Statistics and Probability Letters 104 (2015) 146–152

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

An efficient monotone data augmentation algorithm for Bayesian analysis of incomplete longitudinal data

Yongqiang Tang

Shire, 300 Shire Way, Lexington, MA 02421, USA

ARTICLE INFO

Article history:
Received 12 December 2014
Accepted 21 May 2015
Available online 30 May 2015

Keywords:
Monotone data augmentation
Multivariate normal-gamma distribution
Wishart distribution
MCMC

ABSTRACT

We introduce a new method for sampling from the Wishart distribution by representing the Wishart distributed random matrix as a function of independent multivariate normal-gamma random vectors. An efficient monotone data augmentation (MDA) algorithm is developed for Bayesian multivariate linear regression. For longitudinal outcomes, the proposed method is easier to implement and interpret than that based on Bartlett's decomposition. The proposed algorithm is illustrated by the analysis of an antidepressant trial.

1. Introduction

The data augmentation (DA) algorithm developed by Tanner and Wong (1987) has been widely used to handle missing data in Bayesian inference. The method is useful if the posterior distribution $f(\phi|Y_o)$ of the parameters ϕ given the observed data Y_o is difficult to work with, but it is easy to sample from the augmented data posterior $f(\phi|Y_o, Y_m)$, where Y_m denotes the unobserved data or latent variables. This DA algorithm iterates between an imputation I-step, in which the missing data Y_m are imputed given the current draw of the parameters ϕ, and a posterior P-step, in which the parameters ϕ are generated from $f(\phi|Y_o, Y_m)$ given the current imputed data Y_m.

There are two strategies to implement the DA for incomplete longitudinal data. The monotone data augmentation (MDA) approach imputes only intermittent missing data (i.e. the missing value is followed by an observed value) that destroy monotone missing data pattern while the full DA approach imputes all missing data in the I-step. The MDA algorithm is superior in the sense that it imputes fewer missing values in each iteration, and converges faster with smaller autocorrelation between posterior samples (Schafer, 1997; Liu, 1995). The MDA method is attractive in real applications since the amount of intermediate missing data is generally small in a typical study. The MDA algorithms have been developed for multivariate normal distribution (Schafer, 1997; Liu, 1993) and multivariate linear regression (Liu, 1996). Liu (1995, 1996) also considered MDA for robust inference based on multivariate t or other non-normal distributions.

This note shows that the Wishart distributed random matrix can be represented as a function of independent multivariate normal-gamma variables, and this suggests an alternative way for sampling from the Wishart distribution. We propose a MDA algorithm for multivariate linear regression based on the new decomposition of a Wishart matrix. Unlike the MDA algorithms developed by Schafer (1997) and Liu (1993, 1995, 1996), our method allows the use of informative prior on both the covariance matrix and regression coefficients. Although we use more complex prior distribution, the corresponding posterior distribution has simpler expression than that given by Liu (1996). Liu (1993, 1995, 1996) approaches rely on Bartlett's decomposition, and require organizing the longitudinal data in a time-reverse order. Compared to Liu's work, our method is easier to interpret and implement.

E-mail address: yongqiang_tang@yahoo.com.

http://dx.doi.org/10.1016/j.spl.2015.05.014

The rest of the paper is organized as follows. Section 2 introduces a new method for sampling from the Wishart distribution. The MDA algorithm for Bayesian multivariate linear regression is presented in Section 3. Section 4 illustrates an application of the proposed method to an antidepressant trial.

2. A new method for generating Wishart random matrices

In Bayesian statistics, the Wishart distribution is often used as a conjugate prior for the inverse of a multivariate normal covariance matrix. This section shows that a p-dimensional Wishart matrix can be written as a function of p independent multivariate normal-gamma random vectors, and describes a new method for generating the Wishart random matrix.

Let Σ be a $p \times p$ random symmetric positive definite matrix. Let A be a $p \times p$ fixed positive definite matrix. Then Σ follows an inverse Wishart distribution $\Sigma \sim \mathcal{W}^{-1}(A, n_0)$ if the probability density function (pdf) of Σ is given by

$$f(\Sigma) \propto |\Sigma|^{-\frac{n_0+p+1}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}(A\Sigma^{-1})\right],$$

or equivalently $\Omega = \Sigma^{-1}$ has a Wishart distribution $\Omega \sim \mathcal{W}(A^{-1}, n_0)$

$$f(\Omega) \propto |\Omega|^{\frac{n_0-p-1}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}(A\Omega)\right].$$

Let $\Omega = H'H$ and $A = L^{-1}(L^{-1})'$, where $H = (h_{ij})$ and $L = (l_{ij})$ are low triangular matrices. Let the first $k \times k$ leading submatrix of A and L be denoted respectively by A_k and L_k. Then $A_k = L_k^{-1}(L_k^{-1})'$. Let $\boldsymbol{h}_k = (h_{k1}, \ldots, h_{kk})'$. Appendix A.1 shows that the distribution of H is given by

$$f(\boldsymbol{h}_1, \ldots, \boldsymbol{h}_p) \propto \prod_{k=1}^{p} h_{kk}^{n_0-p+k-1} \exp\left(-\frac{1}{2}\boldsymbol{h}_k' A_k \boldsymbol{h}_k\right). \tag{1}$$

Thus $\boldsymbol{h}_k$'s are independent. The following lemma shows an efficient way to generate $\boldsymbol{h}_k$'s and $\Omega = H'H$. The proofs of all lemmas in this note will be deferred to Appendix A.1.

