PRIOR-LIKELIHOOD FACTORIZATION AND MISSING DATA

D.A.S. Fraser
University of Toronto
York University
University of Waterloo

S.E. Sackey
University of Toronto

Key Words and Phrases: Missing data; absent data; multivariate regression; prior-likelihood factorization.

ABSTRACT

Marginal posterior distributions, when not available analytically, can be at present numerically inaccessible if the number of parameters for integration exceeds 7 to 10. For the normal multivariate regression model, with data absent (missing) in a monotone pattern, some integrations have been accomplished analytically (Guttman and Menzefricke, 1983; Bartlett, 1983; for example).

In this note we show how monotonely missing data support an extended prior-likelihood factorization and the needed posterior can be obtained directly using standard results.

1. INTRODUCTION

Data that are absent or missing at random have been discussed by Rubin (1976). For the multivariate normal regression model,
Guttman and Menzefricke (1983) used the predictive density for missing data to determine the posterior distribution of the regression parameters. Bartlett (1983) worked directly with the prior times likelihood and derived the joint posterior for the regression and covariance parameters and for the regression matrix itself.

In this note we show how a suitable prior-likelihood factorization allows the various posterior distributions to be recorded from standard results for the full-data normal multivariate regression model.

2. BACKGROUND

Consider the normal multivariate regression model

\[(2.1) \quad Y = XB + E = XB + Z\Pi\]

where \(Y\) is the \(n \times p\) matrix of dependent variables, \(X\) is a rank-\(k\) \(n \times k\) design matrix for \(k\) independent variables, \(B\) is a \(k \times p\) matrix of regression parameters, \(E\) has \(n\) independent rows each \(\mathcal{N}_p(Q', \Sigma)\); in the alternative form \(Z\) is an \(n \times k\) matrix of \(\mathcal{N}(0,1)\) variables and \(\Pi\) is \(p \times p\) positive upper triangular with \(\Pi'\Pi = \Sigma\).

The posterior distribution of \(B, \Sigma\) has been studied by Tiao and Zellner (1964), Zellner (1971), Box and Tiao (1973) in terms of the non-informative prior

\[(2.2) \quad d\beta d\Sigma / |\Sigma|^{(p+1)/2}.

This prior can be obtained by Jeffreys standardization with respect to the information matrix (cf. Box and Tiao, 1973) or as the right invariant prior with respect to the positive linear group or from an analysis of the positive linear structural model (eg. Fraser, 1978, ch. 8). We find it convenient to also record the results for the more general prior
(2.3) \[ dBd\Sigma / |\Sigma|^{(p+1+f)/2} \]

where \( f = 0 \) gives (2.2).

The density for (2.1) and the prior (2.3) give the posterior

\[
P(B, \Sigma | Y, X) \propto |\Sigma|^{-\frac{n+p+1+f}{2}} \text{etr}(\frac{1}{2} \Sigma^{-1}) \times \text{etr}(\frac{1}{2} (B-\hat{B})'X'X(B-\hat{B})\Sigma^{-1})
\]

(2.4)

where \( \hat{B} = (X'X)^{-1}X'Y \), \( \Sigma = (Y-X\hat{B})'(Y-X\hat{B}) \). It follows that \( \Sigma \sim W_p^{-1}(S, n-k+f) \), \( B | \Sigma \sim N_{kp} \left( \hat{B}, \Sigma \otimes (X'X)^{-1} \right) \), \( B \sim t_{kp} \left( \hat{B}, (X'X)^{-1}, S, \nu+f \right) \) with \( \nu = n-k-p+1 \) and density

\[
a | S + (B-\hat{B})'X'X(B-\hat{B}) |^{-\frac{n+f}{2}}.
\]

(2.5)

For inference later we record the distribution of the positive upper triangular square root \( \Pi \) of \( \Sigma = \Pi'\Pi \):

\( \Pi \sim \Delta^{-1} \chi_p(S, n-k+f) \), an inverted triangular chi with \( n-k+f \) degrees of freedom, with density

\[
(2.6) \quad \frac{A_{(p)}^{(p)}}{(2\pi)^{\frac{p(n-k+f)}}{2}} \frac{|S|^{n-k+f}}{\text{etr}(\frac{1}{2} (\Pi'\Pi)^{-1}S)} ; \quad |\Pi|^{n-k+f} |\Pi|_{\Delta}
\]

see Appendix A.

