

Comment

Terry Speed

Geoff Robinson is to be congratulated for writing this paper. It is lucidly written, it bridges a number of gulfs that have developed in our subject, and it is provocative. That he wrote it is clearly a Good Thing! I welcome the opportunity to say this and to make a few remarks that he might have made. I believe that these remarks will strengthen his already strong case for a much more explicit recognition of the role of BLUPs in our subject.

1. THE BAYESIAN DERIVATION

In Section 4.2 Robinson describes a Bayesian derivation, stating that the posterior mode is given by the BLUP estimates when β is regarded "as a parameter with a uniform, improper prior distribution and u as a parameter which has a prior distribution which has mean zero and variance $G\sigma^2$, independent of β ." All this is certainly true, but it may be helpful to add that if β is given a proper prior (normal) distribution with mean zero and variance $B\sigma^2$, say, with u as before, then all of the results one could possibly want (posterior means, posterior variances, etc.) can be derived straightforwardly by the standard Bayesian formulae. Then all one has to do to derive the corresponding BLUP formulae is let $B^{-1} \rightarrow 0$. An identity which I have found useful, perhaps even indispensable, for carrying out this last step, is discussed in de Hoog, Speed and Williams (1990). Note that the approach just described is essentially that adopted in Dempster, Rubin and Tsutakawa (1981).

2. FORMULAE FOR \hat{u}

The only actual formulae given in the paper for \hat{u} in the general case is the rather complicated one in Section 4.3. This is a pity, because there is an obvious "plug-in" expression, namely

$$(1) \quad \hat{u} = GZ^T V^{-1}(y - X\hat{\beta}),$$

where $V = ZGZ^T + R$. This may be viewed as the result of regressing u on y , with the mean $X\beta$ of y replaced by its obvious linear estimator.

A variant of (1) is

$$(1') \quad \hat{u} = (Z^T R^{-1} Z + G^{-1})^{-1} Z^T R^{-1}(y - X\hat{\beta}).$$

Terry Speed is Professor, Department of Statistics, University of California, Berkeley, California 94720.

The simpler formulae (5.3) and (5.4) arising when there are no fixed effects also have more general analogues, namely

$$(2) \quad (Z^T A Z + G^{-1}) \hat{u} = Z^T A y,$$

where $A = R^{-1}(I - S)$, $S = P_{\mathcal{R}(X)}^R$ being the projector onto $\mathcal{R}(X)$ orthogonal with respect to $\langle a, b \rangle = a^T R^{-1} b$, and for the variance-covariance matrix of \hat{u} :

$$(3) \quad \{G^{-1} - (Z^T A Z + G^{-1})^{-1}\} \sigma^2.$$

These expressions can be derived readily using the Bayesian approach outlined in (1) above, together with the matrix identity already referred to. I note in passing that Robinson's formulae (5.4) is in fact the variance-covariance matrix of $\hat{u} - u$, not, as stated, of \hat{u} .

3. SOLVING THE BLUP EQUATIONS

Perhaps in order to avoid messy algebra, Robinson has said little about the actual solution of the BLUP equations. I know that he has worked on this problem with some enormous data sets, and so I am hesitant to comment here. However, it does seem worthwhile to make one easy point, in order to connect this topic with another, closely related one. The obvious rearrangement of the first equation in (1.2),

$$(4) \quad X^T R^{-1} X \hat{\beta} = X^T R^{-1}(y - Z\hat{u}),$$

can be combined with either (1) or (1') above, to form the basis of an iterative solution of the BLUP equations, provided, of course, that the separate problems are readily solved. Just such a strategy is recommended more generally in Green (1985) in the context of smoothing, a topic to which I shall return.

It is also worth pointing out that (1') or (2) is to be preferred when G^{-1} has simple structure, whereas if G is simple and V is readily inverted, (1) is more useful. In many animal breeding problems it is G^{-1} which has the simpler structure, as it also does in the Kalman filter case.

