Electronic Journal of Statistics
Vol. 5 (2011) 18–30
ISSN: 1935-7524
DOI: 10.1214/11-EJS594

A Metropolis-Hastings based method for sampling from the G-Wishart distribution in Gaussian graphical models

Nicholas Mitsakakis

Toronto Health Economics and Technology Assessment (THETA) Collaborative
Leslie L. Dan Pharmacy Building
University of Toronto
144 College Street
Toronto, ON, M5S 3M2
e-mail: n.mitsakakis@utoronto.ca

Hélène Massam

Department of Mathematics and Statistics
York University, 4700 Keele St.
Toronto, ON, M3J 1P3
e-mail: massamh@mathstat.yorku.ca

and

Michael D. Escobar

Dalla Lana School of Public Health
Health Sciences Building
University of Toronto
155 College Street
Toronto, ON, M5T 3M7
e-mail: m.escobar@utoronto.ca

Abstract: In Gaussian graphical models, the conjugate prior for the precision matrix K is called G-Wishart distribution, $W_G(\delta, D)$. In this paper we propose a new sampling method for the $W_G(\delta, D)$ based on the Metropolis Hastings algorithm and we show its validity through a number of numerical experiments. We show that this method can be easily used to estimate the Deviance Information Criterion, providing with a computationally inexpensive approach for model selection.

Keywords and phrases: Gaussian graphical models, G-Wishart distribution, Metropolis-Hastings algorithm, non-decomposable graphs, Deviance Information Criterion.

Received September 2010.

1. Introduction

Gaussian graphical models are statistical models for multivariate Gaussian data where dependencies and conditional independencies are represented by means of a graph G. The components of the Gaussian variable $X = (X_1, \ldots, X_p)$ are represented by the set of nodes, $V = \{1, \ldots, p\}$, of the graph and the absence of an edge between nodes i and j in V indicates the conditional independence of X_i and X_j given the remaining variables. In a Bayesian framework, in order to perform model selection or covariance estimation, one often uses the popular Diaconis-Ylvisaker conjugate prior ([4]) for the precision matrix $K = \Sigma^{-1}$ of the Gaussian distribution. This prior distribution is called the G-Wishart and was first given for arbitrary undirected graphs G by Roverato ([14]). Its density is given by

$$f(K|\delta, D) \propto |K|^{(\delta-2)/2} \exp\left\{-\frac{1}{2}\mathrm{tr}(K^t D)\right\}, \tag{1}$$

where $\delta > 0$ and D is symmetric positive definite matrix. The support set of this distribution is the set of all positive definite symmetric matrices K that have zero entries K_{ij} whenever (i, j) is a missing edge in the graph G. This set is denoted with $M^+(G)$. Without loss of generality, we can assume that $X \sim N_p(0, \Sigma)$. Since the $W_G(\delta, D)$ distribution is a conjugate prior, for a sample of size n, the posterior distribution of K given the data matrix x is the $W_G(\delta + n, D + x^t x)$. When G is decomposable, the normalizing constant of the $W_G(\delta, D)$,

$$I_G(\delta, D) = \int_{M^+(G)} |K|^{(\delta-2)/2} \exp\left\{-\frac{1}{2} < K, D >\right\} dK,$$

can be computed explicitly (see [13]). However, this is not true for non-decomposable graphs. In that case the normalizing constant cannot be obtained analytically.

The ability to sample from the G-Wishart (or the Hyper Inverse Wishart) distribution is important and has many applications, including the estimation of the normalizing constant. Another application is the estimation of the covariance matrix Σ and functions of it. Indeed in a Bayesian context, one is often interested in the posterior mean of Σ, $E(\Sigma|x, G)$, given a set of data x of size n and a selected model G. Since the posterior distribution of K is given by the $W_G(\delta + n, D + x^T x)$, we can estimate $E(\Sigma|x, G)$ by the quantity $\hat{J}$, where

$$\hat{J} = \frac{1}{N} \sum_{i=1}^{N} (K_i)^{-1}, \tag{2}$$

where K_i are random samples from the $W_G(\delta+n, D+x^T x)$. In a similar manner, one can estimate any function of K.

