Electronic Journal of Statistics
Vol. 4 (2010) 1470–1475
ISSN: 1935-7524
DOI: 10.1214/10-EJS591

# Simulation of hyper-inverse Wishart distributions for non-decomposable graphs

**Hao Wang**

*Department of Statistics, University of South Carolina, Columbia, South Carolina 29208, U.S.A.*
*e-mail:* haowang@sc.edu

**and**

**Carlos M. Carvalho**

*McCombs School of Business, The University of Texas, Austin, Texas 78712, U.S.A.*
*e-mail:* carlos.carvalho@mccombs.utexas.edu

**Abstract:** We propose an efficient solution to the problem of direct sampling from a hyper-inverse Wishart distribution in non-decomposable graphs. The method relies on local computations based on the standard junction tree representation of graphs and distribution theoretical results of constraint Wishart matrices.

**Keywords and phrases:** Hyper-inverse Wishart, junction trees, non-decomposable graphs, posterior simulation.

Received May 2010.

## 1. Introduction

[3] proposed a general method to direct sample from the hyper-inverse Wishart for both decomposable and non-decomposable graphs. The approach is based on the compositional form of the joint distribution over a sequence of subgraphs defined by the junction tree as if $\Sigma \sim \mathrm{HIW}_G(b, D)$, where $G = (V, E)$, and for a sequence of perfectly ordered prime components $\{P_1, S_2, P_2, \ldots, P_k\}$ the density takes the form

$$p(\Sigma|b, D) = p(\Sigma_{P_1}) \prod_{i=2}^{k} p(\Sigma_{P_i}|\Sigma_{S_i}). \tag{1}$$

The efficiency of the proposed sampling strategy derives from the fact that all matrix operations are done at the component level and therefore the complexity of the algorithm is dependent only on the dimension of the largest prime component of $G$.

In decomposable graphs their methodology follows directly from basic conditioning results of normals and Wishart distributions and it works perfectly. For

non-decomposable graphs, however, they used the general distributional theory for the Cholesky decomposition of a hyper-inverse Wishart defined by [2] in conjunction with their decomposition idea. In this paper, we identify two problems in their approach and propose a direct sampling solution that preserves the computational efficiencies associated with matrices computations being carried out at the level of the prime components.

For clarity of presentation we assume from this point forward that $G = (V, E)$ can be decomposed into two prime components and one separator $\{P_1, S_2, P_2\}$. Assume further that the subgraph $G_{P_1}$ is complete and $G_{P_2}$ is not. This implies no loss of generality as the proposed sampling process can be repeated down the junction tree in the presence of both complete and incomplete additional components.

## 2. Sampling in a single non-decomposable prime component

Let us start by focusing on sampling from the marginal distribution of the $p \times p$ covariance matrix $\Sigma_{P_2} \sim \mathrm{HIW}_{P_2}(b, D_{P_2})$. Remember that $G_{P_2}$ is a non-decomposable subgraph of $G$.

Define $K = \Sigma_{P_2}^{-1}$ so that $K \sim W_{G_{P_2}}(b, D_{P_2})$, a $G$-Wishart distribution with density

$$p(K \mid G_{P_2}) \quad \propto \quad |K|^{(b-2)/2} \exp\left\{-\frac{1}{2}\mathrm{tr}(KD_{P_2})\right\} 1_{\{K \in M^+(G_{P_2})\}}. \tag{2}$$

Write $D_{P_2}^{-1} = T'T$ and $K = \Phi'\Phi$ as Cholesky decompositions and define $\Psi = \Phi T^{-1}$. Following the nomenclature of [2], the free elements of $\Phi$ are those $\phi_{ij}$ such that $(i, j)$ is an edge in $P_2$. From Theorem 1 and equation (38) of [2], these free elements have density defined by

$$p[\psi_{11}^2, \ldots, \psi_{pp}^2, \{\psi_{ij}\}_{\substack{(i,j)\in E \\ i<j}}] \propto \exp\left\{-\frac{1}{2}\sum_{\substack{(i,j)\notin E \\ i<j}} \psi_{ij}^2\right\} \prod_{i=1}^{p} \chi^2_{b+\nu_i}(\psi_{ii}^2) \prod_{\substack{(i,j)\in E \\ i<j}} \mathrm{N}(\psi_{ij} \mid 0, 1) \tag{3}$$

where $p$ is the cardinality of subgraph $G_{P_2}$ and the $\psi_{ij}$s, for $(i, j) \notin E$ and $i < j$, are well-defined functions of the free elements of $G_{P_2}$.