4. REML AND BLUP

In Section 5.4 Robinson states that "REML is the method of estimating variance components that seems to have the best credentials from a Classical

viewpoint." What he does not say, which should be of interest to readers of his paper, is that REML and BLUP are intimately connected. Indeed one view—certainly not the only one—of the REML equations for variance components is that they are simply equating *observed* with *expected* sums of squares of BLUPs. This observation goes back to the original paper by Patterson and Thompson (1971; see also Harville, 1977) and can be concisely stated within Robinson's framework as follows.

Suppose that $Z = [Z_1 : \cdots : Z_c]$ is blocked, corresponding to c random effects, with Z_i being $n \times q_i$, $i = 1, \dots, c$, $u = (u_1 : \cdots : u_c)$ is similarly blocked into q sets of random effects, and finally $G = \text{diag}(G_1, \dots, G_c)$ is diagonally blocked with $G_i = \gamma_i I_{q_i}$, where $\sigma_i^2 = \gamma_i \sigma^2$ is the variance of each independent component of the i th random effect u_i . It is also convenient to denote e by u_0 , put $Z_0 = I_n$ and $\gamma_0 = 1$.

With this notation the REML equations take the form

$$(5) \quad y^T \left\{ V^{-1} \bar{Q} \frac{\partial V}{\partial \sigma_i^2} V^{-1} \bar{Q} \right\} y = \text{tr} \left\{ \frac{\partial V}{\partial \sigma_i^2} V^{-1} \bar{Q} \right\}$$

$i = 0, \dots, c$, where $Qy = X\hat{\beta}$ and $\bar{Q} = I - Q$. (By contrast, the ML equations have no \bar{Q} term in the right-hand expression.)

Turning now to BLUPs in this context, they are (in the form (1) above)

$$(6) \quad \hat{u}_i = \gamma_i Z_i^T V^{-1} \bar{Q} y$$

$i = 0, \dots, c$, and

$$(7a) \quad \text{var}(\hat{u}_i) = (G_i - U_i) \sigma^2$$

$i = 1, \dots, c$, where U_i is the i th diagonal block of the matrix $(Z^T A Z + G^{-1})^{-1}$. Furthermore,

$$(7b) \quad \text{var}(\hat{e}) = V^{-1} \bar{Q} \sigma^2 = (A - A Z U Z^T A) \sigma^2$$

where $A = R^{-1} \bar{S}$ was defined earlier, and $U = (Z^T A Z + G^{-1})^{-1}$. If we write $p_i = \gamma_i^{-1} \text{tr}(U_i)$, $i = 1, \dots, c$, then it follows that for $i = 1, \dots, c$

$$(8a) \quad \mathbb{E} |\hat{u}_i|^2 = (q_i - p_i) \sigma_i^2$$

and

$$(8b) \quad \mathbb{E} |\hat{e}|^2 = \left[(n - p) - \sum_1^c (q_i - p_i) \right] \sigma^2.$$

Now the striking thing is this: the REML equations (5) can rather easily be manipulated into a form just like (8a) and (8b), with the expectation symbol \mathbb{E} omitted. Although this is not necessarily the best way to solve these equations, the repeated calculation of BLUPs and then updating the vari-

ance components is one simple iterative scheme which works quite well.

5. PENALIZED LEAST SQUARES

Suppose that we regard (1.1) as an ordinary ("fixed effects") linear model, and that we wished to estimate β and u by R -weighted least squares with a "penalty" $u^T G^{-1} u$ being added to the sum of squares term being minimized. Then we would obtain just the expression given in Section 4.1, which Henderson minimized. Such penalties are added for many reasons: to smooth, to improve the condition of the matrix to be inverted, and so on, and it has long been recognized that this is a way of making one's linear model "quasi-Bayesian." More precisely, it turns the standard least squares problem into a case of BLUP. This practice has a long history, dating back at least to Whittaker (1923).

6. SMOOTHING SPLINES ARE BLUPS

Continuing with the theme of the previous remark, let us see how the smoothing splines popularized by G. Wahba (see her 1990 monograph for a comprehensive exposition) are just BLUPs. This observation corrects the terminology which has been used in the spline literature for over a decade, for the Bayesian interpretation of the smoothing spline—with a partially improper prior—is just the statement heading this section.

