Journal of Applied Probability & Statistics Vol. **5**, No. 1, xxx–xxx

On Bayesian Inference with Conjugate Priors for Scale Mixtures of Normal Distributions

Vee Ming Ng

School of Chemical and Mathematical Sciences, Murdoch University, Murdoch 6150, Australia.

Email: V.Ng@murdoch.edu.au

Abstract

Bayesian inference is considered for the multivariate regression model with distribution of the random responses belonging to the multivariate scale mixtures of normal distributions. The posterior distribution of the regression parameters and the predictive distribution of future responses for the model are derived when the prior distribution of the parameters is from the conjugate family and they are shown to be identical to those obtained under normally distributed random responses. This gives inference robustness with respect to departures from the reference case of independent sampling from the normal distribution.

Keywords: Multivariate regression, scale mixtures normal, conjugate prior, posterior and predictive distributions.

2010 Mathematics Subject Classification: 62F15, 62H05.

1 Introduction

Linear regression models with normal and non-normal responses have been considered by several authors. The model with normal responses has been considered by Geisser (1965, 1993), Guttman and Hougaard (1985) and Khan (2004) among others, using a Bayesian analysis with noninformative prior, and Fraser (1979) and Fraser and Haq (1969) using a non-Bayesian approach. One of the early papers on non-normal responses is by Zellner (1976) in which a univariate regression model with multivariate Student-t responses is examined. Since then, there have been numerous papers on the multivariate regression model with random responses assumed to have matrix elliptically contoured distributions, the matrix-t distribution being a special case of this family of distributions. Some examples of univariate and multivariate regression models with non-normal responses are Anderson and Fang (1990), Fraser and Ng (1980), Haq and Khan (1990), Khan (2002) and Ng (2000, 2002). Using noninformative prior in a Bayesian analysis, Ng (2002) shows that the posterior distribution of regression parameters and the predictive distribution of future responses are robust with respect to departures of any member of the matrix elliptically contoured distribution from the normal distribution. In this paper, we consider the case in which prior information is from the conjugate family of distributions and the random responses have multivariate scale mixtures of normal distributions, which is a subclass of the elliptical family. We show that the posterior and predictive distributions are also unaffected by a change in the distribution from normal to this larger class of distributions. These results extend those of Jammalamadaka *et al.* (1987) and Zellner (1976).

In Section 2, the posterior distribution of the regression parameters is derived and in Section 3, the Bayesian predictive distribution is obtained. Some concluding comments are made in Section 4.

2 Posterior Distribution of Regression Parameters

$$Y = \mathcal{B}X + \psi(z)^{-1/2}E, \tag{2.1}$$

where the n columns of the $m \times n$ response matrix Y can be regarded as a sample of size n from a m-dimensional population, X is the $k \times n$ design matrix of known values of rank k, $\mathcal{B}$ is the $m \times k$ matrix of unknown regression parameters, $n > m+k$, and $\psi(z)$ is a positive function of a univariate random variable z, with distribution function $W(z)$ independent of E. The $m \times n$ error component E is assumed to have a matrix variate normal distribution, $N(0, \Phi^{-1} \otimes I_n)$, with mean 0 and covariance matrix $\Phi^{-1} \otimes I_n$ where 0 is a matrix of zeros and $\otimes$ is the Kronecker product of the two matrices Φ^{-1} and I_n; I_n is the $n \times n$ identity matrix and Φ is the $m \times m$ precision matrix of each column of E. Model (2.1) implies that conditionally on z, Y is $N(\mathcal{B}X, \psi^{-1}(z)\Phi^{-1} \otimes I_n)$ while the unconditional density function can be written as

$$f(Y|\mathcal{B}, \Phi) \propto \int f(Y|\mathcal{B}, \Phi, z)dW(z), \tag{2.2}$$

where

$$f(Y|\mathcal{B}, \Phi, z) \propto |\psi(z)\Phi\,|^{n/2}\text{etr}\Big\{-\frac{1}{2}\psi(z)\Phi\,(Y - \mathcal{B}X)(Y - \mathcal{B}X)^T\Big\}. \tag{2.3}$$