Based on expression (3), Section 4.3 of [2] suggested a direct algorithm to sample $K$ using chi-squares and normal random variables. This suggestion was adapted in [3] in the context of the decomposition in (1). This is where the first problem with their approach appears. The derivations of [2] are correct but the sampling suggestion used in [3] is not. According to (3), the free elements are not independent normal and chi-square random variates; they are implicitly dependent through the term of $\exp\{-\frac{1}{2}\sum_{\substack{(i,j)\notin E \\ i<j}} \psi_{ij}^2\}$. We suggest the following rejection sampling modification [7] as a fix for the problem:

1. Sample $\Psi$ following the Step 1 and 2 in Section 4.3 of [2], and $u \sim \mathrm{U}[0,1]$.
2. Check whether $u < \exp\{-\frac{1}{2}\sum_{\substack{(i,j)\notin E \\ i<j}} \psi_{ij}^2\}$. If this holds, accept $\Psi$ as a sample from (3); if not, reject the value of $\Psi$ and repeat the sampling step.
3. Construct a sample of $K$ following Step 3 and 4 in Section 4.3 of [2].

The algorithm in [3] only has Step (i) and (iii). However, a sample of $\Psi$ in Step (i) is not from equation (3), but from the instrumental distribution:

$$f[\psi_{11}^2,\ldots,\psi_{pp}^2,\{\psi_{ij}\}_{\substack{(i,j)\in E \\ i<j}}] = \prod_{i=1}^{p} \chi^2_{b+\nu_i}(\psi_{ii}^2) \prod_{\substack{(i,j)\in E \\ i<j}} \mathrm{N}(\psi_{ij} \mid 0,1).$$

Therefore, the Accept-Reject method in Step (ii) is necessary to ensure that this sample is from the true distribution of $\Psi$, that is, equation (3). Note that $\exp\{-\frac{1}{2}\sum_{\substack{(i,j)\notin E \\ i<j}} \psi_{ij}^2\} \leq 1$. Then

$$p[\psi_{11}^2,\ldots,\psi_{pp}^2,\{\psi_{ij}\}_{\substack{(i,j)\in E \\ i<j}}] \leq C f[\psi_{11}^2,\ldots,\psi_{pp}^2,\{\psi_{ij}\}_{\substack{(i,j)\in E \\ i<j}}]$$

where $C$ is the inverse of the normalizing constant of the distribution of equation (3). By Corollary 2.17 and Algorithm A.4 of [7], we can generate a correct sample of $\Psi$ using the rejection sampling of Step (ii).

## 3. Sampling HIW distributions for non-decomposable graphs

With the correct mechanism to sample from a single non-decomposable prime component available, we can revisit the decomposition strategy of [3] to draw samples of $\Sigma$ in all of $G$. It should be clear that sampling $\Sigma_{P_1}$ is a simple task as $\Sigma_{P_1} \sim \mathrm{IW}(b, D_{P_1})$ so the problem is reduced to sampling from the conditional distribution $p(\Sigma_{P_2} \mid \Sigma_{S_2})$ as $\Sigma_{S_2}$ is a complete subset of $\Sigma_{P_1}$. The second mistake in [3] was the suggestion that obtaining samples from this conditional distribution in non-decomposable graphs could be done by matching elements of the Cholesky decomposition of the $\Sigma_{P_2}^{-1}$ with the ones obatained from the decomposition of $\Sigma_{S_2}^{-1}$ [see 3, Section 3.3]. We now fix this problem by defining the appropriate conditioning strategy.

