variables in $\mathbf{r}_{I_{j+}}$. Thus, the Jacobian of the transformation

$\mathbf{r}_{I_j} \rightarrow \mathbf{l}_{I_j}$ depends on $det\left[\frac{\partial \mathbf{r}_{I_j}}{\partial \mathbf{l}_{I_j}}\right]$ which can be obtained from equation (27). This can be simplified as

$$det\left[\frac{\partial \mathbf{r}_{I_j}}{\partial \mathbf{l}_{I_j}}\right] = det\left[-\frac{1}{l_{jj}}R_{\{I_j,I_j\}} + \frac{1}{l_{jj}^3}R_{\{I_j,I_j\}}\mathbf{l}_{I_j}\mathbf{l}_{I_j}^T R_{\{I_j,I_j\}}\right] = \frac{1}{l_{jj}^{n_j}} det\left[R_{\{I_j,I_j\}}\right] det\left[I - \mathbf{u}\mathbf{u}^T\right]$$

where $\mathbf{u} = \frac{1}{l_{jj}}R_{\{I_j,I_j\}}^{1/2}\mathbf{l}_{I_j}$. The last expression can be simplified as

$$\frac{det\left[R_{\{I_j,I_j\}}\right]}{l_{jj}^{n_j}}(1-\mathbf{u}^T\mathbf{u}) = \frac{det\left[R_{\{I_j,I_j\}}\right]}{l_{jj}^{n_j}}\left(1 - \frac{\mathbf{l}_{I_j}^T R_{\{I_j,I_j\}}\mathbf{l}_{I_j}}{l_{jj}^2}\right)$$

$$= \frac{det\left[R_{\{I_j,I_j\}}\right]}{l_{jj}^{n_j}}\left(1 - \frac{l_{jj}^2 - 1}{l_{jj}^2}\right) = \frac{det\left[R_{\{I_j,I_j\}}\right]}{l_{jj}^{n_j+2}}.\square$$

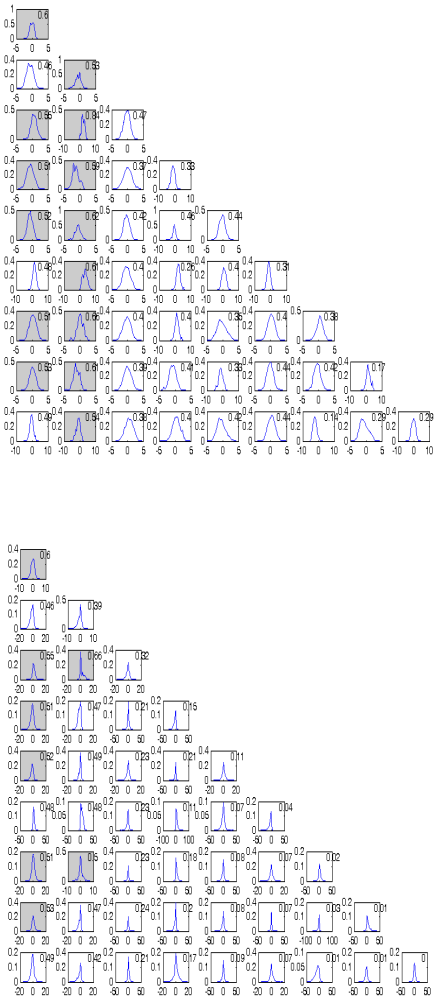

Figure 2: Location-wise Distribution of nonzero values of MCMC sample of lower triangular part of L and W matrix of Androgen pathway genes for chain 1. Grey background indicates posteior mode at 0, and the point probability at 0 is displayed at the top right corner.