It is simplest to deal with cubic smoothing splines on the interval $[0, 1]$. If the observations are taken at $0 \leq t_1 < \cdots < t_n \leq 1$, and are

$$y_i = g(t_i) + \epsilon_i,$$

$i = 1, \dots, n$, where g is an unknown smooth function, then the function g_λ which minimizes

$$n^{-1} \sum_1^n (y_i - g(t_i))^2 + \lambda \int_0^1 \{g''(u)\}^2 du$$

over a suitable class $G_0 \oplus G_1$ of functions, has the values

$$\hat{y} = (g_\lambda(t_i))$$

$$(9) \quad = X(X^T V^{-1} X)^{-1} X^T V^{-1} y + Q_n V^{-1} (I - X(X^T V^{-1} X)^{-1} X^T V^{-1}) y,$$

where

$$X = (t_i^{k-1}), \quad i = 1, \dots, n; \quad k = 1, 2;$$

$$Q_n(i, j) = Q(t_i, t_j), \quad 1 \leq i, j \leq n;$$

and

$$Q(s, t) = \int_0^1 (s-w)_+ (t-w)_+ dw, \quad 0 \leq s, t \leq 1.$$

It is easy to check that (9) is just the fitted value

$$\hat{y} = X\hat{\beta} + Z\hat{u},$$

where, in Robinson's notation, $\hat{\beta}$ and \hat{u} are the BLUPs, X is as given above, $Z = I_n = R$, and $G = (n\lambda)^{-1}Q_n$.

Certainly there is more to smoothing splines than BLUPs; for example, estimates of the value of the function g at values of t other than those observed, but in many applications (9) and related expressions are all that is required.

By now it should come as no surprise to hear that the technique termed generalized maximum likelihood (GML) for estimating the smoothing parameter λ is no other than REML in this BLUP problem. This is readily checked by comparing formulae in this paper with ones in Wahba (1990).

With only very few changes, the identification just made to show that smoothing splines are BLUPs shows that the model robust response surface designs of Steinberg (1985) are also BLUPs. In this case the u term corresponds to sums of tensor products of orthogonal polynomials.

7. LINEAR SMOOTHERS ARE ALMOST BLUPS

There is a sense in which all linear smoothers (see Buja, Hastie and Tibshirani, 1989) are intimately related to BLUPs. A typical linear smoother S satisfies $S^n \rightarrow T$ as $n \rightarrow \infty$, where T is idempotent. This corresponds to a projector onto the subspace $\mathcal{R}(X)$ in Robinson's model (1.1), and so Ty corresponds to $X\hat{\beta}$. Thus $(S - T)y$ corresponds to $Z\hat{u}$, and in some situations it is even possible to construct a covariance matrix V such that this correspondence is precise. Furthermore, many smoothers S have form $S(\lambda)$, where λ is a parameter (bandwidth, variance ratio, smoothness penalty, etc.) that defines a family of similar smoothers. In such cases $S(\infty)$ often has the form $T + W$, where

W is another projector, while $S(0) = T$. Many of the problems and the formulae in the theory of linear smoothers are analogues of ones arising in the theory of BLUPs.

8. INTERVAL ESTIMATES INVOLVING BLUPS

In Section 5.6 Robinson briefly alludes to work done on estimating the precision of BLUP estimates when uncertainty in the dispersion parameter is taken into account. This general problem, and in particular the assignment of interval estimates, has attracted a lot of attention in the literature on smoothing splines (see, e.g., Nychka, 1988, for a recent review). Much concern has been given to the question of what, if any, coverage properties can be expected of a "Bayesian" posterior interval. Making interval statements about an object which is an estimate of the sum of fixed and random effects is bound to cause problems of interpretation to many people, and I would be interested to hear Geoff Robinson's comments on this point. I know that he has studied these matters closely in the past.

9. SUMMARY

In closing these few remarks, I cannot resist paraphrasing I. J. Good's memorable aphorism: "To a Bayesian, all things are Bayesian." How does "To a non-Bayesian, all things are BLUPs" sound as a summary of this fine paper?

ACKNOWLEDGMENT

The author gratefully acknowledges partial support from NSF Grant DMS-88-02378, and the hospitality of the Department of Statistics, IAS, Australian National University whilst preparing this comment.