Write $\Sigma_{P_2}$ and $K = \Sigma_{P_2}^{-1}$ as

$$\Sigma_{P_2} = \begin{pmatrix} \Sigma_{S_2} & \Sigma_{S_2,R} \\ \Sigma_{R,S_2} & \Sigma_R \end{pmatrix} \quad \text{and} \quad K = \Sigma_{P_2}^{-1} = \begin{pmatrix} K_{S_2} & K_{S_2,R} \\ K_{R,S_2} & K_R \end{pmatrix}$$

where $\Sigma_{R,S_2} = \Sigma'_{S_2,R}$ and $K_{R,S_2} = K'_{S_2,R}$. Also define

$$\Sigma_{R\cdot S_2} = \Sigma_R - \Sigma_{R,S_2}\Sigma_{S_2}^{-1}\Sigma_{S_2,R} \quad \text{and} \quad \Gamma_{R\cdot S_2} = \Sigma_{R,S_2}\Sigma_{S_2}^{-1}.$$

By noting that $S_2$ is complete so that the subgraph $G_{P_2}$ is collapsible onto $S_2 \subseteq P_2$ Corollary 1 of [8] implies that $(\Sigma_{R\cdot S_2}, \Gamma_{R\cdot S_2}) \perp\!\!\!\perp \Sigma_{S_2}$. Moreover, note that $K_R = \Sigma^{-1}_{R\cdot S_2}$ and $K_{S_2,R} = -\Gamma'_{R\cdot S_2}\Sigma^{-1}_{R\cdot S_2}$ implying further that

$$(K_R, K_{S_2,R}) \perp\!\!\!\perp \Sigma_{S_2}. \tag{4}$$

The independency result in (4) allow us to define the sampling scheme for $p(\Sigma_{P_2} \mid \Sigma_{S_2})$ as follows:

1. Sample $K \sim W_G(b, D_{P_2})$ using the rejection sampling of Section 2 providing values to the submatrices $K_R$ and $K_{S_2,R}$;
2. Compute $\Sigma_{R\cdot S_2} = K_R^{-1}$ and $\Gamma_{R\cdot S_2} = -\Sigma_{R\cdot S_2}K_{R,S_2}$;
3. Compute the implied values of $\Sigma_R = \Sigma_{R\cdot S_2} + \Sigma_{R,S_2}\Sigma^{-1}_{S_2}\Sigma_{S_2,R}$ and $\Sigma_{R,S_2} = \Gamma_{R\cdot S_2}\Sigma_{S_2}$ which combined with $\Sigma_{S_2}$ create a sample of $\Sigma_{P_2}$.

With samples from $\Sigma_{P_1}$, $\Sigma_{S_2}$ and $\Sigma_{P_2}$ available, the remaining elements of $\Sigma$ where $(i,j) \notin E$ can be computed via the completion operation derived in [5] and presented in equation (6) of [3].

## 4. A simulated example

We consider one simulated example that involves a two prime component non-decomposable graph $G$ in Figure 1. The simulation method was applied to generate 5000 samples from the hyper-inverse Wishart distribution $\mathrm{HIW}_G(203, D)$ where

$$D = \begin{pmatrix} 35.93 & 0.73 & 4.68 & 1.77 & 0.87 & 4.35 & 6.20 \\ 0.73 & 30.88 & 4.47 & 1.87 & -0.39 & 2.30 & 2.05 \\ 4.68 & 4.47 & 19.31 & 2.60 & -0.89 & 0.29 & 1.57 \\ 1.77 & 1.87 & 2.60 & 14.78 & 1.58 & 0.31 & 0.14 \\ 0.87 & -0.39 & -0.89 & 1.58 & 18.03 & 2.91 & 1.48 \\ 4.35 & 2.30 & 0.29 & 0.31 & 2.91 & 9.85 & 6.21 \\ 6.20 & 2.05 & 1.57 & 0.14 & 1.48 & 6.21 & 9.55 \end{pmatrix}.$$

To demonstrate the efficacy of the our sampler, we first compute the theoretically exact value of $\mathrm{E}(\Sigma^E \mid D, d, G)$, where $\Sigma^E$ denotes the free elements of $\Sigma$.

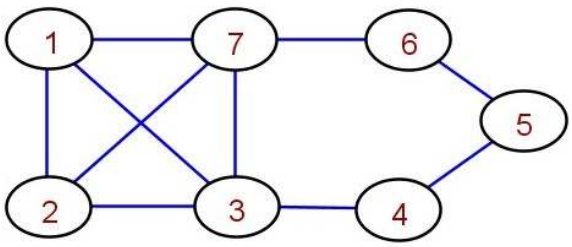


FIG 1. *The underlying non-decomposable graph in the simulated example. The prime components and separators are* $P_1 = \{1,2,3,7\}, S_2 = \{3,7\}$, *and* $P_2 = \{3,4,5,6,7\}$.

