BAYESIAN METHODS IN THE ANALYSIS OF THE GENERAL LINEAR MODEL

by

Charles A. Rohde

Department of Biostatistics and Statistics
The Johns Hopkins University

Institute of Statistics
Mimeograph Series No. 545
1967

Research partially supported by
NSF Science Development Grant GU-1590

BAYESIAN METHODS IN THE ANALYSIS OF THE GENERAL LINEAR MODEL*

By Charles A. Rohde

1. Introduction

Recent interest in the analysis of statistical data using Bayesian methods has concentrated on regression models, goodness of fit tests and to a lesser extent random models in the analysis of variance. The purpose of this paper is to indicate the simple yet pedagogically important extensions to the analysis of the so-called fixed model or Model I. The proofs involved seem to the author to be simple enough to present in statistical theory courses at the senior or first-year graduate level where they permit a unified look at the analysis of linear models using Bayesian methods. As a by-product of the investigation, an explicit relationship between estimates in original and reparametrized models is obtained. It is also shown that the concept of linearly estimable functions largely ignored by applied statisticians appears in the Bayesian analysis as the class of linear parametric functions with non-degenerate posteriof distributions.

---

*Paper Number 388 in the Department of Biostatistics, The Johns Hopkins University

2. Linear model of full rank

In the analysis of linear models using Bayesian methods it is usual (Lindley (5), Box and Tiao (2)) to take independent non-informative prior distributions for the elements of the parameter vector and an (independent) non-informative prior distribution for log θ where θ is the variance of an observation. A set of assumptions underlying a Bayesian analysis of the general linear model can thus be written as

$$(1) \quad \underline{Y} \text{ is } MVN(X\underline{\beta},\ I_n\theta)$$

$$(2) \quad \pi(\underline{\beta},\theta)\ \alpha\ \frac{1}{\theta}$$

where $\underset{\sim}{X}$ is an nxp matrix of fixed or known variables of rank $r \leq p < n$ $\underline{\beta}$ is a px1 vector of parameters and I_n denotes an identity matrix of order n. When the rank of $\underset{\sim}{X}, \rho(\underset{\sim}{X})$, is equal to p the Bayesian analysis proceeds rather routinely and is fully discussed in Lindley (5). The joint posterior distribution of $\underline{\beta}$ and θ is clearly proportional to

$$\theta^{-\frac{n+2}{2}}\ e^{-\frac{1}{2\theta}\ (\underline{y}-\underset{\sim}{X}\underline{\beta})'\ (\underline{y}-\underset{\sim}{X}\underline{\beta})}$$

noting that we can write

$$(3) \quad (\underline{y}-\underset{\sim}{X}\underline{\beta})'\ (\underline{y}-\underset{\sim}{X}\underline{\beta}) \ = \ (\underline{y}-\underset{\sim}{X}\underline{b})'\ (\underline{y}-\underset{\sim}{X}\underline{b}) \ + \ (\underline{\beta}-b)'\underset{\sim}{X}'\underset{\sim}{X}(\underline{\beta}-\underline{b})$$

where $\underline{b}$ satisfies the normal equations $\underset{\sim}{X}'\underset{\sim}{X}\underline{b} = \underset{\sim}{X}'\underline{y}$ we have

$$(4) \quad \pi(\underline{\beta},|/y)\ \alpha\ \theta^{-\frac{n+2}{2}}\ \exp\ -\frac{1}{2\theta}\left[S_e^2 + (\underline{\beta}-\underline{b})'(\underset{\sim}{X}'\underset{\sim}{X})\ (\underline{\beta}-b)\right]$$

where $S_e^2 = (\underline{y}-\underset{\sim}{X}\underline{b})'\ (\underline{y}-\underset{\sim}{X}b)$. Integration of (4) with respect to θ shows that

$$\pi(\underline{\beta})\ \ \alpha\ \left[1 + \frac{(\underline{\beta}-\underline{b})'(X'X)\ (\underline{\beta}-b)}{S_e^2}\right]^{-\frac{n}{2}}$$

i.e., β has a multivariate t distribution as discussed by Geisser and Cornfield (3). Integration of (4) with respect to β shows that the posterior distribution of S_e^2/θ is chi-square with (n-p) d.f. It can be directly established (Lindley (5)) using the above results that the posterior distribution of

$$\frac{(\beta-b)'X'X(\beta-b)}{S_2^2} \quad \frac{(n-p)}{p}$$

is F with p and n-p degrees of freedom. Alternatively one can easily establish the following Lemma.