A method to sample from the G-Wishart distribution when G is non-decomposable was proposed in [3]. Unfortunately, this method is wrong. The purpose of this paper is to offer a new method. In Section 2 we review the existing methods for the simulation of the G-Wishart. In Section 3, we present our new

method, which is based on the Metropolis-Hastings algorithm. In Section 4, we investigate its performance through a number of experiments and use it together with the DIC criterion for model selection.

2. Existing sampling methods

2.1. The decomposable case

Sampling from the Hyper Inverse Wishart distribution for decomposable graphs has been previously discussed in [3]. If $C_1, C_2, \ldots, C_k$ is a perfect ordering of the cliques of G and $S_2, \ldots, S_k$ are the corresponding separators, we use the notation $R_i = C_i \setminus S_i, i = 2, \ldots, k$,

$$\Sigma_{C_i} = \begin{pmatrix} \Sigma_{S_i} & \Sigma_{S_i,R_i} \\ \Sigma_{R_i,S_i} & \Sigma_{R_i} \end{pmatrix}, \tag{3}$$

$$D_{C_i} = \begin{pmatrix} D_{S_i} & D_{S_i,R_i} \\ D_{R_i,S_i} & D_{R_i} \end{pmatrix}, \tag{4}$$

$\Sigma_{R_i|S_i} = \Sigma_{R_i} - \Sigma_{R_i,S_i}\Sigma_{S_i}^{-1}\Sigma_{S_i,R_i}$ and $D_{R_i|S_i} = D_{R_i} - D_{R_i,S_i}D_{S_i}^{-1}D_{S_i,R_i}$. Then the sampling scheme is as follows:

1. Sample $\Sigma_{C_1} \sim IW(\delta, D_{C_1})$, which yields Σ_{S_2}.
2. For $i = 2, \ldots, k$,
   (a) sample $\Sigma_{R_i|S_i} \sim IW(\delta + |S_i|, D_{R_i|S_i})$, $U_i \sim N(D_{R_i,S_i}D_{S_i}^{-1}, \Sigma_{R_i|S_i} \otimes D_{S_i}^{-1})$, with $\otimes$ denoting the Kronecker product,
   (b) compute $\Sigma_{R_i,S_i} = U_i\Sigma_{S_i}$ and $\Sigma_{R_i} = \Sigma_{R_i|S_i} + \Sigma_{R_i,S_i}\Sigma_{S_i}^{-1}\Sigma_{S_i,R_i}$, and subsequently obtain Σ_{C_i} from Equation (4).
3. The precision matrix K is obtained through the formula $K = \sum_{i=1}^{k}[\Sigma_{C_i}^{-1}]^0 - \sum_{i=2}^{k}[\Sigma_{S_i}^{-1}]^0$ ([8]), where for a $|I| \times |J|$ matrix A with $|I|, |J| < p$ we denote with $[A]^0$ the matrix obtained from A by filling up with zeros entries in order to obtain full dimension, i.e. $([A]^0)_{\kappa\lambda} = A_{\kappa\lambda}$ if $(\kappa, \lambda) \in I \times J$, otherwise $([A]^0)_{\kappa\lambda} = 0$.

2.2. The general case

The method in 2.1 cannot be used when the graph is not decomposable, since in that case the marginals for the submatrices of Σ corresponding to prime components do not follow an inverse Wishart distribution [14]. Recently a Gibbs Sampling based method has been proposed for the generation of random samples for the G-Wishart distribution that can be applied to non-decomposable graphs [1]. This method is based on the theoretical results of [12] describing sufficient conditions in regular exponential families for the construction of a block Gibbs sampler for sampling from their natural conjugate densities. When applied to Gaussian graphical models, the block Gibbs sampler becomes the

Bayesian Iterative Proportional Scaling (BIPS) ([6]). For this cyclical method, the number of iterations to obtain one sample point is equal to the number of cliques of the graph. If $C_1, C_2, \ldots, C_k$ is the set of cliques of G, the algorithm is as follows:

- Set $K^0 = I_p$
- For iteration $r = 0, 1, \ldots$ do
  1. Set $K^{r,0} = K^r$
  2. For each $j = 1, \ldots, k$:
     (a) Sample A from Wishart distribution $W(\delta, D_{C_j})$
     (b) Set $K^{r,j}$ so that $K^{r,j}_{C_j,C_j} = A + K^{r,j-1}_{C_j,\bar{C}_j}\left[K^{r,j-1}_{\bar{C}_j,\bar{C}_j}\right]^{-1} K^{r,j-1}_{\bar{C}_j,C_j}$, while $K^{r,j}_{\kappa,\lambda} = K^{r,j-1}_{\kappa,\lambda}$, for $(\kappa, \lambda) \notin C_j \times C_j$
  3. Set $K^{r+1} = K^{r,k}$

After some burn-in time r_0, the sequence $(K^r)_{r>r_0}$ is a set of random samples from the $W_G(\delta, D)$.

The method has been implemented successfully, for example in [9]. However, sampling using BIPS presents various limitations. The algorithm needs the set of cliques of G to be enumerated. This problem is known to be NP-hard ([10]). Also, significant computational time is needed since for the generation of one sample a series of matrix inversions is needed (see step 2b above). In the following section we propose a new sampling method that does not suffer from those limitations. This method is based on the Metropolis Hastings (MH) algorithm.

A very recent addition to the litterature of sampling methods from the Hyper Wishart distribution for non-decomposable graphs is presented in [17], where a rejection sampling method is used for the first prime component in a perfect ordering of the prime components and the same method together with conditioning on the separators is also used for the subsequent prime components.

3. A new sampling method for the $W_G(\delta, D)$ distribution

3.1. The proposed MH-based sampling method

In our proposed method, we make use of the fact that, given a positive definite matrix D, there is bijection between $M^+(G)$ and $M_*^{\triangleleft}(G)$, the space of all the upper triangle matrices incomplete with respect to the graph G. The mapping is described in detail in [2]. In summary, each K in $M^+(G)$ is mapped to $\psi^{\mathcal{V}}$, the projection of the upper triangle matrix ψ on to the space of $\mathcal{V}$-incomplete matrices, where $\psi = \phi T^{-1}$, and ψ, T upper triangle matrices such that $K = \phi^T \phi, D^{-1} = T^T T$. Conversely, the completion ψ of $\psi^{\mathcal{V}}$ can be done with the use of the following equations (see [2], Lemma 2):

$$\psi_{1s} = \sum_{j=1}^{s-1} (-\psi_{1j} h_{js}), \tag{5}$$

while for $(r,s) \in \bar{\mathcal{V}}, r > 1$,

$$\psi_{rs} = \sum_{j=r}^{s-1}(-\psi_{rj}h_{js}) - \sum_{i=1}^{r-1}\left(\frac{\psi_{ir} + \sum_{j=i}^{r-1}\psi_{ij}h_{jr}}{\psi_{rr}}\right)\left(\psi_{is} + \sum_{j=i}^{s-1}\psi_{ij}h_{js}\right), \quad (6)$$

where $h_{ij} = t_{ij}/t_{jj}$, t_{ij} being the (i,j) entry of matrix T. Once ψ is completed, K is given by $(\psi T)^T \psi T$.

The density of $\psi^{\mathcal{V}}$ is such that

$$p(\psi^{\mathcal{V}}|G,\delta,D) \propto \exp\left\{-\frac{1}{2}\sum_{(i,j)\in\bar{\mathcal{V}}}\psi_{ij}^2\right\} \times \prod_{i=1}^{p}\chi_{\delta+\nu_i} \times N_{|E|}(\mathbf{0}_{|E|}, I_{|E|}), \quad (7)$$

where ν_i is the number of nodes j connected with node i such that $j > i$, $\chi_{\delta+\nu_i}$ denotes the distribution of ψ_{ii} when ψ_{ii}^2 follows the chi-square distribution with $\delta + \nu_i$ degrees of freedom, and $N_{|E|}(\mathbf{0}_{|E|}, I_{|E|})$ denotes the multivariate normal distribution with zero mean and covariance matrix equal to the identity matrix, of dimension equal to the number of edges $|E|$. The proposed MH algorithm described here can be used to sample from the distribution of $\psi^{\mathcal{V}}$. Then, samples from the G-Wishart can be obtained using the mapping described above. The proposed algorithm is an *independence chain* ([16]), with proposal density

$$h(\psi^{\mathcal{V}}) = \prod_{i=1}^{p}\chi_{\delta+\nu_i} \times N_{|E|}(\mathbf{0}_{|E|}, I_{|E|}) \quad (8)$$

