JOURNAL OF MULTIVARIATE ANALYSIS 46, 198–206 (1993)

Bartlett's Decomposition of the Posterior Distribution of the Covariance for Normal Monotone Ignorable Missing Data

CHUANHAI LIU

Harvard University

This paper presents a decomposition for the posterior distribution of the covariance matrix of normal models under a family of prior distributions when missing data are ignorable and monotone. This decomposition is an extension of Bartlett's decomposition of the Wishart distribution to monotone missing data. It is not only theoretically interesting but also practically useful. First, with monotone missing data, it allows more efficient drawing of parameters from the posterior distribution than the factorized likelihood approach. Furthermore, with nonmonotone missing data, it allows for a very efficient monotone date augmentation algorithm and thereby multiple imputation of the missing data needed to create a monotone pattern.

1. INTRODUCTION

Missing data problems occur frequently in practice. An efficient way to deal with missing values in a multivariate data set is to sort the data into a monotone pattern such that it contains all the observed values with "minimum" number of missing values that destroy the monotone pattern (see [9, 12, 14, 16]). A monotone pattern of missing data is depicted in Fig. 1 and can be represented by

$$X_{MP} = \{(x_{i,k}^{(k)}, ..., x_{i,p}^{(k)}) : i = 1, ..., n_k; k = 1, ..., p\}, \tag{1}$$

where $n_k \geqslant 0$ for $k = 1, ..., p$, the superscript indexes the pattern, and $\sum n_k = n$. That is, the data are sorted so that the jth variable is at least as observed as the $(j-1)$th variable for $j = 2, 3, ..., p$.

Received September 28, 1992; revised December 12, 1992.

AMS 1970 classification numbers: primary 62A15, 62J10, 62E99, 62H99; secondary 65C05, 65C10, 65B99.

Key words and phrases: Bartlett's decomposition, monotone data augmentation, monotone pattern, multiple imputation, posterior distribution.

0047-259X/93 $5.00

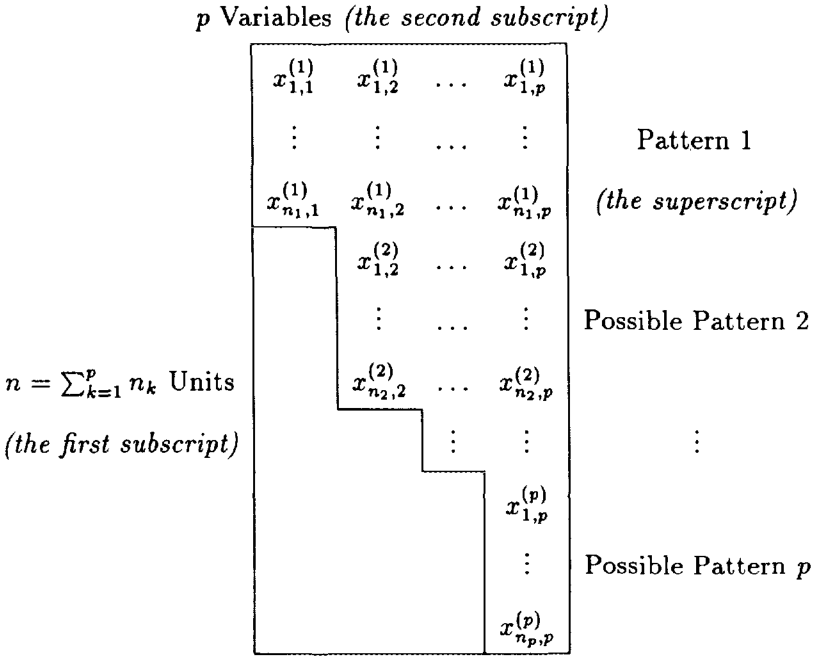

FIG. 1. Data sorted into a monotone pattern.