Lemma 1. (a) *Let $\boldsymbol{t} = (t_1, \ldots, t_m)'$ and $D = C^{-1}(C^{-1})'$, where C is an $m \times m$ lower triangular matrix. If $t_j \sim N(0, 1)$ for $j < m$, $t_m^2 \sim \chi^2_{e+1}$, and t_j's are independent, the distribution of $\boldsymbol{h} = (h_1, \ldots, h_m)' = C'\boldsymbol{t}$ is given by*

$$f(\boldsymbol{h}) \propto h_m^e \exp\left[-\frac{1}{2}\boldsymbol{h}'D\boldsymbol{h}\right]. \tag{2}$$

(b) *Let $\gamma = h_m^2$, $\beta_j = -h_j/h_m$ for $j < m$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{m-1})'$, and $\tilde{\boldsymbol{\beta}} = (-\boldsymbol{\beta}', 1)'$. Let $D = \begin{bmatrix} D_{m-1} & \boldsymbol{d}_m \\ \boldsymbol{d}_m' & d_{mm} \end{bmatrix}$ and $C = \begin{bmatrix} C_{m-1} & \boldsymbol{0} \\ \boldsymbol{c}_m' & c_{mm} \end{bmatrix}$. If $\boldsymbol{h} = (h_1, \ldots, h_m)' = h_m\tilde{\boldsymbol{\beta}}$ is distributed as (2), the joint distribution of $(\boldsymbol{\beta}, \gamma)$ is given by*

$$f(\boldsymbol{\beta}, \gamma) \propto \gamma^{\frac{e+m-2}{2}} \exp\left[-\frac{\gamma}{2}\tilde{\boldsymbol{\beta}}'D\tilde{\boldsymbol{\beta}}\right]. \tag{3}$$

The marginal distribution of γ is $\gamma \sim s^{-1}\chi^2_{e+1}$, and the conditional distribution of $\boldsymbol{\beta}$ given γ is normal with mean $D_{m-1}^{-1}\boldsymbol{d}_m = -c_{mm}^{-1}\boldsymbol{c}_m$, and variance $(\gamma D_{m-1})^{-1} = \gamma^{-1}C_{m-1}'C_{m-1}$, where $s = d_{mm} - \boldsymbol{d}_m'D_{m-1}^{-1}\boldsymbol{d}_m = c_{mm}^{-2}$.

(c) *Let $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$ denote respectively the first w and last $m-1-w$ elements of $\boldsymbol{\beta}$, and $\tilde{\boldsymbol{\beta}}^{(2)} = (-\boldsymbol{\beta}^{(2)'}, 1)'$. Suppose D can be partitioned as $D = \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}$, where the size of D_{11} is $w \times w$. If the distribution of $(\boldsymbol{\beta}, \gamma)$ is given by (3), the marginal distribution of $(\boldsymbol{\beta}^{(2)}, \gamma)$ is*

$$f(\boldsymbol{\beta}^{(2)}, \gamma) \propto \gamma^{\frac{m+e-2-w}{2}} \exp\left[-\frac{\gamma}{2}\tilde{\boldsymbol{\beta}}^{(2)'}(D_{22} - D_{21}D_{11}^{-1}D_{12})\tilde{\boldsymbol{\beta}}^{(2)}\right],$$

and the conditional distribution of $\boldsymbol{\beta}^{(1)}$ given $(\boldsymbol{\beta}^{(2)}, \gamma)$ is $N\left[D_{11}^{-1}D_{12}\tilde{\boldsymbol{\beta}}^{(2)}, (\gamma D_{11})^{-1}\right]$.

Lemma 1(a) implies that $\boldsymbol{h}_k$ can be generated as $\boldsymbol{h}_k = L_k'\boldsymbol{t}_k$, where $\boldsymbol{t}_k = (t_{k1}, \ldots, t_{kk})'$, $t_{kj} \sim N(0, 1)$ for $j < k$, $t_{kk}^2 \sim \chi^2_{n-p+k}$ and t_{ij}'s are independent (when $k = 1$, $\boldsymbol{h}_1 \sim l_{11}\sqrt{\chi^2_{n-p+1}}$ is a scalar). The Wishart matrix can be generated as $\Omega = H'H$, where $T = \begin{bmatrix} t_{11} & 0 & \cdots & 0 \\ t_{21} & t_{22} & 0 & 0 \\ & \cdots & & \\ t_{p1} & t_{p2} & \cdots & t_{pp} \end{bmatrix}$ and $H = TL$.

Lemma 1(b) indicates that $\boldsymbol{h}_k$ can be reparameterized as multivariate normal-gamma random variables, where $\gamma_k = h_{kk}^2$ follows a Gamma or Chi-square distribution, and the conditional distribution of $\boldsymbol{\beta}_k = -(h_{k1}, \ldots, h_{k,k-1})'/h_{kk}$ given γ_k is