The Corollary 2 of [8] implies that this expectation can be calculated as

$$\mathrm{E}(\Sigma^E \mid D, d, G) = D^E/(d-2).$$

The theoretically exact value and the Monte Carlo estimate based the sample mean of the 5000 simulated covariance matrices are

$$\begin{pmatrix} 0.1787 & 0.0036 & 0.0233 & \cdot & \cdot & \cdot & 0.0309 \\ 0.0036 & 0.1536 & 0.0223 & \cdot & \cdot & \cdot & 0.0102 \\ 0.0233 & 0.0223 & 0.0961 & 0.0130 & \cdot & \cdot & 0.0078 \\ \cdot & \cdot & 0.0130 & 0.0735 & 0.0078 & \cdot & \cdot \\ \cdot & \cdot & \cdot & 0.0078 & 0.0897 & 0.0145 & \cdot \\ \cdot & \cdot & \cdot & \cdot & 0.0145 & 0.0490 & 0.0309 \\ 0.0309 & 0.0102 & 0.0078 & \cdot & \cdot & 0.0309 & 0.0475 \end{pmatrix}$$

and

$$\begin{pmatrix} 0.1790 & 0.0038 & 0.0234 & \cdot & \cdot & \cdot & 0.0308 \\ 0.0038 & 0.1536 & 0.0222 & \cdot & \cdot & \cdot & 0.0101 \\ 0.0234 & 0.0222 & 0.0962 & 0.0130 & \cdot & \cdot & 0.0078 \\ \cdot & \cdot & 0.0130 & 0.0737 & 0.0078 & \cdot & \cdot \\ \cdot & \cdot & \cdot & 0.0078 & 0.0897 & 0.0146 & \cdot \\ \cdot & \cdot & \cdot & \cdot & 0.0146 & 0.0490 & 0.0309 \\ 0.0308 & 0.0101 & 0.0078 & \cdot & \cdot & 0.0309 & 0.0475 \end{pmatrix},$$

respectively, where $\cdot$ denotes non-free elements to highlight structure.

The mean and median number of samples required to accept one sample in the rejection sampling of Step (ii) in Section 2 is 1.35 and 1 respectively.

## 5. Final remarks

We have identified and fixed two problems involving the sampling of hyper-inverse Wishart random variables conditional on non-decomposable graphical models as proposed by [2] and [3]. Their ideas are still used in our approach and it now clear that direct samples can be obtained in a computationally efficient way by using the junction tree of a graph. Finally, we would like to point the reader to two recently proposed alternative solutions to this problem: [1], [6] and [4]. It is important to highlight, however, that all of these strategies are based on iterative Markov chain Monte Carlo and therefore have to rely on the eventual convergence of the chain. Our method is a direct sampling strategy that, at each step, delivers exact samples from the hyper-inverse Wishart distribution. Upon publication of this note, a R-package for sampling from general hyper-inverse Wishart distributions will be available at the author's website.

## 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). The marginal likelihood for decomposable and non-decomposable graphical Gaussian models. *Biometrika* **92** 317–35. MR2201362

[3] Carvalho, C., Massam, H. and West, M. (2007). Simulation of hyper-inverse Wishart distributions in graphical models. *Biometrika* **94** 647–659. MR2410014

[4] Dobra, A., Lenkoski, A. and Rodriguez, A. (2010). Bayesian inference for general Gaussian graphical models with application to multivariate lattice data Technical Report, University of Washington.

[5] Massam, H. and Neher, E. (1998). Estimation and Testing for Lattice Conditional Independence Models on Euclidean Jordan Algebras. *Ann. Statist.* **26** 1051–82. MR1635442

[6] Mitsakakis, N., Massam, H. and Escobar, M. (2010). A Metropolis-Hastings based method for sampling from G-Wishart distribution in Gaussian graphical Models Technical Report, University of Toronto.

[7] Robert, C. and Casella, G. (2010). *Monte Carlo Statistical Methods*, 2 ed. Springer-Verlag, New York. MR2080278

[8] 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