The acceptance probability is equal to

$$min\left\{\frac{w(\psi^{\mathcal{V}}_{prop})}{w(\psi^{\mathcal{V}}_{cur})}, 1\right\},$$

where

$$w(\psi^{\mathcal{V}}) = \exp\left\{-\frac{1}{2}\sum_{(i,j)\in\bar{\mathcal{V}}}\psi_{ij}^2\right\}.$$

Since $w(\psi^{\mathcal{V}})$ is uniformly bounded (by 1) the chain is *uniformly ergodic* (the strongest convergence rate condition in use, [11]).

We now present the pseudo code of the method:

- Initialize the chain by sampling $\psi_0^{\mathcal{V}}$ from h, as in Equation (8); set $\psi^{\mathcal{V}}_{cur} = \psi_0^{\mathcal{V}}$
- For i =1, 2, ... N do:
  1. Sample $\psi^{\mathcal{V}}_{prop}$ from h
  2. Set
$$\log\alpha = \frac{1}{2}\sum_{(i,j)\in\bar{\mathcal{V}}}\left\{(\psi_{cur})_{ij}^2 - (\psi_{prop})_{ij}^2\right\}$$

3. If $\log \alpha > 0$ then do:
   - $\psi^{\mathcal{V}}_{cur} = \psi^{\mathcal{V}}_{prop}$
   - $accept = 1$
4. Else do:
   - Sample b from $Bernoulli(\alpha)$
   - If $b = 1$ then do:
     * $\psi^{\mathcal{V}}_{cur} = \psi^{\mathcal{V}}_{prop}$
     * $accept = 1$
   - Else $accept = 0$

For each $\psi^{\mathcal{V}}_{cur}$ there is a unique corresponding matrix K_{cur} that can be constructed following the reverse procedure of what is described at the beginning of this section. The sequence of K_{cur} gives a sample from the G-Wishart distribution.

The proposed sampling method has been implemented in R and the code is available from the authors.

4. Experiments

4.1. Experiment 1: Sampling from the prior

In order to investigate the performance of the MH sampling method under different scenarios we conducted a number of numerical experiments. For the first experiment, we sampled from the prior distribution of K, using a seven-nodes non-decomposable graph, as shown on Figure 1). We use as matrix D the identity matrix and $\delta = 3$. We ran the chain for 10,000 iterations, discarding the first 2,000 as burn-in samples. For the assessment of the samples we compare with exact distributions that are available, since in this example graph G contains complete prime components. From [14] we know that, if $K \sim W_G(\delta, D)$ and C is