The density function above defines a family of scale mixtures of normal distributions which includes the matrix-t, Cauchy and logistic distributions, see Ravishanker *et al.* (2002, pp.182-183). When z has a gamma distribution, $G(v/2, v/2)$, with density function

$$f(z|v) \propto z^{v/2-1}\text{exp}(-vz/2)$$

and $\psi(z) = z$, it can be easily shown that the unconditional distribution of Y is matrix-t with density function given by, Kowalski *et al.* (1999),

$$f(Y|\mathcal{B}, \Phi) \propto |\Phi|^{n/2}(1 + \text{tr}[\Phi(Y - \mathcal{B}X)(Y - \mathcal{B}X)^T/v])^{-(v+mn)/2}.$$

Let $\Gamma = \psi(z)\Phi$ with Jacobian of transformation, $J(\Gamma \to \Phi) = \psi(z)^{m(m+1)/2}$ and assume a normal-Wishart prior for $(\mathcal{B}, \Gamma)$ i.e.,

$$\begin{aligned} \pi(\mathcal{B}|\Gamma) &\propto |\Gamma|^{k/2}\text{etr}\Big\{-\frac{1}{2}\Gamma(\mathcal{B}-\mathcal{B}^*)A(\mathcal{B}-\mathcal{B}^*)^T\Big\} \\ \pi(\Gamma) &\propto |\Gamma|^{(v^*-m-1)/2}\text{etr}\Big\{-\frac{1}{2}D\Gamma\Big\} \end{aligned}$$

with known hyperparameters v^*, $\mathcal{B}^*$ ($m \times k$ matrix), A ($k \times k$ matrix) and D ($m \times m$ matrix). Suppose

$$\begin{aligned} \pi(\mathcal{B}, \Phi|z) &\propto \pi(\mathcal{B}|\Gamma)\pi(\Gamma)|J(\Gamma \to \Phi) \\ &\propto \psi(z)^{m(v^*+k)/2}|\Phi|^{(v^*+k-m-1)/2}\text{etr}\Big\{-\frac{1}{2}\psi(z)\Phi[(\mathcal{B}-\mathcal{B}^*)A(\mathcal{B}-\mathcal{B}^*)^T+D]\Big\}. \end{aligned} \tag{2.4}$$

Extending the result of Jammalamadaka *et al.* (1987), the conjugate prior density for the scale mixtures of normal distributions can be obtained as

$$\pi(\mathcal{B}, \Phi) \propto \int \pi(\mathcal{B}, \Phi|z)dW(z). \tag{2.5}$$

If $\psi(z) = z$ and z has the gamma distribution $G(v/2, v/2)$, then evaluation of (2.5) yields the conjugate prior for the family of matrix-t distributions since $\psi(z) = z$ results in Y having a matrix-t distribution. Evaluation of (2.5) gives the conjugate prior density for $(\mathcal{B}, \Phi)$ as

$$\begin{aligned} \pi(\mathcal{B}, \Phi) \propto\ & |(v+\text{tr}\Phi D)^{-1}\Phi|^{k/2}(1+\text{tr}[(v+\text{tr}\Phi D)^{-1}\Phi(\mathcal{B}-\mathcal{B}^*)A(\mathcal{B}-\mathcal{B}^*)^T])^{-[m(k+v^*)+v]/2} \\ & \times |\Phi|^{(v*-m-1)/2}(v+\text{tr}[\Phi D])^{-(mv^*+v)/2}. \end{aligned}$$

Thus $\pi(\mathcal{B}, \Phi)$ is of the form of a matrix-t density function of $\mathcal{B}$, given Φ, multiplied by a multivariate inverted Beta density (see Johnson *et al.* 1972, p. 236) of Φ. For $m = 1$, the prior reduces to a product of a multivariate Student-t density and a F density as in Jammalamadaka *et al.* (1987) and Zellner (1976).