Under the assumptions that the missing-data mechanism is ignorable (see [9, 13, 14]) and that the observations of the p variables for the n units are independently identically distributed with a normal distribution $N_p(\mu, \Psi)$, the model for X_{MP} in (1) is

$$(x_{i,k}^{(k)}, ..., x_{i,p}^{(k)}) \mid (\mu, \Psi) \sim N_k(\mu^{(k)}, \Psi^{(k)}), \tag{2}$$

where $\mu = (\mu_1, ..., \mu_p)'$, $\Psi_{(p \times p)} > 0$, $\mu^{(k)} = (\mu_k, ..., \mu_p)'$, and $\Psi^{(k)}$ is the lower right $(p-k+1) \times (p-k+1)$ submatrix of Ψ. In Bayesian inference, a common prior distribution for μ and Ψ is Jeffrey's prior distribution [2], $pr(\mu, \Psi) \propto |\Psi|^{-(p+1)/2}$. This paper considers the following prior distribution of μ and Ψ, (see [3]),

$$pr(\mu, \Psi) \propto |\Psi|^{-(q+1)/2}, \tag{3}$$

where q is a integer. When $q = p$, (3) becomes the noninformative prior or Jeffrey's prior, and when $q = -1$, it is the flat prior, i.e., $pr(\mu, \Psi) \propto c$, where c is a constant. The result of this paper can be applied to the case with the inverse Wishart prior (see Section 2).

The posterior distribution of (μ, Ψ) is fundamental in Bayesian inference. If $n_2 = \cdots = n_p = 0$, that is, for the complete data case, the marginal posterior distribution of Ψ^{-1} is the Wishart distribution with $n-1+(q-p)$ degrees of freedom, the marginal distribution of μ is a multivariate t-distribution, and the conditional distribution of μ given Ψ is normal. Bartlett's (see [1, 6]) decomposition of the Wishart distribution can thus be used to derive quantities of interest such as posterior means, variances, and distributions that are relevant to Ψ; it also can be used to simulate Ψ. For example, in the P-step implementing the Data Augmentation (DA) algorithm of Tanner and Wong [19], the crucial step of DA is to draw Ψ from its posterior distribution, the inverse Wishart. To generate inverse Wishart deviates, Jones [8] adapted Smith and Hocking's [18] computer program, which is developed from Odell and Feiveson's [11] algorithm. Odell and Feiveson's algorithm basically is Barlett's [1] (see also [6]) decomposition although they credit it to Hartley and Harris [4] (see also [7]).

Instead of imputing all the missing values in the rectangular data matrix in the I-step of DA, Rubin and Schafer [16] proposed the monotone data augmentation (MDA) algorithm that only imputes those missing values destroying a monotone pattern. They implemented MDA by using factorized likelihood approach [14]. A brief review of DA and MDA can be found in Section 3 of this paper. For more details, please see Schafer [17].

A new approach to implementing MDA is presented in this paper based on a decomposition of the posterior distribution of Ψ^{-1} for the monotone data set (1) and the conditional posterior distribution of μ given Ψ in Section 2, which is similar to Bartlett's decomposition. Compared with the approach of Rubin [14] (see also [16, 17]), the new approach has the following advantages: (1) convenience with Jeffrey's prior and the flat prior that are widely preferred in statistics; (2) direct simulation of μ and Ψ; in other words, it does not have to use reparametrization of Rubin and Schafer [16] (see also [17]); (3) efficiency and simplicity of the resultant statistical algorithm [10] for imputing monotone missing data and thereby for the MDA. Details are given in Section 3.

2. Theoretical Results

A decomposition of the marginal posterior distribution of Ψ for monotone data (1) is presented in Theorem 1, follows by Corollary 1, a decomposition of the conditional posterior distribution of Ψ given μ. Other useful results are listed in Corollaries 2 and 3.