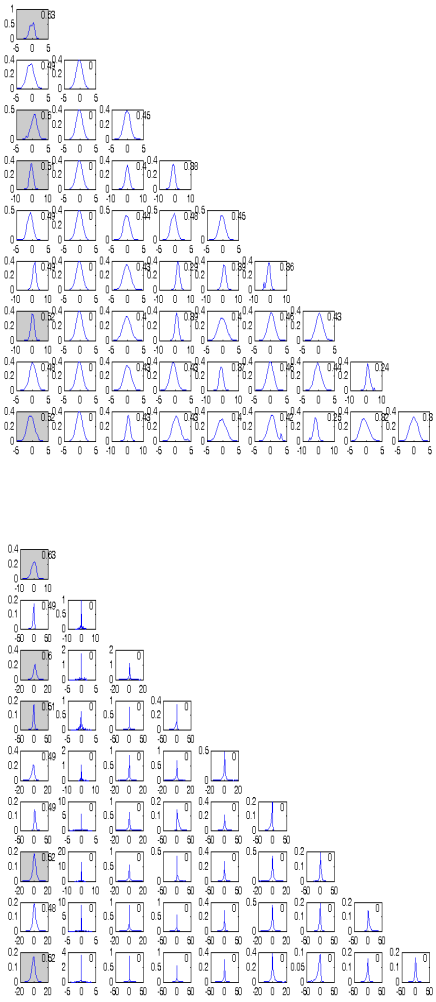

Figure 3: Location-wise Distribution of nonzero values of MCMC sample of lower triangular part of L and W matrix of Androgen pathway genes for chain 2. Grey background indicates posteior mode at 0, and the point probability at 0 is displayed at the top right corner.

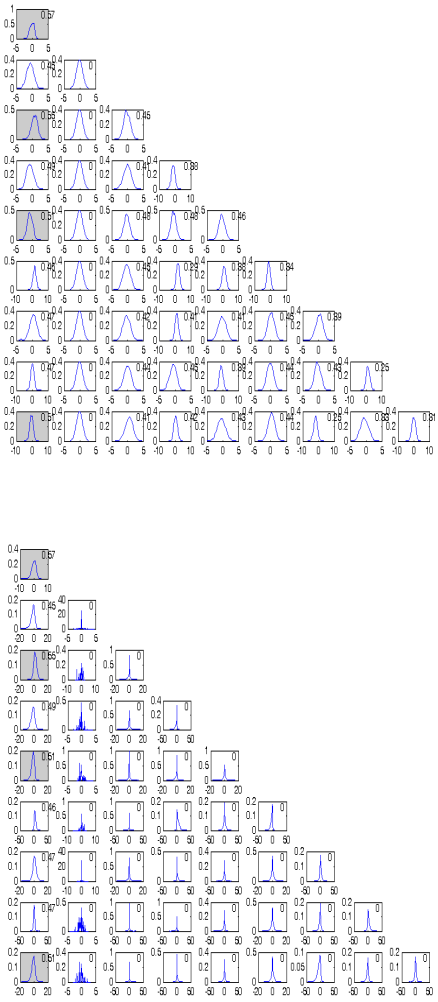

Figure 4: Location-wise Distribution of nonzero values of MCMC sample of lower triangular part of L and W matrix of Androgen pathway genes for chain 3. Grey background indicates posteior mode at 0, and the point probability at 0 is displayed at the top right corner.

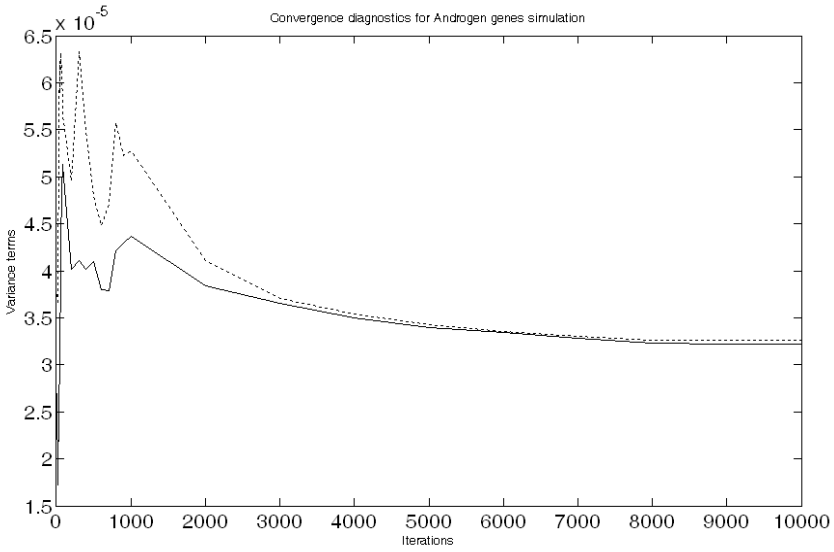

Figure 5: Mixture of sequence and within sequence variance estimates for the covariance matrix estimation of the Androgen genes