Using the prior information above, we obtain the following result on the posterior distribution $\mathcal{B}$.

Theorem 2.1. *Using the conjugate prior on $(\mathcal{B}, \Phi)$ given in (2.4), the posterior distribution of the regression parameters $\mathcal{B}$ for the multivariate regression model (2.1) with scale mixtures of normal distributions for Y is matrix-t:*

$$t_{mk}(\mathcal{B}_1, (XX^T + A)^{-1}, R, n + v^* - m + 1),$$

where

$$\mathcal{B}_1 = (YX^T + \mathcal{B}^*A)(XX^T + A)^{-1},$$

$$R = YY^T + \mathcal{B}^* A \mathcal{B}^{*T} - (YX^T + \mathcal{B}^* A)(XX^T + A)^{-1}(YX^T + \mathcal{B}^* A)^T + D.$$

(see Box and Tiao 1973, p. 441-442 for notation).

Proof. To obtain the above result, note that

$$\begin{aligned} f(\mathcal{B}|Y) &\propto \int\int f(Y|\mathcal{B},\Phi,z)\pi(\mathcal{B},\Phi|z)d\Phi dW(z) \\ &\propto \int f(\mathcal{B}|Y,z)dW(z), \end{aligned}$$

where

$$\begin{aligned} f(\mathcal{B}|Y,z) &= \int f(Y|\mathcal{B},\Phi,z)\pi(\mathcal{B},\Phi|z)d\Phi \\ &\propto \int \psi(z)^{m(n+v^*+k)/2}|\Phi|^{(n+v^*+k-m-1)/2} \\ &\quad \times \mathrm{etr}\Big\{-\frac{1}{2}\psi(z)\Phi(Y-\mathcal{B}X)(Y-\mathcal{B}X)^T\Big\} \\ &\quad \times \mathrm{etr}\Big\{-\frac{1}{2}\psi(z)\Phi[(\mathcal{B}-\mathcal{B}^*)A(\mathcal{B}-\mathcal{B}^*)^T + D]\Big\}d\Phi. \end{aligned} \tag{2.6}$$

Combining the quadratic forms in (2.6), we have

$$(Y-\mathcal{B}X)(Y-\mathcal{B}X)^T + (\mathcal{B}-\mathcal{B}^*)A(\mathcal{B}-\mathcal{B}^*)^T + D = (\mathcal{B}-\mathcal{B}_1)(XX^T+A)(\mathcal{B}-\mathcal{B}_1)^T + R.$$

The integral (2.6) becomes

$$\begin{aligned} f(\mathcal{B}|Y,z) \propto \int & \psi(z)^{m(n+v^*+k)/2}|\Phi|^{(n+v^*+k-m-1)/2} \\ & \times \mathrm{etr}\Big\{-\frac{1}{2}\psi(z)[(\mathcal{B}-\mathcal{B}_1)(XX^T+A)(\mathcal{B}-\mathcal{B}_1)^T + R]\Phi\Big\}d\Phi. \end{aligned}$$

Let $U = \psi(z)[(\mathcal{B}-\mathcal{B}_1)(XX^T+A)(\mathcal{B}-\mathcal{B}_1)^T + R]\Phi$ with Jacobian of transformation $|\psi(z)[(\mathcal{B}-\mathcal{B}_1)(XX^T+A)(\mathcal{B}-\mathcal{B}_1)^T + R]|^{-(m+1)/2}$. Hence