Theorem 1. *For $1 \leqslant k \leqslant p$, let $\bar{\mathbf{y}}_k$ be the $(p-k+1)$-dimensional sample mean of $\{(x_{i,k}^{(j)}, ..., x_{i,p}^{(j)}) : i = 1, ..., n_j; j = 1, ..., k\}$ (i.e., the completely observed* *unit by variable matrix for patterns* $\{1, ..., k\}$ *in Fig.* 1) *and* $\mathbf{S}_k$ $(>\mathbf{0})$ *be the corresponding total sum of squares and cross products matrix about the sample mean* $\bar{\mathbf{y}}_k$ *with the Cholesky factorization* $\mathbf{S}_k^{-1} = \mathbf{L}_k \mathbf{L}_k'$, *where* $\mathbf{L}_k$ *is lower triangular. Let* $\mathbf{H}$ *be a lower triangular* $p \times p$ *matrix with its lower triangular part formed by columns* $\mathbf{L}_1\mathbf{t}_1, \mathbf{L}_2\mathbf{t}_2, ..., \mathbf{L}_p\mathbf{t}_p$, *where* $\mathbf{t}_k = (t_{k,k}, ..., t_{p,k})'$, *with* $\mathbf{t}_1, ..., \mathbf{t}_p$ *satisfying*

(a) $t_{i,j}$ *are independent for* $1 \leqslant j \leqslant i \leqslant p$;

(b) $t_{i,j} \sim N(0, 1)$ *for* $1 \leqslant j < i \leqslant p$;

(c) $t_{j,j}^2 \sim \chi^2_{n_1 + n_2 + \cdots + n_j - j + (q-p)}$ *for* $j = 1, ..., p$.

If $n_1 + \cdots + n_k > k + (p - q)$ *for* $k = 1, ..., p$, *then the marginal posterior distribution of* Ψ^{-1} *determined by* (1), (2) *and* (3) *is distributed as* $\mathbf{HH}'$.

To prove Theorem 1, we introduce the following Lemma 1, which can be proved easily by algebraical calculations.

LEMMA 1. *Let* $\Psi^{-1} = \mathbf{HH}'$ *with* $\mathbf{H}$ *lower triangular; then* $(\Psi^{(k)})^{-1} = \mathbf{H}^{(k)}(\mathbf{H}^{(k)})'$, *where* $\mathbf{H}^{(k)}$ *is lower right* $(p-k+1) \times (p-k+1)$ *submatrix of* $\mathbf{H}$.

Proof of Theorem 1. From (1), (2), and (3), the posterior distribution of μ and Ψ is

$$pr(\mu, \Psi \mid X_{MP}) \propto |\Psi|^{-(q+1)/2} \exp\left\{-\tfrac{1}{2} \sum_{k=1}^{p} tr[(\Psi^{(k)})^{-1} \mathbf{R}^{(k)}]\right\} \times \prod_{k=1}^{p} |\Psi^{(k)}|^{-n_k/2} (d\Psi\, d\mu),$$

where $\mathbf{R}^{(k)}$ is the total sum of squares and cross products matrix of samples $\{(x_{i,k}^{(k)}, ..., x_{i,p}^{(k)}) : i = 1, ..., n_k\}$ about $(\mu_k, ..., \mu_p)$. The Jacobian of transformation from Ψ to Ψ^{-1}, $J_{(\Psi \to \Psi^{-1})} = |\Psi|^{(p+1)}$, leads to

$$pr(\mu, \Psi^{-1} \mid X_{MP}) \propto |\Psi|^{(2p-q+1)/2} \exp\left\{-\tfrac{1}{2} \sum_{k=1}^{p} \mathrm{tr}[(\Psi^{(k)})^{-1} \mathbf{R}^{(k)}]\right\} \times \prod_{k=1}^{p} |\Psi^{(k)}|^{-n_k/2} (d\Psi^{-1}\, d\mu).$$