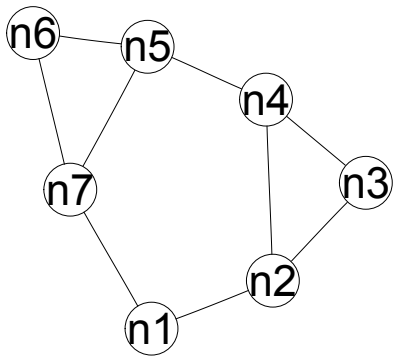

FIG 1. *Non-decomposable graph used in Experiment 1.*

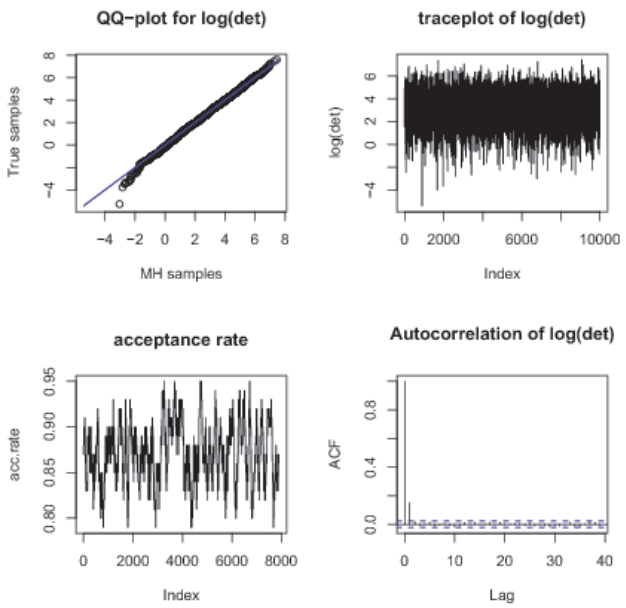

FIG 2. *Results from Experiment 1, sampling from the prior distribution of K.*

a complete prime component of G, the submatrix Σ_C, where $\Sigma = K^{-1}$, follows the inverse Wishart distribution, with parameters δ and D_C, with D_C being the corresponding submatrix of D. If $\{K^{(i)}\}_i$ is the sample generated by the MH method, we compute the determinants

$$d^{(i)} = \det([\Sigma_C^{(i)}]^{-1}), \tag{9}$$

where $\Sigma^{(i)} = [K^{(i)}]^{-1}$. The empirical distribution of the sample $\{d^{(i)}\}_i$ is then compared with the empirical distribution of $\det(X)$, where $X \sim W(\delta, D_C)$. Q-Q plots can be used for the comparison. We also calculate the acceptance rates, autocorrelation and generate the trace plots. From the results shown in Figure 2 we see that under the simple case of $D = I$ and despite the fact that G is non-decomposable, the sampling distribution is very similar to the true distribution. Also the acceptance rate is quite high, suggesting that the samples are approximately independent. The trace plot indicates that the chain seems to be mixing well.

4.2. Experiment 2: Sampling from the posterior

We also perform a similar experiment using the same graph, but this time sampling from the posterior distribution of K. For that we use a simulated dataset of 50 observations generated from a multivariate normal distribution with zero mean and covariance matrix Σ such that $\Sigma^{-1} \in M^+(G)$. Because of

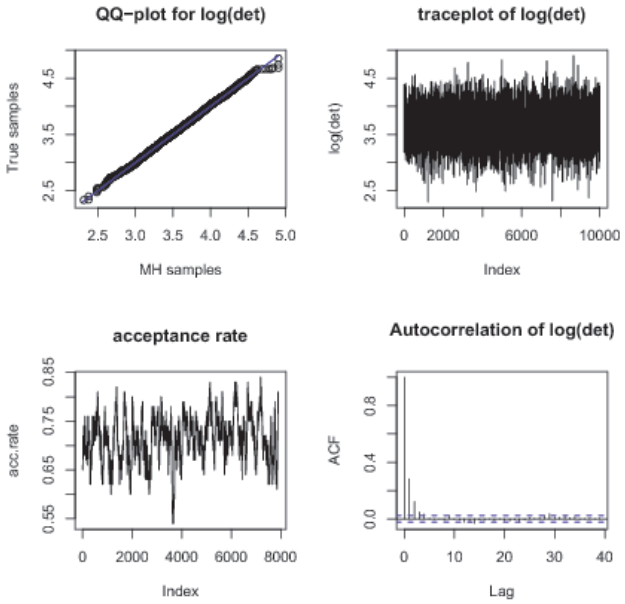

FIG 3. *Results from Experiment 2, sampling from the posterior distribution of K.*

the way the simulated data are generated, the graph G is a plausible model for X. It is important therefore to ensure that the algorithm is sampling well from the posterior distribution. As in Experiment 1, we used $D = I$ and $\delta = 3$. Again, we ran the chain for 10,000 iterations, discarding the first 2,000 as burn-in samples. For the assessment of the samples we use similar methods as in Experiment 1. The results presented in Figure 3 show good mixing of the chain, high acceptance rate, low autocorrelation and high proximity to the true distribution, indicating that the sampler performs very well.