3. ABSENT DATA AND COMPONENT MODELS

Consider the normal regression model (2.1) and suppose that the response is available for all \( n = n_1 + n_2 \) observations on
the first \( p_1 \) variables and available only for the last \( n_2 \) observations on the last \( p_2 \) variables \((p_1+p_2=p)\).

The first \( p_1 \) columns of the modified (2.1) can be written as

\[
(3.1) \quad Y_1 = XB_1 + E_1 = XB_1 + Z_1 \Pi_{11}
\]

where \( E_1 \) has \( n = n_1 + n_2 \) independent rows each \( N_{p_1} \left( 0, \Sigma_{11} \right) \)

where \( \Sigma_{11} \) is the top-left \( p_1 \times p_1 \) submatrix of \( \Sigma \), \( B_1 \)
records the first \( p_1 \) columns of \( B \), and \( \Pi_{11} \) is the top-left \( p_1 \times p_1 \) submatrix of \( \Pi \), with \( \Sigma_{11} = \Pi_{11} \Pi_{11} \); see Appendix B.

For the remaining \( p_2 \) columns, we have the last \( n_2 \) observations from the modified (2.1) which can be written as

\[
(3.2) \quad Y_2 = X_{(n_2)} B_2 + E_{(n_2)} H_{12} + E_{2 \cdot 1}
\]

\[
= X_{(n_2)} B_2 + Z_{1(n_2)} \Pi_{12} + Z_{2 \cdot 1} \Pi_{22}
\]

where \( Y_2 \) is \( n_2 \times p_2 \), \( M_{(n_2)} \) designates the matrix \( M \) as
trimmed to its last \( n_2 \) observations (rows); \( E_{2 \cdot 1} \) is an
\( n_2 \times p_2 \) matrix of residual errors (after regression on \( E_{1(n_2)} \))
with independent rows each \( N_{p_2} \left( 0, \Sigma_{22 \cdot 1} \right) \); and

\[
E_{22 \cdot 1} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \quad \Pi_{12} = \Pi_{11} H_{12} \quad H_{12} = \Sigma_{11}^{-1} \Sigma_{12}
\]

The preceding equations can be expressed as

\[
(3.3) \quad Y_2 = X_{(n_2)} B_2 + \left( Y_{1(n_2)} - X_{(n_2)} B_1 \right) H_{12} = E_{2 \cdot 1}
\]

\[
= X_{(n_2)} B_2^* + Y_{1(n_2)} H_{12} + E_{2 \cdot 1}
\]

\[
= \begin{pmatrix} X_{(n_2)} & Y_{1(n_2)} \end{pmatrix} \begin{pmatrix} B_2^* \\ H_{12} \end{pmatrix} + E_{2 \cdot 1}
\]
\[ = X(n_2)^{B^*} + Y_1(n_2)^{H_{12}} + Z_{2\cdot 1}^{\Pi_{12}} \]
\[ = (X(n_2) ; Y_1(n_2)) \begin{bmatrix} B^* \\ H_{12} \end{bmatrix} + Z_{2\cdot 1}^{\Pi_{12}} \]

where \( B^* = B_2 - B_1 H_{12} \).

Now consider the general prior (2.3) and its factorization in accordance with the \( p = p_1 + p_2 \) variables:

\[
\frac{d B d \Sigma}{|\Sigma| (p+1+f)/2} = \frac{d B_1 d \Sigma_1}{|\Sigma_1| (p_1+1+f_1)/2} \cdot \frac{d B_2^* d H_{12} d \Sigma_{2\cdot 1}}{|\Sigma_{2\cdot 1}| (p_2+1+f_2)/2}
\]

where \( f_1 = f - p_2 \), \( f_2 = f + p_1 \), and the differential changes follow from

\[ H_{12} = \Sigma_{11}^{-1} \Sigma_{12} \quad \Sigma_{2\cdot 1} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \]

The posteriors from (3.2) and (3.3) with the factored prior in (3.4) can be recorded directly from the standard results in section 2.

For the submodel (3.1) we have

\[ \Sigma_{11} \sim W_{p_1}^{-1} (S_{11}, n-k+f_1) \]

where

\[ S_{11} = (Y_1 - \hat{X}_1 \hat{B}_1)'(Y_1 - \hat{X}_1 \hat{B}_1) \quad \hat{B}_1 = (X'X)^{-1}X'Y_1 \]

and we have \( B_1 | \Sigma_{11} \sim N_{k \times p_1} (\hat{B}_1, \Sigma_{11} \otimes (X'X)^{-1}) \)

\[ B_1 \sim t_{kp_1} (\hat{B}_1, (X'X)^{-1}, S_{11}, v_1+f_1) \]
where \( Q = (I^*_{B_1} M^T_{B_1} I_{B_1}) \).