Let $\Psi^{-1} = \mathbf{HH}'$ with $\mathbf{H} = (h_{i,j})$ lower triangular; then $J_{(\Psi^{-1} \to \mathbf{H})} = 2^p h_{1,1}^p h_{2,2}^{p-1} \cdots h_{p,p}$. Applying Lemma 1 to the lower triangular decomposition of Ψ^{-1}, we have

$$pr(\mu, \mathbf{H} \mid X_{MP}) \propto \exp\left\{-\tfrac{1}{2} \sum_{k=1}^{p} \mathbf{h}_k' \mathbf{R}_k \mathbf{h}_k\right\} \prod_{k=1}^{p} h_{k,k}^{(\sum_{j=1}^{k} n_j - k + q - p)} (d\mathbf{H}\, d\mu), \qquad (4)$$

where $\mathbf{h}_k = (h_{k,k}, ..., h_{p,k})'$ for $k = 1, ..., p$; and $\mathbf{R}_k$ is the total sum of squares and cross products matrix of samples $\{(x_{i,k}^{(j)}, ..., x_{i,p}^{(j)}) : i = 1, ..., n_j; j = 1, ..., k\}$ about the mean $(\mu_k, ..., \mu_p)$. With Lemma 1 and some algebraic calculations, (4) can be factorized into

$$pr(\mu, \mathbf{H} \mid X_{MP}) \propto \exp\left\{-\tfrac{1}{2} \sum_{k=1}^{p} \mathbf{h}_k' \mathbf{S}_k \mathbf{h}_k\right\} \prod_{k=1}^{p} h_{k,k}^{(\sum_{j=1}^{k} n_j - k + q - p)} \times \exp\{-\tfrac{1}{2}(\mu - \theta)' \mathbf{D}^{-1}(\mu - \theta)\}(d\mathbf{H}\, d\mu), \tag{5}$$

where $\mathbf{D} = (\mathbf{H} \Lambda \mathbf{H}')^{-1}$ with $\Lambda = \mathrm{diag}\{n_1, n_2 + n_2, ..., n_1 + n_2 + \cdots + n_p\}$; and $\theta = (\mathbf{H}')^{-1} (\mathbf{h}_1' \bar{\mathbf{y}}_1, ..., \mathbf{h}_p' \bar{\mathbf{y}}_p)'$. Thus, $pr(\mu \mid \Psi; X_{MP})$ is normal with mean θ and covariance $\mathbf{D}$, and then

$$pr(\mathbf{H} \mid X_{MP}) \propto \exp\left\{-\tfrac{1}{2} \sum_{k=1}^{p} \mathbf{h}_k' \mathbf{S}_k \mathbf{h}_k\right\} \prod_{k=1}^{p} h_{k,k}^{(\sum_{j=1}^{k} n_j - k - 1 + q - p)}(d\mathbf{H}). \tag{6}$$

Finally because $J_{(\mathbf{H} \to (\mathbf{t}_1, ..., \mathbf{t}_p))} = \prod |\mathbf{L}_k|$, we get

$$P(\mathbf{t}_1, ..., \mathbf{t}_p \mid X_{MP}) \propto \exp\left\{-\tfrac{1}{2} \prod_{k=1}^{p} \mathbf{t}_k' \mathbf{t}_k\right\} \prod_{k=1}^{p} t_{k,k}^{(\sum_{j=1}^{k} n_j - k - 1 + q - p)}(d(\mathbf{t}_1, ..., \mathbf{t}_p))$$

and Theorem 1 is thus proved. ▮

Theorem 1 can be adapted when one uses the inverse Wishart prior

$$pr(\mu, \Psi) \propto |\Psi|^{-(q+1)/2} \exp\{-\tfrac{1}{2} \operatorname{tr} \Psi^{-1}\mathbf{W}\}, \tag{7}$$

where $\mathbf{W}$ is a known $p \times p$ non-negative matrix.

With $\mathbf{S}_k$ in Theorem 1 replaced by $\mathbf{S}_k + \mathbf{W}_k$, where $\mathbf{W}_k$ is the lower right $(p-k+1) \times (p-k+1)$ submatrix of $\mathbf{W}$, one can show that Theorem 1 is still valid by inserting $-\frac{1}{2} \operatorname{tr} \Psi^{-1}\mathbf{W}$ in the corresponding exponential parts of the densities in the proof of Theorem 1.