$$\begin{aligned} f(\mathcal{B}|Y,z) &\propto \int \psi(z)^{m(n+v^*+k)/2}\Big|[(\mathcal{B}-\mathcal{B}_1)(XX^T+A)(\mathcal{B}-\mathcal{B}_1)^T + R]^{-1} \\ &\quad \times \psi(z)^{-1}U\Big|^{(n+v^*+k-m-1)/2} \\ &\quad \times \mathrm{etr}\Big\{-\frac{1}{2}U\Big\}|\psi(z)[(\mathcal{B}-\mathcal{B}_1)(XX^T+A)(\mathcal{B}-\mathcal{B}_1)^T + R]|^{-(m+1)/2}dU \\ &\propto |(\mathcal{B}-\mathcal{B}_1)(XX^T+A)(\mathcal{B}-\mathcal{B}_1)^T + R|^{-(n+v^*+k)/2} \\ &\quad \times \int |U|^{(n+v^*+k-m-1)/2}\mathrm{etr}\Big\{-\frac{1}{2}U\Big\}dU \\ &\propto |(\mathcal{B}-\mathcal{B}_1)(XX^T+A)(\mathcal{B}-\mathcal{B}_1)^T + R|^{-(n+v^*+k)/2}, \end{aligned}$$

which is independent of z. □

The posterior distribution of $\mathcal{B}$ under matrix normal responses is identical to the matrix-t distribution above (see Broemeling 1985, p.379).

3 The Prediction Distribution

The Bayesian approach to the prediction problem can be described as follows. Suppose Y in equation (2.1) is observable and Y_f in

$$Y_f = \mathcal{B}X_f + \psi(z)^{-1/2}E_f \tag{3.1}$$

is an unobserved $m \times n_f$ matrix of future responses with a $k \times n_f$ design matrix X_f of known values also of rank k with $n + n_f > m + k$.

The density function of (Y, Y_f) is given by

$$f(Y, Y_f|\mathcal{B}, \Phi) \propto \int f(Y, Y_f|\mathcal{B}, \Phi, z)dW(z),$$

where

$$f(Y, Y_f|\mathcal{B}, \Phi, z) \propto |\psi(z)\Phi|^{(n+n_f)/2}\text{etr}\Big\{-\frac{1}{2}\psi(z)\Phi[||Y-\mathcal{B}X||^2+||Y_f-\mathcal{B}X_f||^2]\Big\} \tag{3.2}$$

and $||M||^2 = MM^T$ for any matrix M.

The Bayesian predictive density function of Y_f is defined as

$$f(Y_f|Y) \propto \int\int\int f(Y, Y_f|\mathcal{B}, \Phi, z)\pi(\mathcal{B}, \Phi|z)d\mathcal{B}d\Phi dW(z), \tag{3.3}$$

where $\pi(\mathcal{B}, \Phi|z)$ is given in (2.4).

We have the following result based on a conjugate prior on $(\mathcal{B}, \Phi)$.

Theorem 3.1. *For the multivariate regression model (2.1) where Y has a scale mixtures of normal distributions and the prior information on $(\mathcal{B}, \Phi)$ is the conjugate prior given in (2.4), the predictive distribution of Y_f is the matrix-t distribution with $n + v^* - m + 1$ degrees of freedom, i.e., $t_{n_f m}[M, H^{-1}, R_1, n + v^* - m + 1]$, where*

$$\begin{aligned}
L &= XX^T + X_fX_f^T + A \\
H &= I_{n_f} - {X_f}^T L^{-1} X_f \\
M &= (YX^T + \mathcal{B}^*A)L^{-1}X_fH^{-1} \\
R_1 &= YY^T + \mathcal{B}^*A{\mathcal{B}^*}^T + D + (YX^T + \mathcal{B}^*A)L^{-1}(YX^T + \mathcal{B}^*A)^T \\
&\quad -(YX^T + \mathcal{B}^*A)L^{-1}X_fM^T.
\end{aligned}$$

I_{n_f} is the $n_f \times n_f$ identity matrix.

