An importance sampler for graphical Gaussian model inference

Alberto Roverato*, University of Modena
Joe Whittaker†, Lancaster University

21 September 1997

Abstract

While inference in decomposable graphical Gaussian models is completely solved the non-decomposable case still poses difficulties concerned with the evaluation of normalising constants. We propose an importance sampler based on the asymptotic distribution of the maximum likelihood estimates, or equivalently of the posterior distribution in a conjugate analysis, to solve this integration numerically. An example of model comparison based on Bayes factors involving non-decomposable models is given.

Keywords: Bayes factor, graphical Gaussian model, importance sampling, Isserlis matrix, matrix completion, normalising constant.

1 Introduction

Let $\boldsymbol{X}$ be a p-dimensional Normal random vector with mean zero and positive definite concentration matrix $\Omega = \Sigma^{-1}$. In graphical Gaussian models the conditional independence structure of X is represented by means of an undirected graph $G = (V, \mathcal{V})$ with $V = \{1, \ldots, p\}$. A missing edge from the graph, $(i, j) \notin \mathcal{V}$, denotes pairwise conditional independence and corresponds to a zero element in Ω, $\omega_{ij} = 0$, (see, for instance, Whittaker, 1990; Lauritzen, 1996).

It is natural for a model selection procedure to choose candidates from the set of all graphs and a restriction to the subset of decomposable models with chordal graphs is somewhat artificial. While the analysis of decomposable models may be reduced to the analysis of saturated marginal models corresponding to the cliques of the graph, in the non-decomposable case this is not possible and inference for the incomplete prime components of the graph still presents open problems. Large sample approximation to standard errors of non-decomposable models (Roverato and

*Dipartimento di Economia Politica, Viale J. Berengario n.51, 41100 Modena, Italy; email: roverato@unimo.it

†Department of Mathematics and Statistics, Lancaster University, LA1 4YF, UK; email: joe.whittaker@lancaster.ac.uk

Whittaker, 1996) are available, but in the Bayesian approach the difficulties due to the evaluation of normalising constants (see Dellaportas *et al.*, 1996) has proscribed the computation of the Bayes factor when at least one of the competing models is non-decomposable.

In this paper we propose an importance sampler for the numerical evaluation of the normalising constant in a conjugate analysis of the graphical Gaussian model. The problems are twofold. First the support of the required distribution is that of a positive definite matrix constrained by a given zero structure. Sampling from standard matrix distributions such as the Wishart give zero probability to this support (in the decomposable case this problem may be circumvented by factorising the support and sampling from the clique marginals). Secondly, the shape of the required distribution over its support needs to be closely tracked by the sampler as the high dimensional spaces (p variables give $p(p+1)/2$ parameters) quickly lead to numerical inaccurancy. We solve these problems by using the asymptotic normal distribution (Roverato and Whittaker, 1997) and checking that the acceptance probability on the support is sufficiently high.

The notation and a description of the problem are presented in Section 2. In Section 3 we describe the importance sampling procedure. In Section 4 we present an example of near neighbour model comparison based on the Bayes factor. We end with a discussion.

2 Background

2.1 Notation

Let V be a finite set with $|V| = p$, and let Γ be a $p \times p$ symmetric invertible matrix. The rows and columns of Γ are indexed by the elements of V, so that Γ itself is indexed by $V \times V$. When $V = \{1, \ldots, p\}$, Γ is indexed by row and column numbers.

The Isserlis matrix of Γ, Iss(Γ), (Isserlis, 1918; Roverato and Whittaker, 1997) is the symmetric matrix indexed by $\mathcal{W} \times \mathcal{W}$ where $\mathcal{W} = \{(i,j) : i,j \in V, i \leq j\}$, with elements

$$\{\mathrm{Iss}(\Gamma)\}_{(i,j),(r,s)} = \gamma_{ir}\gamma_{js} + \gamma_{is}\gamma_{jr}.$$

For an arbitrary undirected graph $G = (V, \mathcal{V})$ we use the convention that for all $i \in V$ the pair (i,i) is included in the edge set $\mathcal{V}$ and that if $(i,j) \in \mathcal{V}$ then $i \leq j$. The set $\mathcal{W}$ is therefore the edge set of the complete graph and we denote by $\bar{\mathcal{V}} = \mathcal{W} \backslash \mathcal{V}$ the set of edges not in G. For example, for the graph in Figure 1, with $V = \{1,2,3\}$, $\mathcal{V} = \{(1,1),(2,2),(3,3),(1,2),(2,3)\}$ and $\bar{\mathcal{V}} = \{(1,3)\}$.