4.3. Experiment 3: Comparison with the Block Gibbs Sampler

In a third experiment, we compare the MH method with the only other available method, the Block Gibbs Sampler. We use an 11-nodes graph, taken from [7]. The graph was selected by the method of *graphical lasso* to represent the conditional independencies of a dataset, containing flow cytometry of 11 proteins measured on 7466 cells. The graph is shown in Figure 4. It is decomposable, but we prefer testing the sampler with a non-decomposable graph. We therefore modify the graph, by removing the edge (PIP2, P38). The resulting graph is non-decomposable, since it contains a chordless cycle of length 4, {PIcg, PIP2, Akt, P38} [8]. We sample from the posterior distribution using as prior hyperparameters, $D = I_{11}$ and $\delta = 3$. Figure 5 shows the autocorrelation plots of the $\log\det$ of the precision matrix for the two method. The MH method required 38.1 seconds of elapsed CPU time per 1000 effective samples, while the Block

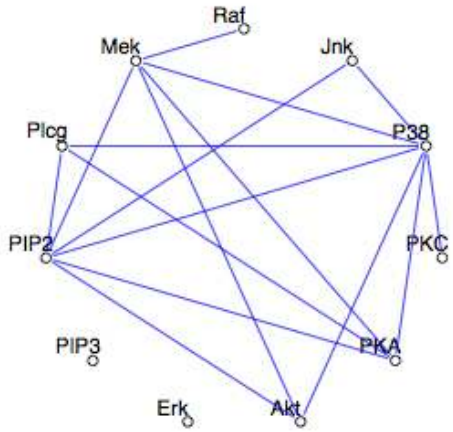

FIG 4. *Graphical model estimated from a flow cytometry dataset, with $p = 11$ proteins measured on $N = 7466$ cells. Taken from [7].*

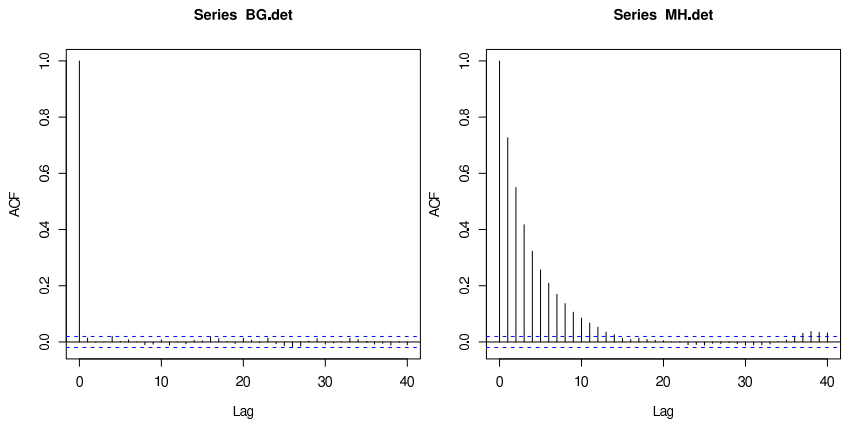

FIG 5. *Autocorrelation plots for* $\log\det(K)$ *using Metropolis-Hastings and Block Gibbs Sampling methods.*

Gibbs Sampler only 4.4. Using samples from equal computational times (40sec) we also compare the empirical densities of the det of the precision matrix. The result in Figure 6 shows that the posterior density estimates of $\det(K)$ from the two methods are very similar. However, the estimate from MH shows signs of multiples modes. This could be because the estimate is based on an effective sample size 9.3 times smaller than the one from the Block Gibbs sampler.

Following the results of those experiments we can conclude that the MH algorithm seems to provide with a reliable method for sampling from the G-Wishart distribution, with relatively small computational expense.

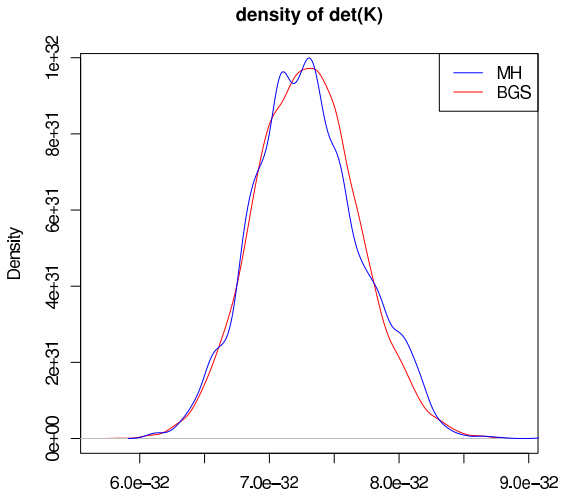

FIG 6. *Density estimation for* $\det(K)$ *using Metropolis-Hastings and Block Gibbs Sampling methods.*