Note that this involves \( B_1 \) and thus is conditional on \( B_1 \), from the first submodel. The marginal for \( B_2 \) from the present submodel is then

\[
B_2 \sim t_{k_p^2} (\hat{B}_2^* + B_1 \hat{H}_{12}, Q, S_{22}, \nu_2 + f_2)
\]

where \( \nu_2 = n_2 - k - p_1 - p_2 + 1 \).

4. ABSENT DATA AND COMBINED POSTERIORS

Consider the combined model from (3.1) and (3.2), and the general posterior (3.4).

We use the results from section 3 to obtain the posterior distribution of the regression parameters \( B \) and the variance parameters \( \Sigma \).

For the regression parameters \( B \) we have that

\[
B_1 \sim t_{k_p^1} (\hat{B}_1, (X'X)^{-1}, S_{11}, \nu_1 + f_1)
\]

and

\[
B_2 | B_1 \sim t_{k_p^2} (\hat{B}_2^* + B_1 \hat{H}_{12}, Q, S_{22}, \nu_2 + f_2).
\]

It follows that the joint density \( g_B \) for \( B = (B_1, B_2) \) is

\[
g_B = g_{B_1} \cdot g_{B_2 | B_1},
\]

giving

\[
g_B = \frac{(p_1!)}{(p_1!)^A} \frac{(n + f_1 - k)^{n + f_1 - k}}{2^{n + f_1 - k}} \frac{1}{|X'X|^{p_1/2} |S_{11}|^{\nu_1/2}} \frac{1}{|S_{11} + (B_1 - \hat{B}_1)' X' (B_1 - \hat{B}_1)|^{\nu_2 + f_2/2}}.
\]
\[
\begin{align*}
\rho & = \frac{A_{n-k+f_1}}{p_1} \left| \frac{S_{11}}{2} \right|^2 \\
\Delta & = \frac{n-k+f_1}{p_1} \\
\sigma_1 & = \frac{A_{n-k+f_1}}{p_1} \left| \frac{S_{11}}{2} \right|^2 \\
\sigma_2 & = \frac{n}{p_1} \frac{S_{11}}{2} \\
\end{align*}
\]

For the variance parameter \( \Sigma \), we have that

\[
\Sigma_{11} \sim W^{-1}(S_{11}, n-k+f_1) \\
\Sigma_{22} \sim W^{-1}(S_{22}, n_2-k-p_1+f_2)
\]

and

\[
H_{12} = \Sigma_{11}^{-1} \Sigma_{12} \cdot \Sigma_{22} \sim \mathcal{N}_{p_1 \times p_2} (\hat{H}_{12}, \Sigma_{22} \otimes M_{22})
\]

where \( M_{22} \) designates the lower right \( p_2 \times p_2 \) submatrix of \( M \).

The density \( g_\Sigma \) for \( \Sigma \) can be obtained by combining the preceding components in the conditioning order \( \Sigma_{11}, \Sigma_{22}, \Sigma_{12} \).

Thus

\[
(4.2)
\]

\[
g_\Sigma = g_{\Sigma_{11}} \cdot g_{\Sigma_{22}} \cdot g_{\Sigma_{12}}
\]