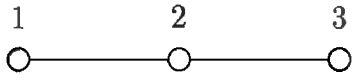

Figure 1: Example of undirected graph.

For any undirected graph $G = (V, \mathcal{V})$ the pair $(\mathcal{V}, \bar{\mathcal{V}})$ is a partition of $\mathcal{W}$. To

this correspond the submatrices $\mathrm{Iss}(\Gamma)_{\nu\nu}$, $\mathrm{Iss}(\Gamma)_{\nu\bar{\nu}}$ and $\mathrm{Iss}(\Gamma)_{\bar{\nu}\bar{\nu}}$ as well as the partial matrix $\mathrm{Iss}(\Gamma)_{\nu\nu|\bar{\nu}} = \mathrm{Iss}(\Gamma)_{\nu\nu} - \mathrm{Iss}(\Gamma)_{\nu\bar{\nu}}[\mathrm{Iss}(\Gamma)_{\bar{\nu}\bar{\nu}}]^{-1}\mathrm{Iss}(\Gamma)_{\bar{\nu}\nu}$.

For a set $\mathcal{C} \subseteq \mathcal{W}$ we define the $\mathcal{C}$-incomplete matrix $\Gamma^{\mathcal{C}}$ as the symmetrised matrix indexed by $V \times V$ with elements $\{\gamma_{ij}\}$ for all $(i,j) \in \mathcal{C}$, and with the remaining elements unspecified. For example, in the graph in Figure 1 above the incomplete matrices corresponding to the sets $\mathcal{V}$ and $\bar{\mathcal{V}}$ are respectively

$$\Gamma^{\mathcal{V}} = \begin{pmatrix} \gamma_{11} & \gamma_{12} & * \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \\ * & \gamma_{32} & \gamma_{33} \end{pmatrix} \quad \text{and} \quad \Gamma^{\bar{\mathcal{V}}} = \begin{pmatrix} * & * & \gamma_{13} \\ * & * & * \\ \gamma_{31} & * & * \end{pmatrix},$$

where asterisks denote unspecified elements. The matrix $\Gamma^{-\mathcal{V}}$ is a shorthand for $(\Gamma^{-1})^{\mathcal{V}}$. If it is possible to fill an incomplete matrix $\Gamma^{\mathcal{C}}$ to obtain a (full) positive definite matrix we say that $\Gamma^{\mathcal{C}}$ admits a positive completion.

Let $\Gamma^{\mathcal{V}}$ be a $\mathcal{V}$-incomplete matrix, with $G = (V, \mathcal{V})$, which admits a positive completion. We say that Γ_G is *the completion of* $\Gamma^{\mathcal{V}}$ if it is the unique positive definite matrix such that

$$(\Gamma_G)^{\mathcal{V}} = \Gamma^{\mathcal{V}} \qquad \text{and} \qquad \{\Gamma_G^{-1}\}_{ij} = 0 \ \text{ for all } \ (i,j) \in \bar{\mathcal{V}}. \tag{1}$$

See Grone *et al.* (1984) for a proof of the existence and uniqueness of such matrix.

When G is non-chordal matrix completion can be performed using the standard iterative proportional fitting algorithm (Whittaker, 1990). Note that (1) can be reformulated as the equations satisfied by the maximum likelihood estimate of Σ for a given graphical model.

For an undirected graph $G = (V, \mathcal{V})$, we denote by $\mathcal{M}_*(G)$ the set of all $\mathcal{V}$-incomplete matrices and by $\mathcal{M}_*^+(G)$ the set of all $\mathcal{V}$-incomplete matrices that admit positive completion. Furthermore we denote by $\mathcal{M}_0(G)$ the set of all symmetric matrices indexed by $V \times V$ with element (i,j) equal to zero whenever $(i,j) \in \bar{\mathcal{V}}$. The intersection of $\mathcal{M}_0(G)$ with the set of all positive definite matrices is denoted by $\mathcal{M}_0^+(G)$.