4.4. Application to model selection using the DIC

In this section we present an application of the proposed MH-based sampling method to model selection. The selection is performed with the use of the Deviance Information Criterion (DIC) ([15]), a criterion similar in spirit to the AIC and BIC. It provides with a measure of how well a particular model fits some existing data, while taking into account how *easily* the model fits the data (how many parameters it effectively has). It can therefore be used for Bayesian model selection. Given a dataset x and an unknown parameter θ, the deviance is defined as $D(\theta) = -2\log(p(x|\theta)) + C$, where $p(x|\theta)$ is the likelihood function and C a constant that depends only on the data. Under a specific model and posterior distribution of θ given x one can define the expectations $\bar{\theta} = E_{\theta|x}[\theta]$ and $\bar{D} = E_{\theta|x}[D(\theta)]$. The Deviance Information Criterion can be calculated as

$$DIC = p_D + \bar{D}, \tag{10}$$

where $p_D = \bar{D} - D(\bar{\theta})$, is the effective number of parameters of the model, a Bayesian measure of model complexity ([15]). The larger p_D is, the textiteasier for the model to fit the data. On the other hand the quantity $\bar{D}$ measures how *well* the model fits the data, with larger $\bar{D}$ indicating a worse fit. Overall, models with smaller DIC are preferred to those with larger DIC values. One of the charasteristics of DIC that makes it attractive over other model selection criteria is the fact that it can be calculated when an MCMC or other sampling method for the posterior distribution of θ is available. In that case $\bar{\theta}$ and $\bar{D}$ can be estimated by

$$\frac{1}{N}\sum_{i=1}^{N}\theta_i$$

and

$$\frac{1}{N}\sum_{i=1}^{N} D(\theta_i)$$

respectively, where $\theta_i, i = 1, \ldots, N$ are samples from the posterior distribution of θ given x. The standard error of $\bar{D}$ can be estimated by

$$\sqrt{\frac{\sum_{i=1}^{N}(D(\theta_i) - \bar{D})^2}{N(N-1)}}.$$

In Gaussian graphical Models, the precision matrix K plays the role of the parameter θ, with known posterior distribution $W_G(\delta + n, D + x^T x)$. The DIC of a graph G can be defined as in Equation (10), and $\bar{K}$ and $\bar{D}$ can be estimated after sampling efficiently from the $W_G(\delta + n, D + x^T x)$ with the use of the proposed MH-method.

We now present a numerical example of the calculation of DIC for various graphical models. We used the well known 4-dimensional *Iris flower data set* ([5]), containing the measurements of the length and width of sepals and petals from 150 samples of Iris flowers. The samples are taken equally from three species of the flower, *Iris setosa*, *Iris virginica* and *Iris versicolor*. For our experiment only the 50 samples of *Iris virginica* were used. Since $p = 4$, there are $2^{p(p-1)/2} = 2^6 = 64$ possible four-node graphical models. We excluded the no-edge model and we estimated the DIC for the 63 remaining models. For the hyperparameters of the prior we use the values $D = I_4, \delta = 3$. Under each model we generated 10,000 samples from the posterior distribution of K, using the MH-based sampling method. The initial 2,000 iterations are discarded as "burn-in". Using those samples we estimate $\bar{K}$ and $\bar{D}$, and subsequently calculate the DIC values. We examine the consistency of the DIC values with the values of the log of the ratio of normalizing constants (equal to the marginal likelihood $p(x|G)$ up to a constant multiplying factor), denoted with λ. Figure 7 shows a scatterplot of the values of -DIC against λ. The values of the two measures seem to be well correlated, which gives the indication that model selection based on DIC value is similar to the one based on the marginal likelihood and that our sampler from the G-Wishart distribution worked well. Similar indication is offered by the observation that 9 out of the 10 best models based on DIC belong to the set of the 10 best models according to λ values.