Theorem 1 implies that q in (3) needs to satisfy the condition $n_1 + \cdots + n_k > k + (p - q)$ for $k = 1, ..., p$. If this condition is not met, there is no proper posterior distribution of Ψ. In MDA, the missing values destroying a monotone pattern might result in an improper posterior distribution of Ψ even if this condition is satisfied. However, if this condition is met and the sum of squares and cross products matrix of the complete samples in pattern 1 is positive, then Ψ has a proper posterior distribution. Otherwise, it is necessary to use the inverse Wishart prior (7), instead of (3), with an appropriate $\mathbf{W}$.

Theorem 1 provides a straightforward algorithm generating Ψ from its posterior distribution. (1) Independently create the standard normal deviates $t_{i,j}$ for $1 \leqslant j < i \leqslant p$ and $t_{i,j} \sim \chi_{n_1 + n_2 + \cdots + n_j - j + (q-p)}$ for $j = 1, ..., p$;

(2) form the lower triangular matrix $\mathbf{H} = (\mathbf{L}_1 \mathbf{t}_1, ..., \mathbf{L}_p \mathbf{t}_p)$; (3) $\Psi = (\mathbf{HH}')^{-1}$ is then a random draw from the posterior distribution of Ψ with prior (3).

If $n_{j+1} = n_{j+2} = \cdots = n_{j+k} = 0$ for some j and k, where $1 \leq j < j+k \leq p$, $\mathbf{S}_{j+l}$ $(1 \leq l \leq k)$ is then the lower right $(p-j-l+1) \times (p-j-l+1)$ submatrix of $\mathbf{S}_j$. From Lemma 1 $\mathbf{L}_{j+l}$ is the lower right $(p-j-l+1) \times (p-j-l+1)$ submatrix of $\mathbf{L}_j$. In this case, Theorem 1 allows a more efficient algorithm drawing Ψ. When $n_2 = n_3 = \cdots = n_p = 0$, Theorem 1 thus becomes Bartlett's decomposition [1, 6].

BARTLETT'S DECOMPOSITION. *Let $\bar{\mathbf{y}}$ be the sample mean of the complete samples and $\mathbf{S}$ $(>\mathbf{0})$ be the corresponding total sum of squares and cross products matrix about the sample mean $\bar{\mathbf{y}}$ with the Cholesky factorization $\mathbf{S}^{-1} = \mathbf{LL}'$, where $\mathbf{L}$ is lower triangular. Let $\mathbf{T}$ be a lower triangular $p \times p$ matrix with entries $t_{i,j}$ satisfying*

(a) *$t_{i,j}$ are independent for $1 \leq j \leq i \leq p$;*

(b) *$t_{i,j} \sim N(0, 1)$ for $1 \leq j < i \leq p$;*

(c) *$t_{j,j}^2 \sim \chi^2_{n-j+(q-p)}$ for $j = 1, ..., p$.*

If $n > p + (p-q)$, then the marginal posterior distribution of Ψ^{-1} given the complete data and prior (3) is distributed as $\mathbf{LTT}'\mathbf{L}'$.

The following results follow from Theorem 1 and its proof.

COROLLARY 1. *For $1 \leq k \leq p$, denote $\mathbf{R}_k$ $(>\mathbf{0})$ to be the total sum of squares and cross products matrix of the $(p-k+1)$-dimensional samples $\{(x_{i,k}^{(j)}, ..., x_{i,p}^{(j)}): i = 1, ..., n_j; j = 1, ..., k\}$ about the mean $(\mu_k, ..., \mu_p)'$ with the Cholesky factorization of $\mathbf{R}_k^{-1}$: $\mathbf{R}_k^{-1} = \mathbf{L}_k \mathbf{L}_k'$, where $\mathbf{L}_k$ is lower triangular. Let $\mathbf{H}$ be a lower triangular matrix defined as in Theorem 1, except that the condition (c) is replaced by*

(c′) *$t_{j,j}^2 \sim \chi^2_{n_1+n_2+\cdots+n_j-j+(q-p)+1}$ for $j = 1, ..., p$.*

If $n_1 + \cdots + n_k > k + (p-q)$ for $k = 1, ..., p$, then the conditional posterior distribution of Ψ^{-1} given μ and X_{MP} is distributed as $\mathbf{HH}'$.