The trace of the product of two square matrices is $\mathrm{tr}(\Phi\Gamma) = \sum_{i=1}^{p}\sum_{j=1}^{p}\phi_{ij}\gamma_{ij}$. If $\Phi \in \mathcal{M}_0(G)$ only the specified elements of $\Gamma^{\mathcal{V}}$ enter into this sum, and so we can write $\mathrm{tr}(\Phi\Gamma) = \mathrm{tr}(\Phi\Gamma^{\mathcal{V}})$.

2.2 Normalising constants

The function $k(\Omega|A^{\mathcal{V}}, b) = C(A^{\mathcal{V}}, b)g(\Omega|A^{\mathcal{V}}, b)$ on $\Omega \in \mathcal{M}_0^+(G)$ specified by

$$g(\Omega|A^{\mathcal{V}}, b) = \exp\left\{-\frac{b}{2}\mathrm{tr}(\Omega A^{\mathcal{V}}) + \frac{b}{2}\log|\Omega|\right\}, \tag{2}$$

with the normalising constant

$$C(A^{\mathcal{V}}, b) = \left[\int_{\mathcal{M}_0^+(G)} g(\Omega|A^{\mathcal{V}}, b)\, d\Omega\right]^{-1}, \tag{3}$$

is relevant in the statistical inference of graphical Gaussian models. Its parameters are an incomplete matrix $A^{\mathcal{V}} \in \mathcal{M}_*^+(G)$ and a positive constant b.

With $A = S$, the sample covariance matrix, and $b = n$, the sample size, $k(\Omega|S^{\mathcal{V}}, n)$ is proportional to the likelihood function and has maximum value at $\widehat{\Omega} = S_G^{-1}$ (Speed and Kiiveri, 1986) the inverse of the completion of $S^{\mathcal{V}}$ with respect to G. In Bayesian inference, for the choice of hyper parameters $A^{\mathcal{V}} = D^{\mathcal{V}}$ and $b = h$, $k(\Omega|D^{\mathcal{V}}, h)$ is the density of the conjugate prior for Ω. The corresponding posterior distribution has density $k(\Omega|T^{\mathcal{V}}, d)$ where $T = (hD^{\mathcal{V}} + nS^{\mathcal{V}})/(h+n)$ and $d = h+n$. When compatible conjugate priors are used, the Bayes factor for comparing models $G_1 = (V_1, \mathcal{V}_1)$ and $G_2 = (V_2, \mathcal{V}_2)$ can be written as

$$B(G_1, G_2) = \frac{P(D|G_1)}{P(D|G_2)} = \frac{C(D^{\mathcal{V}_1}, h)C(T^{\mathcal{V}_2}, d)}{C(T^{\mathcal{V}_1}, d)C(D^{\mathcal{V}_2}, h)}. \tag{4}$$

The multivariate normal distribution

$$N\left(A_G^{-\mathcal{V}}, b^{-1}\mathrm{Iss}(A_G^{-1})_{\mathcal{V}\mathcal{V}|\bar{\mathcal{V}}}\right) \tag{5}$$

is the asymptotic sampling distribution of the maximum likelihood estimates, and the limiting distribution of the conjugate posterior, when $A^{\mathcal{V}}$ and b are appropriately chosen (Roverato and Whittaker, 1997).

When G is chordal the conjugate prior is hyper Wishart (Dawid and Lauritzen, 1993), and it is the hyper Markov combination with Wishart marginals $W(b + |C| + 1, (bA_{CC})^{-1})$ over the cliques C of the graph. In this case the solution in closed form of integral (3) is available. In the non-decomposable case the integral (3) can be simplified by decomposing it according to the prime components of the graph and their separators. However for an incomplete prime component the integral is intractable and we consider its numerical evaluation in the next section.

3 The importance sampler

Importance sampling evaluates $C = \int g(x)dx$ approximately by drawing N samples $\{x_1, \ldots, x_N\}$ from a distribution with density $q(x)$ and computing $\widehat{C} = N^{-1}\sum_{i=1}^{N} \frac{g(x_i)}{q(x_i)}$, for instance see Fishman (1996).