\[
= \frac{(p_1)}{A_{n-k+f_1}} \cdot \frac{n-k+f_1}{S_{11}}^2 \cdot \frac{1}{2^{\frac{p_1}{2} \left| \Sigma_{11} \right|^2}}
\]
\[
\begin{align*}
\frac{(p_2)}{A_{n_2-k-p_1+f_2}} \cdot \frac{n_2-k-p_1+f_2}{p_2(n_2-k-p_1+f_2)} \frac{|s_{22\cdot1}|}{2} \\
\frac{1}{(n_2-k-p_1+f_2+p_2+1)} \text{etr}\left\{\frac{1}{2} S_{22\cdot1} \Sigma_{22\cdot1}^{-1}\right\} \frac{p_2}{2} |\Sigma_{22\cdot1}|^{-p_1/2} \frac{|\Sigma_{22\cdot1}|^{-p_1/2}}{p_2/2} \\
\frac{1}{\Sigma_{11} M_{22} \Sigma_{11}'} \text{etr}\left\{\frac{1}{2} \Sigma_{22\cdot1}^{-1} \left(\Sigma_{12} - \Sigma_{11} \hat{H}_{12}\right) \left[\Sigma_{11} M_{22} \Sigma_{11}' \right]^{-1} \left(\Sigma_{12} - \Sigma_{11} \hat{H}_{12}\right)\right\} \\
(4.3) \\
= c^* \left|\Sigma_{11}\right|^{-2} \left|\Sigma_{22\cdot1}\right|^{-2} \left|M_{22}\right|^{-2} \\
\cdot \text{etr}\left\{\frac{1}{2} \Sigma_{11}^{-1} S_{11}\right\} \cdot \text{etr}\left\{\frac{1}{2} \Sigma_{22\cdot1}^{-1} S_{22\cdot1}\right\} \\
\cdot \text{etr}\left\{\frac{1}{2} \Sigma_{22\cdot1}^{-1} \left(\Sigma_{12} - \Sigma_{11} \hat{H}_{12}\right) \left[\Sigma_{11} M_{22} \Sigma_{11}' \right]^{-1} \left(\Sigma_{12} - \Sigma_{11} \hat{H}_{12}\right)\right\}
\end{align*}
\]

where

\[
c^* = \frac{(p_1) A_{n-k+f_1} \cdot (p_2) A_{n_2-k-p_1+f_2} \left|S_{11}\right|^{2} \left|S_{22\cdot1}\right|^{2}}{(2\pi)^{m/2} 2^p} \frac{n-k+f_1}{n_2-k-p_1+f_2}
\]
and

\[ m = p_1(n-k+f) + p_2(n_2-k-p_1+f) + p_1p_2. \]

**BIBLIOGRAPHY**


**APPENDIX A**

For \( S \sim W_p^{-1}(S, n-k+f) \), the corresponding density can be written as
PRIOR-LIKELIHOOD FACTORIZATION

\begin{equation}
\frac{A_{d}^{(p)}}{p(n-k+f)} \cdot \frac{|S|^{n-k+f}}{(2\pi)^{\frac{n-k+f}{2}} \cdot 2^{p} |\Sigma|^{\frac{n-k+f}{2}}} \cdot \text{etr} \left\{ -\frac{1}{2} \left( \Pi' \Pi - l_s \right) \right\}
\end{equation}

where \( A_{d}^{(p)} = A_{d}A_{d-1} \ldots A_{d-p+1} \), a descending product of \( A_{d} = \frac{2\pi^{d/2}}{\Gamma(d/2)} \) which is the surface volume of the unit sphere in \( \mathbb{R}^{d} \). The relationship \( \Sigma = \Pi' \Pi \) yields

\begin{equation}
J(\Sigma + \Pi) = 2^{p} |\Pi|_{\nabla}
\end{equation}

where \( |\ \ |_{\nabla} \) is the descending determinant. From (A.1) we then obtain the density of \( \Pi \):

\begin{equation}
\frac{A_{d}^{(p)}}{p(n-k+f)} \cdot \frac{|S|^{n-k+f}}{(2\pi)^{\frac{n-k+f}{2}} \cdot |\Pi|^{n-k+f+p+1}} \cdot \text{etr} \left\{ -\frac{1}{2} \left( \Pi' \Pi - l_s \right) \right\}
\end{equation}

\begin{equation}
= \frac{A_{d}^{(p)}}{p(n-k+f)} \cdot \frac{|S|^{n-k+f}}{(2\pi)^{\frac{n-k+f}{2}} \cdot |\Pi|^{n-k+f} |\Pi|_{\Delta}} \cdot \text{etr} \left\{ -\frac{1}{2} \left( \Pi' \Pi - l_s \right) \right\}
\end{equation}

where \( |\ \ |_{\Delta} \) is the increasing determinant.

APPENDIX B

The component parts of \( \Pi \) are easily related to components of the regression and variance matrices:

\( \Pi' \Pi_{11} = \Sigma_{11} \),

\( \Pi_{11} \Pi_{12} = \Sigma_{12} = \Sigma_{11} \Pi_{12} \), \( \Pi_{12} = \Sigma_{11}^{-1} \Pi_{12} = \Pi_{11}^{-1} \Pi_{12} \).

Received April, 1986; Retyped June, 1986.
Recommended Anonymously.