COROLLARY 2. *The conditional posterior distribution of μ given Ψ is normal with mean $\theta = (\mathbf{H}')^{-1} (\mathbf{h}_1' \bar{\mathbf{y}}_1, ..., \mathbf{h}_p' \bar{\mathbf{y}}_p)'$ and covariance $\mathbf{D} = (\mathbf{H}\Lambda\mathbf{H}')^{-1}$, where $\mathbf{HH}' = \Psi^{-1}$, $\Lambda = \operatorname{diag}\{n_1, n_2 + n_2, ..., n_1 + n_2 + \cdots + n_p\}$, and $\bar{\mathbf{y}}_k$ is defined in Theorem 1 for $k = 1, ..., p$.*

Let $Z = (z_1, ..., z_p)' \sim N(0, I)$; then the conditional posterior distribution of μ given Ψ is distributed as $(\mathbf{H}^{-1})' (\mathbf{h}_1' \bar{\mathbf{y}}_1 + z_1/N_1^{1/2}, ..., \mathbf{h}_p' \bar{\mathbf{y}}_p + z_p/N_p^{1/2})$, where $N_k = n_1 + \cdots + n_k$ for $k = 1, ..., p$. This expression allows a simple drawing of μ given $\mathbf{H}$ in simulations of μ, where $\mathbf{HH}' = \Psi^{-1}$.

COROLLARY 3. *The posterior distribution of* $|\Psi|$ *is proportional to the product of* p *independent inverse* χ^2*'s (see* [2]*) with degrees of freedom* $(n_1+q-p-1), ..., (n_1+\cdots+n_p+q-2p)$, *respectively.*

Note that the distinct prior specification of Rubin [14] (see also [16]) has the same property as Corollary 3.

Proofs of the corollaries. Corollary 1 follows the comparison of (4) and (6). Corollary 2 has been proved in (5). Corollary 3 follows the fact that $|\Psi| = \prod l_{k,k}^{-2} \prod t_{k,k}^{-2}$, where $l_{k,k} = (\mathbf{L}_k)_{1,1}$ for $k = 1, ..., p$.

3. APPLICATION TO MONOTONE DATA AUGMENTATION

A useful technique for inference with missing values in multivariate data sets is multiple imputation [14], which can be implemented using the data augmentation (DA) of Tanner and Wong [19] or the monotone data augmentation of Rubin and Schafer [16] (see also [17]); both employ iterative simulation techniques. Each iteration in either of DA or MDA proceeds in two steps: (1) I-step: missing data imputation; and (2) P-step: posterior simulation. In each step of the iteration, DA fills in all the missing values in the rectangular data base while MDA fills in only the missing values that destroy the monotone pattern. MDA can be quite efficient because the number of missing values contained in a resorted monotone pattern can be small and thus iterative simulations for the monotone data augmentation converge much faster than those for the data augmentation. Moreover, the random variates drawn by MDA are more efficient for estimation using Monte Carlo technique than those drawn by DA because their sequential correlation is less than those drawn by DA. If there are no missing values in X_{MP}, MDA does not require any iterative sampling.

In MDA, the I-step, the imputation of the missing data that destroy the monotone pattern, uses the same algorithm as that of DA, which is typically implemented in terms of the regression method via the Gaussian sweep operator. The P-step in MDA is more difficult than that in DA. By using the factorization of the likelihood [14], Rubin and Schafer [16, 17] reparameterize the parameters μ and Ψ by $\Phi = \Phi(\mu, \Psi)$, where Φ is a 1–1 mapping. The transformations from (μ, Ψ) to Φ and from Φ to (μ, Ψ) are conducted in terms of the Gaussian sweep operator. With distinct specification of the prior distribution of Φ, Φ are also a posterior distinct (see [14]) and can be generated via the Gaussian sweep operator. For Jeffrey's prior, the factorized likelihood approach needs adjustment such as by SIR of Rubin of [15] or the generalized Metropolis algorithm of Hastings [5]. Although the algorithm for the SIR or Metropolis is simple, the rejection steps in SIR or Metropolis are time consuming. For more details, please see Rubin [14], Schafer [17], and Rubin and Schafer [16].