The importance sampler $q(x)$ we propose for the numerical evaluation of the normalising constant $C(A^{\mathcal{V}}, b)$ is based on (5). The rationale is that as $b \to \infty$ the distribution of Ω is asymptotically equal to (5).

Pseudo code outlining the procedure is:

initialise:
- Sum=0
- input b, $A^{\mathcal{V}}$
- evaluate the mean of (5): $\psi = A_G^{-1}$
- evaluate the variance of (5): $\Phi = b^{-1}\mathrm{Iss}(A_G^{-1})_{\mathcal{V}\mathcal{V}|\bar{\mathcal{V}}}$
- make a small sample correction: $\psi_c = \frac{b+p_s+1}{b}\psi$, $\Phi_c = \frac{b+p_s+1}{b}\Phi$

loop over $i = 1$ to N
- generate $\Omega_i^{\mathcal{V}} \sim N(\psi_c, \Phi_c)$
- fill $\Omega_i^{\mathcal{V}}$ with zeros to give Ω_i

evaluate the Cholesky decomposition: Chol(Ω_i)
if the Cholesky decomposition fails (i.e. $\Omega_i \not> 0$) then reject Ω_i
else
evaluate $g = g(\Omega_i | A^\nu, b)$ from (2) (use Chol(Ω_i) for the determinant)
evaluate $q = q(\Omega_i^\nu)$ from the density of $N(\psi_c, \Phi_c)$
increment Sum= Sum $+g/q$
end loop
return Sum/N

The efficiency of the procedure crucially depends on the acceptance rate, determined by the proportion of random incomplete matrices that, filled with zeros, are positive definite. There are two points to note: firstly the Cholesky decompostion is efficient because used in sequential form it allows an early rejection of a non positive matrix; and if the matrix is positive the procedure evaluates a required determinant. Secondly, the small sample correction to transform the mean and variance of (2) by the multiplying factor $(b + p_s + 1)/b$ is a tuning constant. If the prime component is complete then $p_s = p$, the number of variables of the component, and makes the first two moments equal to those of the corresponding Wishart. Otherwise our experience suggests that p_s be chosen equal to the largest clique size of the prime component. This substantially improves the acceptance rate when b is small.

Table 1 gives some indication of the accuracy of the procedure. The logarithm of the normalising constant, $\log C$, for decomposable (saturated) models with $p = 6$ (21 parameters) are calculated exactly and compared to estimates, $\log \widehat{C}$, from the importance sampler. Increasing values of b are considered while A is given an intraclass correlation form (unitary diagonal elements and all off-diagonal elements equal to ρ with $-1/(p-1) < \rho < 1$ so as to assure $A > 0$). For $p = 6$ we set $\rho = 0.4$, the mid point of the interval $[-1/(p-1), 1]$. The approximation is really rather good, but as the acceptance rate is low when b is small, additional sampling is required. In

Table 1: Exact and estimated constants for the saturated model with $p = 6$ (21 parameters).

b	$\log C(A,b)$	$\log \widehat{C(A,b)}$	CPU time	sample size	accept. rate
1	-34.8929	-34.8415	91.74	600	16.74%
2	-21.5596	-21.5385	63.20	400	21.94%
4	-6.9924	-6.9857	33.45	200	33.45%
6	2.9681	2.9656	26.75	150	45.32%
10	18.4858	18.4871	29.50	150	66.22%
15	34.6974	34.6964	32.75	150	83.70%
20	49.3941	49.3928	33.75	150	92.88%
30	76.7027	76.7029	23.29	100	98.85%
50	127.7852	127.784	9.69	40	99.98%
75	188.9897	188.989	5.62	25	100.00%

Table 2 we compute $\log \widehat{C}$ for chordless cycle models of increasing dimensions. We set

$b = p/2$, half of the minimum sample size required for the estimation of the model, and this keeps the acceptance rate nearly constant as p increases. Figure 3 shows

Table 2: Normalising constant for chordless cycle models.

p	b	no. param.	$\log \widehat{C(A^{\nu}, b)}$	CPU time	sample size	accept. rate
4	2	8	-5.1780	21.65	400	45.87%
8	4	16	3.1423	55.30	400	36.64%
12	6	24	20.2257	108.85	400	38.49%
16	8	32	45.3430	197.08	400	44.59%
20	10	40	78.0100	324.33	400	52.56%

the convergence of the procedure for the chordless cycle with $p = 20$. Although an acceptable approximation is reached quite quickly the low acceptance rate makes the variance of the procedure decrease rather slowly.