Lemma 1: If the posterior distribution of β for fixed θ is MVN(b, $V\theta$) and the posterior distribution of S_e^2/θ is chi-square with (n-p) d.f. then the posterior distribution of

$$\frac{\beta-b)'V^{-1}(\beta-b)}{S_e^2} \quad \frac{n-p}{p}$$

is F with p and (n-p) d.f.

Armed with Lemma 1 it is easily established that the posterior distribution of a set of linear functions $L\beta$ (L is qxp with $\rho(L) = q \leq p$) is such that

$$\frac{(L\beta - Lb)'[L(X]X)^{-1}L']^{-1}[L\beta - Lb]}{S_e^2} \quad \frac{n-p}{q}$$

has an F distribution with q and (n-p) d.f. Posterior confidence regions for $L\beta$ or tests of hypotheses could thus be based on the F distribution or on the multivariate t distribution. The posterior confidence regions would not be highest posterior density regions using the usual F tables. Tables for obtaining such regions do not appear to be available at present. In particular inference on a subset of $\underline{\beta}$, say $\underline{\beta}_2$ can be based on

$$\frac{(\underline{\beta}_2-\underline{b}_2)'(X_2'D_1X_2)^{-1}(\underline{\beta}_2-\underline{b}_2)}{s_e^2} \quad \frac{n-p}{q}$$

which has a posterior F distribution with q and n-p d.f. Here $\underline{\beta}_2(\underline{b}_2)$ is a q component subset of $\underline{\beta}(\underline{b})$, $X = [X_1 \mid X_2]$ has been rewritten to conform with the partition of $\underline{\beta}$, and $D_1 = I - X_1(X_1'X_1)^{-1}X_1'$.

The reader familiar with classical least squares theory will recognize that $\underline{b}_2' (X_2'D_1X_2)^{-1}\underline{b}_2$ is the so-called sum of squares for $\underline{\beta}_2$ adjusted for $\underline{\beta}_1$ and as in the classical case can be used to test the hypothesis that $\underline{\beta}_2 = 0$. The concept of adjusted sums of squares arises naturally in the Bayesian treatment when the marginal posterior distribution of $\underline{\beta}_2$ is considered. Intuitively the reason is that the effect of $\underline{\beta}_1$ must be removed before the posterior distribution of $\underline{\beta}_2$ can be considered. Technically the presence of X'X in the exponent of (4) and its "inversion" upon integration with respect to $\underline{\beta}_1$ accounts for the appearance of the adjusted sum of squares.

3. Linear model of less than full rank

When the rank of X is not full (i.e. $\rho(X) = r<p$) the analysis of Section 2 is plausible but the proofs need some alteration. The reason for the difficulty is the appearance of degenerate or singular distributions due to the singularity of $X'X$. If degenerate prior distributions are not allowed, the singular distributions do not appear. Such an approach is discussed in section 4. The method of proof used here is to consider linear combinations of the elements of $\underline{\beta}$. Specifically we restrict attention to sets of linear functions $L\underline{\beta}$ such that

$$L = CX$$

where L is qxp of rank q. The resulting theory leads to results analogous to that of section 2. Specifically we shall show that the posterior distribution of

$$\frac{(L\underline{\beta}-L\underline{b})'[L(X'X)^{-}L']^{-1}(L\underline{\beta}-L\underline{b})}{S_e^2} \quad \frac{n-r}{q} \tag{5}$$

is F with q and (n-r) d.f. Here $(X'X)^{-}$ is any generalized inverse of $X'X$,[1] $\underline{b}$ is any solution to the normal equations $X'X\underline{b} = X'\underline{y}$ and $S_e^2=\underline{y}'\underline{y}-\underline{y}'X\underline{b}$ is the error sum of squares.