Proof. To obtain the above result, note that substituting (2.4) and (3.2) in (3.3) yields the predictive density function of Y_f as

$$f(Y_f|Y) \propto \int f(Y_f|Y, z)dW(z)$$

where

$$\begin{aligned} f(Y_f|Y,z) &\propto \int\int f(Y,Y_f|\mathcal{B},\Phi,z)\pi(\mathcal{B},\Phi|z)d\mathcal{B}d\Phi \\ &\propto \int\int \psi(z)^{m(n+n_f+v^*+k)/2}|\Phi|^{(n+n_f+v^*+k-m-1)/2} \\ &\quad \times \mathrm{etr}\Big\{-\frac{1}{2}\psi(z)\Phi[||Y-\mathcal{B}X||^2+||Y_f-\mathcal{B}X_f||^2]\Big\} \\ &\quad \times \mathrm{etr}\Big\{-\frac{1}{2}\psi(z)\Phi[(\mathcal{B}-\mathcal{B}^*)A(\mathcal{B}-\mathcal{B}^*)^T+D]\Big\}d\mathcal{B}d\Phi. \end{aligned} \tag{3.4}$$

By combining the quadratic forms, the matrix expression in (3.4) can be written as (see Broemeling, 1985)

$$\begin{aligned} &||Y_f-\mathcal{B}X_f||^2+||Y-\mathcal{B}X||^2+(\mathcal{B}-\mathcal{B}^*)A(\mathcal{B}-\mathcal{B}^*)^T+D \\ &\qquad = (Y_f-M)H(Y_f-M)^T+(\mathcal{B}-\mathcal{B}_2)L(\mathcal{B}-\mathcal{B}_2)^T+R_1 \end{aligned}$$

where $\mathcal{B}_2=(Y_fX_f^T+\mathcal{B}_1(XX^T+A))L^{-1}$ with $\mathcal{B}_1$ defined in Section 2. The integral (3.4) becomes

$$\begin{aligned} f(Y_f|Y,z) \propto \int\int &\psi(z)^{m(n+n_f+v^*)/2}|\Phi|^{(n+n_f+v^*-m-1)/2} \\ &\times \mathrm{etr}\Big\{-\frac{1}{2}\psi(z)\Phi[(Y_f-M)H(Y_f-M)^T+R_1]\Big\}d\Phi \\ &\times \psi(z)^{mk/2}|\Phi|^{k/2}\mathrm{etr}\Big\{-\frac{1}{2}\psi(z)\Phi(\mathcal{B}-\mathcal{B}_2)L(\mathcal{B}-\mathcal{B}_2)^T\Big\}d\mathcal{B}. \end{aligned}$$

Integrating out $\mathcal{B}$, we have

$$\begin{aligned} f(Y_f|Y,z) \propto \int &\psi(z)^{m(n+n_f+v^*)/2}|\Phi|^{(n+n_f+v^*-m-1)/2} \\ &\times \mathrm{etr}\Big\{-\frac{1}{2}\psi(z)[(Y_f-M)H(Y_f-M)^T+R_1]\Phi\Big\}d\Phi. \end{aligned}$$

The transformation $U=\psi(z)[(Y_f-M)H(Y_f-M)^T+R_1]\Phi$ with Jacobian of transformation $|\psi(z)[(Y_f-M)H(Y_f-M)^T+R_1]|^{-(m+1)/2}$ yields

$$\begin{aligned} f(Y_f|Y,Z) &\propto |(Y_f-M)H(Y_f-M)^T+R_1|^{-(n+n_f+v*)/2} \\ &\quad\times\int|U|^{(n+n_f+v^*-m-1)/2}\mathrm{etr}\Big\{-\frac{1}{2}U\Big\}dU \\ &\propto |(Y_f-M)H(Y_f-M)^T+R_1|^{-(n+n_f+v*)/2}, \end{aligned}$$

which is independent of z. □

The predictive distribution of Y_f is the same as that for matrix normal responses (see Broemeling, 1985) and extends the result in Jammalamadaka *et al.* (1987).