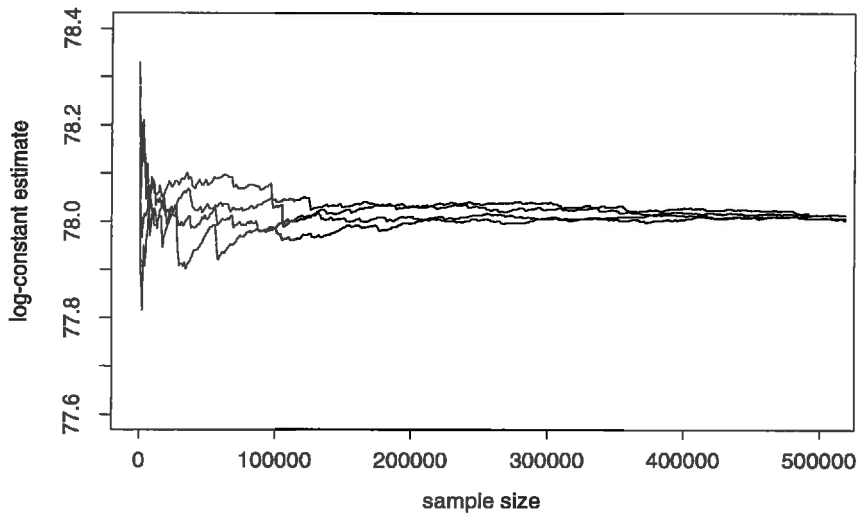

Figure 2: Convergence of procedures initialised with different random seeds for a chordless cycle model with $p = 20$ (40 parameters) and $b = 10$.

Figure 3 gives some instances of the CPU time required for the computations.

4 Example

In this Section we present an application of the procedure to the computation of Bayes factors for real data. Where the comparison involves only decomposable models the exact Bayes factor is also given.

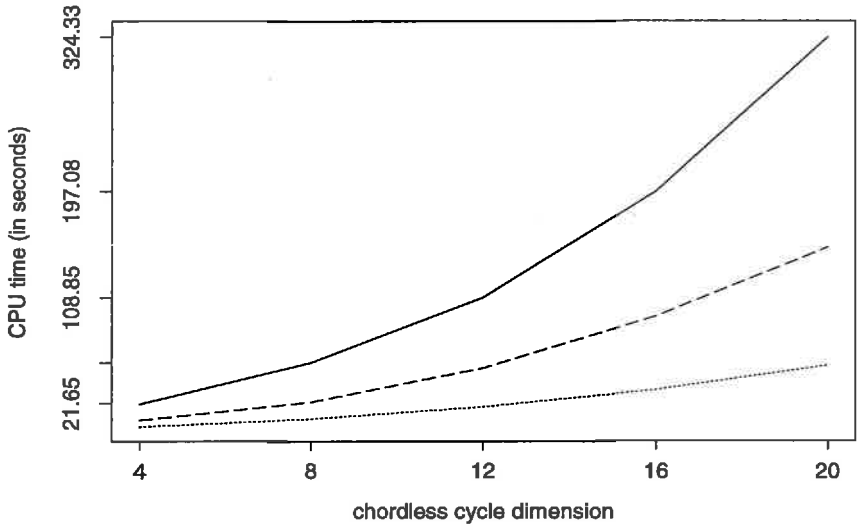

Figure 3: Time required on a Sun Sparc Ultra 2 220 station for the estimation of the normalising constant for chordless cycle models; full line: $b = p/2$, dashed line: $b = p$, dotted line: $b = 2p$.

Table 3 gives summary statistics for six variables measured in Genoa and Padua paediatric hospitals on 107 three month old babies. These data come from a larger

Table 3: Summary statistics for the HIV data: sample variances (main diagonal), correlations (lower triangle) and partial correlations (upper triangle).