To establish (5) we first reparametrize the original model. Following Graybill (4) we note that there exists a non-singular matrix P such that

$$P'X'XP = \begin{bmatrix} B & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{0} \end{bmatrix}$$

where B is non-singular of rank r and $\underset{\sim}{0}$ denotes null matrices of appropriate

---

[1] A generalized inverse of a matrix A is a matrix A^{-} such that $AA^{-}A = A$. No standard definition appears to exist since some authors define a generalized inverse by the equations $AA^{-}A=A$, $A^{-}AA^{-}=A^{-}$, $(AA^{-})'=AA^{-}$ and $(A^{-}A)'=A^{-}A$ while others call such an A^{-} a pseudo-inverse of A. It is well known (Rao (7)) that if $AA'A = A$ then $A(A'A)^{-}A'$ is unique, symmetric and idempotent.

orders. If we let $P = [P_1 | P_2]$ where P_1 is pxr and P_2 is px(p-r) it follows that

$$P_1'X'XP_1 = B$$
$$XP_2 = \underset{\sim}{0}$$

Defining

$$Q = P^{-1} = \begin{bmatrix} Q_1 \\ \hline Q_2 \end{bmatrix}$$

where Q_1 is rxp and Q_2 is (p-r) xp we can write

$$X\underline{\beta} = XPQ\underline{\beta} = X[P_1 | P_2] \begin{bmatrix} Q_1 \\ \hline Q_2 \end{bmatrix} \underline{\beta} = [XP_1 | \underset{\sim}{0}] \begin{bmatrix} Q_1 \\ \hline Q_2 \end{bmatrix} \underline{\beta}$$
$$= XP_1Q_1\underline{\beta}$$

Letting $Z = XP_1$ and $\underline{\alpha} = Q\underline{\beta}$ yields

$$X\underline{\beta} = Z\underline{\alpha}$$

where Z is nxr of rank r and $\underline{\alpha}$ is an rx1 vector of parameters.

It is now clear that we can write the model in the form

$$\underline{Y} \sim MVN(Z\underline{\alpha}, I\theta)$$

which is of the full rank form discussed in Section 2. The assumption of a non-informative prior distribution for $\underline{\beta}$ is clearly equivalent to the assumption of a non-informative prior distribution for $\underline{\alpha}$ if the range of $\underline{\beta}$ is infinite so that the posterior distribution of α and θ is

$$\pi(\underline{\alpha},\theta) \propto \theta^{-\frac{n+2}{2}} \exp - \frac{1}{2\theta} \quad (\underline{y}-Z\underline{\alpha})'(\underline{y}-Z\alpha)$$

Since Z is of full rank we can write

$$(\underline{y}-Z\underline{\alpha})'(\underline{y}-Z\underline{\alpha}) \equiv (\underline{y}-Z\underline{a})'(\underline{y}-Z\underline{a}) + (\underline{\alpha}-\underline{a})'Z'Z(\underline{\alpha}-\underline{a})$$

where $\underline{a}$ satisfies the normal equations $Z'Z\underline{a} = Z'\underline{y}$. Note that

$$(\underline{y}-Z\underline{a})'(\underline{y}-Z\underline{a}) = \underline{y}'\underline{y} - \underline{y}'Z\underline{a} = \underline{y}'[I-Z(Z'Z)^{-1}Z']\underline{y}$$

Since X and Z have the same column space, it follows that $X\underline{b} = Z\underline{a}$ so that $S_e^2 = (\underline{y} - Z\underline{a})'(\underline{y}-Z\underline{a}) = (\underline{y}-X\underline{b})'(\underline{y}-X\underline{b})$ and $X(X'X)^{-}X' = Z(Z'Z)^{-1}Z'$.