A more efficient and straightforward approach to drawing μ and Ψ is to use Theorem 1 and Corollary 2 in Section 2; i.e., first generate Ψ with Theorem 1 and then draw μ based on Corollary 2. The detailed statistical algorithm is given in Liu [10]. This method is quite efficient for large data sets with missing values because each drawing of μ and Ψ needs only $p(p+1)/2$ independent $N(0, 1)$ deviates, p independent χ random draws, and about $p^4/12$ multiplications, whereas the distinct factorized likelihood approach requires the same number of random draws but $13p^4/24$ multiplications (about six times as many as the Bartlett approach).

ACKNOWLEDGMENTS

I am indebted to my adviser Professor Donald B. Rubin for his extensive and stimulating comments as well as for his helpful suggestions for writing this paper. I thank Professor Arthur P. Dempster, Professor Hal S. Stern, and a referee for their useful comments. This research was supported by NSF Grant SES-88-05433.

REFERENCES

[1] BARTLETT, M. S. (1933). On the theory of statistical regression. *Proc. Roy. Soc. Edinburgh* **53** 260–283.

[2] BOX, G. E. P., AND TIAO, G. C. (1973). *Bayesian Inference in Statistical Analysis.* Addison-Wesley, Reading, MA.

[3] DEMPSTER, A. P. (1963). On a paradox concerning inference about a covariance matrix. *Ann. Math. Statist.* **34** 1414–1418.

[4] HARTLEY, H. O., AND HARRIS, D. L. (1963). Monte Carlo computations in normal correlation problems. *J. Assoc. Comput. Mach.* **10** 301–306.

[5] HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109.

[6] KSHIRSAGAR, A. M. (1959). Bartlett decomposition and Wishart distribution. *Ann. Math. Statist.* **30** 239–241.

[7] Johnson, M. E. (1987). *Multivariate Statistical Simulation.* Wiley, New York.

[8] JONES, M. C. (1985). Generating inverse Wishart matrices. *Comm. Statist., Simulation Comput.* **14** 511–514.

[9] LITTLE, R. J. A., AND RUBIN, D. B. (1987). *Statistical Analysis with Missing Data.* Wiley, New York.

[10] LIU, C. (1992). Statistical Algorithm: Efficiently drawing the posterior mean and covariance from normal monotone missing data. Technical Report, Department of Statistics, Harvard University, Cambridge, MA 02138, U.S.A.

[11] ODELL, P. L., AND FEIVESON, A. H. (1966). A numerical procedure to generate a sample covariance matrix. *J. Amer. Statist. Assoc.* **61** 199–203.

[12] RUBIN, D. B. (1974). Characterizing the estimation of parameters in incomplete data problems. *J. Amer. Statist. Assoc.* **69** 467–474.

[13] RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63** 581–592.

[14] RUBIN, D. B. (1987a). *Multiple Imputation for Nonresponse in Surveys.* Wiley, New York.

[15] RUBIN, D. B. (1987b). Using the SIR algorithm to simulate posterior distributions. In *Bayesian Statistics* (J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith Eds.), Vol. 3, pp. 395–402, Oxford Univ. Press, London.

[16] RUBIN, D. B., AND SCHAFER, J. L. (1990). Efficiently creating multiple imputations for incomplete multivariate normal data. In *Proceedings of Statistical Computing Section of the American Statistical Association.*

[17] SCHAFER, J. L. (in preparation). *Analysis by Simulation of Incomplete Multivariate Data: Algorithms and Examples.* Chapman & Hall, New York.

[18] SMITH, W. B., AND HOCKING, R. R. (1972). Wishart variate generation. *Appl. Statist.* **21** 341–345.

[19] TANNER, M. A., AND WONG, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *J. Amer. Statist. Assoc.* **82** 528–550.