X_1	8.8374	0.479	-0.043	-0.033	0.356	-0.236
X_2	0.483	0.1919	0.068	-0.084	-0.224	-0.110
X_3	0.220	0.057	8924231.9	0.085	0.552	-0.330
X_4	-0.040	-0.133	0.149	20392.4	0.091	0.013
X_5	0.253	-0.124	0.523	0.179	1952795.2	0.384
X_6	-0.276	-0.314	-0.183	0.064	0.213	1.378
	X_1	X_2	X_3	X_4	X_5	X_6

Italian study (Boccuzzo, 1991) investigating early diagnosis of HIV infection in children from HIV positive mothers. The variables are related to various measures on blood and its components: X_1 and X_2 to immunoglobin G and A, respectively; X_4 to the platelet count; X_3, X_5 lymphocyte B and T4, respectively; and X_6 to the T4/T8 lymphocyte ratio.

Discussion with the experts running the study suggests the presence of a strong association between variables X_1, X_2 and between variables X_3, X_5, X_6; together with an association structure of these variables compatible with the graph of Figure 4.

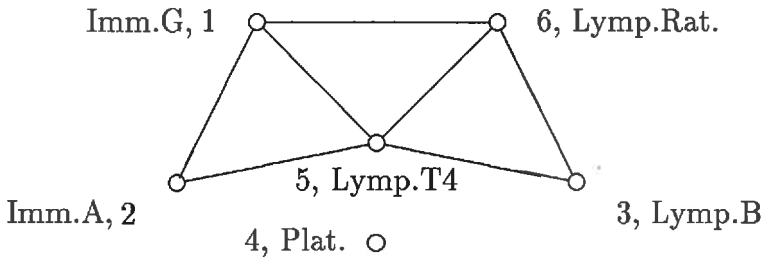

Figure 4: Hypothesised model for HIV diagnosis.

A minimal form of model checking is to compare the decomposable model suggested by the experts with graph in Figure 4 to neighbouring models that differ exactly by one edge. Some of these are non-decomposable.

We used the conjugate prior and set the hyper parameter h equal to **3.** The hyper parameter D is specified in the form $\Delta^{1/2}P\Delta^{1/2}$, where P is the intraclass correlation matrix given above and $\Delta = \text{diag}(10, 10^0, 10^7, 10^5, 10^7, 10)$ incorporates the information on the scale of the variables.

Table 4 gives the Bayes factors for the nearest neighbouring comparison where G_1 is always the smaller model. The values are logged for comparison with Jeffreys'scale (Jeffrey, 1961, Appendix B).

Table 4: Hypothesized model for HIV diagnostic data: nearest neighbour comparisons.

(i,j)	$\log_{10}(B)$	$\log_{10}(\widehat{B})$	(i,j)	$\log_{10}(B)$	$\log_{10}(\widehat{B})$
excluded edges			included edges		
$(1,2)$	-6.4414	-6.4440	$(1,3)$	0.9582	0.9411
$(1,5)$	*non-d.*	-3.1855	$(1,4)$	0.7425	0.7314
$(1,6)$	-1.6479	-1.6403	$(2,3)$	*non-d.*	0.4855
$(2,5)$	-0.7296	-0.6940	$(2,4)$	0.3283	0.3001
$(3,5)$	-8.5698	-8.5547	$(2,6)$	0.2194	0.2036
$(3,6)$	-1.8332	-1.8114	$(3,4)$	0.0959	0.1033
$(5,6)$	*non-d.*	-4.3846	$(4,5)$	-0.6297	-0.6518
			$(4,6)$	0.0675	0.0551

The entries in Table 4 show that the approximation error is of the order of the second decimal place. This is due to the high variance of the procedure evaluating the constant of the prior and would be reduced if the prior were more informative. However it is far from affecting the inference made.

5 Discussion

There are other ways to construct an importance sampler.