Thus the posterior distribution of $\underline{\alpha}$ and θ is

$$\pi(\underline{\alpha},\theta) \propto \theta^{-\frac{n+2}{2}} \exp -\frac{1}{2\theta}\left[S_e^2 + (\underline{\alpha}-\underline{a})'Z'Z(\underline{\alpha}-\underline{a})\right]$$

so that for fixed θ, $D\underline{\alpha}$ is $MVN(D\underline{a},[D(Z'Z)^{-1}D']\phi)$ where D is qxr of rank $q \leq r$ while S_e^2/θ has a posterior χ^2 distribution with (n-r) d.f. The only thing remaining to prove is that one can get $L\underline{\beta}$ (where L = CX for some C) uniquely from $D\underline{\alpha}$ and conversely for some D and some reparametrization P. The obvious affirmative answer to this question is contained in the following Lemma.

<u>Lemma 2</u>: Consider a general linear model $\mathcal{E}[\underline{Y}] = X\underline{\beta}$ which is reparametrized so that $\mathcal{E}[\underline{Y}] = Z\underline{\alpha}$ where

$$Z = XP_1, \begin{bmatrix} P_1' \\ \hline P_2' \end{bmatrix} X'X[P_1 \mid P_2] = \begin{bmatrix} P_1'X'XP_1 & \underset{\sim}{0} \\ \hline \underset{\sim}{0} & \underset{\sim}{0} \end{bmatrix}, P = [P_1|P_2]$$

being non-singular. Then if $D = L(X'X)^{-}X'Z = LP_1$ we have

(a) $L\underline{\beta} = D\underline{\alpha}$

(b) $L\underline{b} = D\underline{a}$

(c) $L(X'X)^{-}L' = D(Z'Z)^{-1}D'$

<u>Proof</u>: Since L = CX for some C we have

$$D = CX(X'X)^{-}X'Z = CX(X'X)^{-}X'XP_1 = CXP_1 = LP_1$$

so that

(a) $D\underline{\alpha} = LP_1\underline{\alpha} = CXP_1\underline{\alpha} = C[XP_1|\underset{\sim}{0}]\begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix}\underline{\beta} = CX[P_1|P_2]\begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix}\underline{\beta}$

$= L\underline{\beta}$

(b) $D\underline{a} = CX(X'X)^{-}X'Z(Z'Z)^{-1}Z'\underline{y} = L\underline{b}$

(c) $D(Z'Z)^{-1}D' = CX(X'X)^{-}X'Z(Z'Z)^{-1}Z'X(X'X)^{-}X'C'$

$= CX(X'X)^{-}X'C'$

$= L(X'X)^{-}L'$

Lemma 2 thus provides us with an explicit one to one passage from $L\underline{\beta}$ to $D\underline{\alpha}$ in the reparametrized model. Such a one-to-one passage does not appear to have been previously obtained. In particular, since the functions $L\underline{\beta}$ for which $L = CX$ have been called linearly estimable functions in classical least squares theory we have established an explicit one to one correspondence between the linearly estimable functions $L\underline{\beta}$ and the same linearly estimable functions in the reparametrized model given by

$$D = LP_1.$$

In the Bayesian analysis it is clear that we have shown that the collection of linear parametric functions $L\underline{\beta}$ which have non-degenerate posterior distributions coincides with the class of linearly estimable functions in the classical theory. The Bayesian analysis for testing purposes is summarized in the following theorem.

Theorem 1: Let $\underline{Y}$ be $MVN(X\underline{\beta}, I\theta)$, $\rho(X) = r$, with independent non-informative prior distributions for $\underline{\beta}$ and $\log\theta$. Also let $L\underline{\beta}$ be a set of q linearly independent estimable functions. Then the posterior distribution of

$$\frac{(L\underline{\beta}-L\underline{b})'\ [L(X'X)^{-}L']^{-1}(L\underline{\beta}-L\underline{b})}{S_e^2} \qquad \frac{n-r}{q}$$

if F with q and (n-r) d.f.

Proof: From the reparametrized form of the model it is clear that $D\underline{\alpha}$ for fixed θ has a $MVN[Z\underline{\alpha},(Z'Z)^{-1}\theta]$ posterior distribution. Since S_e^2/θ has a posterior χ^2 distribution with n-r d.f. the conclusion follows by applying Lemma 1.