4 Concluding Remarks

When random responses in a multivariate regression model are assumed to have a multivariate scale mixtures of normal distributions, the Bayesian analysis using a prior in the conjugate family yields posterior distribution of the regression parameters and predictive distribution of future responses that are identical to those obtained under independently distributed normal responses. Hence inference on regression parameters and future responses is unaffected by departures from the normality assumption in the direction of scale mixtures of normal distributions. The marginal distribution of the regression parameters and predictive distribution are therefore invariant to a wider class of distributions of the responses in a Bayesian analysis using informative prior.

5 Acknowledgment

The author could like to thank an anonymous referee for valuable comments.

References

[1] T.W. Anderson and K.T. Fang, (1990). Inference in multivariate elliptically contoured distribution based on maximum likelihood, in: *Statistical Inference in Elliptically Contoured and Related Distribution,* (K.T. Fang and T.W. Anderson Ed.), pp. 201–216, Allerton Press, New York.

[2] G.E.P. Box and G.C. Tiao, (1973). *Bayesian Inference in Statistical Analysis,* Addison-Wesley, Reading, Mass.

[3] L.D. Broemeling, (1985). *Bayesian Analysis of Linear Models,* Marcel Dekker, New York.

[4] D.A.S. Fraser, (1979). *Inference and Linear Models*, McGraw-Hill, New York.

[5] D.A.S. Fraser and M.S. Haq, (1969). Structural probability and prediction for the multilinear model, *J. R. Stat. Soc. Ser. B Stat. Methodol.* **31**, 317–331.

[6] D.A.S. Fraser and K.W. Ng, (1980). Multivariate regression analysis with spherical error, in: *Multivariate Analysis-V,* (P.R. Krishnaiah Ed.), pp.369–386, North-Holland, New York.

[7] S. Geisser, (1965). Bayesian estimation in multivariate analysis, *Ann. Math. Statist.* **36**, 150–159.

[8] S. Geisser, (1993). *Predictive Inference: An Introduction*, Chapman & Hall, London.

[9] I. Guttman and P. Hougaard, (1985). Studentization and prediction problems in multivariate multiple regression, *Comm. Statist. Theory Methods* **14**, 1251–1258.

[10] M.S. Haq and S. Khan, (1990). Prediction distribution for a linear regression model with multivariate Student-t error distribution, *Comm. Statist. Theory Methods* **19**, 4705–4712.

[11] S.R. Jammalamadaka, R.C. Tiwari and S. Chib, (1987). Bayes prediction in the linear model with spherically symmetric errors, *Econom. Lett.* **24**, 39–44.

[12] N.L. Johnson and S. Kotz, (1972). *Distributions in Statistics: Continuous Multivariate Distributions,* Wiley, New York.

[13] S. Khan, (2002). Distribution of sum of squares and products matrices for generalized multilinear matrix-T model, *J. Multivariate Anal.* **83**, 124–140.

[14] S. Khan, (2004). Predictive distribution of regression vector and residual sum of squares for normal multiple regression model, *Comm. Statist. Theory Methods* **33**, 2423–2443.

[15] J. Kowalski, X.M. Tu and L.J. Gleser, (1999). On the difference in inference and prediction between the joint and independent t-error models for seemingly unrelated regression, *Comm. Statist. Theory Methods* **28**, 2199–2140.

[16] V. M. Ng, (2000). A note on predictive inference for multivariate elliptically contoured distributions, *Comm. Statist. Theory Methods* **29**, 477–483.

[17] V. M. Ng, (2002). Robust Bayesian inference for seemingly unrelated regressions with elliptical errors, *J. Multivariate Anal.* **83**, 409–414.

[18] N. Ravishanker and D. K. Dey, (2002). *A First Course in Linear Model Theory,* Chapman & Hall/CRC, Boca Raton.

[19] A. Zellner, (1976). Bayesian and non-Bayesian analysis of the regression model with multivariate Student-t error terms, *J. Amer. Statist. Assoc.* **71**, 400–405.