For a graphical Gaussian model with graph $G = (V, \mathcal{V})$ function (2) can be alternatively expressed as either a function of the moment parameter $\Sigma^{\mathcal{V}} \in \mathcal{M}_*^+(G)$ or of the canonical parameter $\Sigma_G^{-1} = \Omega \in \mathcal{M}_0^+(G)$. Consider the following diagram involving the sets described in Section 2.1,

$$\begin{array}{ccccc} \Omega & \in & \mathcal{M}_0^+(G) & \subset & \mathcal{M}_0(G) \\ & & \updownarrow & & \updownarrow \\ \Sigma^{\mathcal{V}} & \in & \mathcal{M}_*^+(G) & \subset & \mathcal{M}_*(G). \end{array}$$

The difference between $\mathcal{M}_0(G)$ and $\mathcal{M}_*(G)$ is only notational and these two sets have an obvious bijective mapping. However the bijective mapping between $\mathcal{M}_0^+(G)$ and $\mathcal{M}^+{}_*(G)$ is not trivial because if $\Gamma^{\mathcal{V}} \in \mathcal{M}_*^+(G)$ the corresponding full matrix completed with zeros is not necessarily positive definite. So the mapping between $\mathcal{M}_*^+(G)$ and $\mathcal{M}_0(G)$ requires the operations of matrix completion and inversion: any matrix $\Sigma^{\mathcal{V}} \in \mathcal{M}_*^+(G)$ uniquely identifies a matrix $\Omega = \Sigma_G^{-1} \in \mathcal{M}_0^+(G)$.

Constructing an importance sampler based on the moment parameter has the advantage that samples from $\mathcal{M}_*^+(G)$ are easy to draw, for example, by marginalising the sample obtained from an appropriate Wishart distribution. However in the non-decomposable case the evaluation of (2) requires a matrix completion operation, to be performed numerically, for each sample. This makes the procedure highly inefficient.

On the other hand the approach based on the canonical parameter allows an efficient evaluation of (2) but the standard matrix distributions give zero probability to $\mathcal{M}_0^+(G)$.

Our proposed solution is based on obtaining a random realisation $\Omega^{\mathcal{V}}$ from $\mathcal{M}_*(G)$, considering the corresponding $\Omega \in \mathcal{M}_0(G)$ and checking if this element belongs to $\mathcal{M}_0^+(G)$. This is only feasible if the importance sampler gives positive probability to $\mathcal{M}_0^+(G)$; it is efficient if this probability is high and its shape closely resembles that of $g(\cdot)$.

Our importance sampler works well when b is large, for instance, when evaluating the constant for a posterior distribution. It is less efficient when b is small relative to the number of variables in the largest incomplete prime component of the graph.

Acknowledgements: This work was completed when the authors were visitors at the Isaac Newton Mathematical Institute, Cambridge.

References

BOCCUZZO, G. (1991). *Funzioni Diagnostiche dell'Infezione Connatale da HIV.* Thesis. Department of Statistics. University of Padua.

DAWID, A. P. & LAURITZEN, S. L. (1993). Hyper Markov laws in the statistical analysis of decomposable graphical models. *Annals of Statistics*, **21**, 1272-1317.

DELLAPORTAS, P., GIUDICI, P. & GARETH, R. (1996). Bayesian inference for nondecomposable graphical Gaussian models. *Research Report Series n. 96-17,*

Cambridge University Statistical Laboratory.

FISHMAN G.S. (1996). *Monte Carlo: Concepts, Algorithms and Applications.* Springer-Verlag, New York.

GRONE, R., JOHNSON, C.R., SÀ, E.M. & WOLKOWICE, H. (1984). Positive definite completions of partial Hermitian matrices. *Linear Algebra and its Applications*, **58**, 109-124.

ISSERLIS, L. (1918). On a formula for the product-moment correlation of any order of a normal frequency distribution in any number of variables. *Biometrika*, **12**, 134-139.

JEFFREYS, H. (1961). *Theory of Probability*, 3rd ed. Oxford University Press, Oxford.

LAURITZEN, S. L. (1996). *Graphical Models.* Oxford University Press, Oxford.

ROVERATO, A. & WHITTAKER, J. (1996). Standard errors for the parameters of graphical Gaussian models. *Statistics and Computing*, **6**, 297-302.

ROVERATO, A. & WHITTAKER, J. (1997). The Isserlis matrix and its application to non-decomposable graphical Gaussian models. *Technical report.* Department of Mathematics and Statistics. Lancaster University.

SPEED, T.P. & KIIVERI, H. (1986). Gaussian Markov distributions over finite graphs. *Annals of Statistics*, **14**, 138-150.

WHITTAKER, J. (1990). *Graphical Models in Applied Multivariate Statistics.* Wiley, Chichester.