Confidence regions for $L\beta$ with highest posterior density (following Box and Tiao (2)) could be obtained using Theorem 1 (although tables are not yet available) or using the posterior distribution of $L\beta$ (which is obviously a multivariate t distribution).

4. An alternative development

As an alternative development of the results in Section 3, we can suppose that β has a $MVN(\beta_o, \phi V)$ posterior distribution. Then under the assumption that log ϕ has a non-informative prior distribution independent of β we have

$$\pi(\beta,\theta/y) \ \alpha \ \theta^{-\frac{n+2}{2}} e^{-\frac{1}{2}Q}$$

where $Q = (y-X\beta)'\theta^{-1}(y-X\beta) + (\beta-\beta_o)'V^{-1}\phi^{-1}(\beta-\beta_o)$.

Thus

$$\Phi(\beta,\theta y) \ \alpha \ \theta^{-\frac{n+2}{2}} e^{-\frac{1}{2}\beta'[\phi^{-1}V^{-1}+\theta^{-1}X'X]\beta+(\theta^{-1}X'y+\phi^{-1}V^{-1}\beta_o)'\beta}$$

from which it follows that for fixed θ, β has a

$MVN\left\{(\theta^{-1}X'X+\phi^{-1}V^{-1})^{-1}(\theta^{-1}X'y+\phi^{-1}V^{-1}\beta_o),[\theta^{-1}X'X+\phi^{-1}V^{-1}]^{-1}\right\}$ posterior distribution.

If $L\beta$ is a set of q linearly estimable functions, then for fixed ϕ the posterior distriubtion of $L\beta$ is clearly

$$MVN\left\{L(X'X + \theta\phi^{-1}V^{-1})^{-1}(X'y+\theta\phi^{-1}V^{-1}\beta_o),\ L[\phi^{-1}V^{-1}+\theta^{-1}X'X]^{-1}L'\right\}$$

If in the prior distribution of β (or more rigorously $L\beta$) we let $\phi^{-1} = \epsilon$ and let $\epsilon \to 0$ then the prior distribution of β tends to the non-informative prior and hence the posterior distribution of $L\beta$ for fixed θ becomes

$$MVN\left\{LRX'y,\ \phi LRL'\right\}$$

where $R = \lim_{\epsilon \to 0} [X'X+\theta\epsilon V^{-1}]^{-1}$

But it is shown in Albert (1) that R is the pseudo-inverse[2] of X'X, hence the distribution becomes

[2] We write $R = (X'X)^*$ to denote a pseudo-inverse--see **footnote** 1.

$$\text{MVN} \quad L(X'X)^{*}X'\underline{y}, L(X'X)^{*}L'\theta$$

Since $X)X'X)^{-}X'$ is unique for any generalized inverse and since $(X'X)^{*}$ is a generalized inverse it follows that for fixed θ the posterior distribution of $L\underline{\beta}$ is

$$\text{MVN}(L\underline{b}, L(X'X)^{-}L'\theta)$$

in agreement with the results of Section 3. The above development clearly has merit in its simplicity but tends to obscure certain connections with classical theories of importance due to their widespread use. It is of interest to note that Raiffia and Schalifer (6) (Chapter 13) indicate the importance of the uniqueness of the projection $X\underline{b}$ in the analysis of the general linear model although relation to the classical formulae of least squares is not discussed.

5. References

1. Albert, A., An Introduction and Beginner's Guide to Matrix Pseudo-Inverses, Arcon Report, Lexington, Massachusetts.

2. Box, G. E. P. and Tiao, G., Multiparameter Problems from a Bayesian Point of View, Annals of Mathematical Statistics, Sept., 1965.

3. Geisser, S. and Cornfield, J., Posterior Distributions for Multivariate Normal Parameters J.R.S.S., B, 25, pp. 368-376.

4. Graybill, T., An Introduction to Linear Statistical Models, McGraw-Hill, 1960.

5. Lindley, D. V., Probability and Statistics from a Bayesian Point of View, Cambridge University Press, 1964.

6. Raiffia and Schlaifer, Applied Statistical Decision Theory.

7. Rao, C. R. Linear Statistical Inference and Its Applications. John Wiley and Sons.