5. Conclusions

The main contribution of this paper is the proposal of a new sampling method from the G-Wishart distribution, based on the Metropolis-Hastings algorithm. A first series of experiments showed satisfactory results and improved efficiency over existing sampling methods, such as the Block Gibbs Sampler. In addition, efficient sampling using this method can be used for the estimation of the Deviance Information Criterion (DIC), which, as our experiments suggest, can

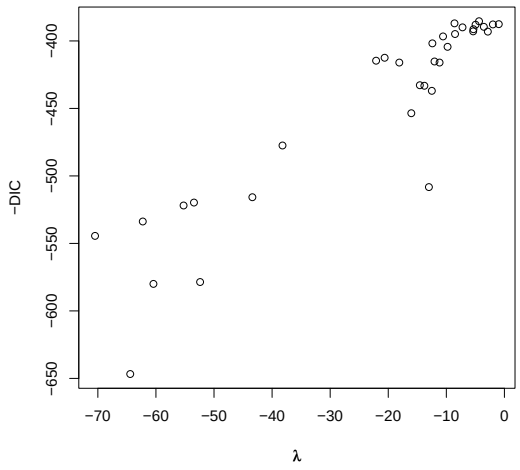

FIG 7. *Scatterplot between the log ratio of the normalizing constants, λ, and the value of -DIC.*

provide with a computationally inexpensive method for model selection. Further investigation is needed for fine tuning of the method in order to improve its performance.

Acknowledgements

The authors thank Alex Lenkoski for the R code for the Block Gibbs Sampler. This study was funded by a CIHR New Emerging Teams (NETs) grant and NSERC Discovery Grant A8947.

References

[1] ASCI, C. and PICCIONI, M. (2007). Functionally compatible local characteristics for the local specification of priors in graphical models. *Scandinavian Journal of Statistics* **34** 829–840. MR2396941

[2] ATAY-KAYIS, A. and MASSAM, H. (2005). A Monte Carlo method for computing the marginal likelihood in nondecomposable Gaussian graphical models. *Biomertika* **92** 317–335. MR2201362

[3] CARVALHO, C. M., MASSAM, H. and WEST, M. (2007). Simulation of Hyper Inverse-Wishart Distributions in Graphical Models. *Biometrika* **94** 647–659. MR2410014

[4] DIACONIS, P. and YLVISAKER, D. (1979). Conjugate priors for exponential families. *Annals of Statistics* **7** 269–281. MR0520238

[5] FISHER, R. A. (1936). The Use of Multiple Measurements in Taxonomy Problems. *Annals of Eugenics* **7** 179–188.

[6] GELMAN, A., CARLIN, J. B., STERN, H. S. and RUBIN, D. B. (1995). *Bayesian data analysis*. Chapman and Hall. MR1385925

[7] Hastie, T., Tibshirani, R. and Freedman, J. (2009). *Elements of Statistical Learning*, Second ed. Springer-Verlag. MR2722294
[8] Lauritzen, L. S. (1996). *Graphical Models*. Oxford University Press. MR1419991
[9] Lenkoski, A. and Dobra, A. (2010). Computational Aspects Related to Inference in Gaussian Graphical Models with the G-Wishart Prior. *Journal of Computational and Graphical Statistics*. To appear.
[10] McHugh, J. (1990). *Algorithmic Graph Theory*. Prentice-Hall.
[11] Mengersen, K. L. and Tweedie, R. L. (1996). Rates of Convergence of the Hastings and Metropolis Algorithms. *The Annals of Statistics* **24** 101–121. MR1389882
[12] Piccioni, M. (2000). Independence structure of natural conjugate densities to exponential families and the Gibbs Sampler. *Scandinavian Journal of Statistics* **27** 111–127. MR1774047
[13] Roverato, A. (2000). Cholesky decomposition of a hyper inverse Wishart matrix. *Biometrika* **87** 99–112. MR1766831
[14] Roverato, A. (2002). Hyper Inverse Wishart Distribution for Non-decomposable Graphs and its Application to Bayesian Inference for Gaussian Graphical Models. *Scandinavian Journal of Statistics* **29** 391–411. MR1925566
[15] Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **64** 583–639. MR1979380
[16] Tierney, L. (1994). Markov Chains for Exploring Posterior Distributions. *The Annals of Statistics* **22** 1701–1762. MR1329166
[17] Wang, H. and Carvalho, C. M. (2010). Simulation of Hyper-Inverse Wishart Distributions for Non-decomposable Graphs. *Electronic Journal of Statistics* **4** 1470–1475.