That BLUP is a Good Thing: The Estimation of Random Effects
Author(s): G. K. Robinson
Source: *Statistical Science*, Vol. 6, No. 1 (Feb., 1991), pp. 15-32
Published by: Institute of Mathematical Statistics
Stable URL: https://www.jstor.org/stable/2245695
Accessed: 01-11-2019 07:01 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve and extend access to *Statistical Science*

This content downloaded from 14.139.222.72 on Fri, 01 Nov 2019 07:01:24 UTC
All use subject to https://about.jstor.org/terms

Statistical Science
1991, Vol. 6, No. 1, 15–51

That BLUP Is a Good Thing: The Estimation of Random Effects

G. K. Robinson

Abstract. In animal breeding, Best Linear Unbiased Prediction, or BLUP, is a technique for estimating genetic merits. In general, it is a method of estimating random effects. It can be used to derive the Kalman filter, the method of Kriging used for ore reserve estimation, credibility theory used to work out insurance premiums, and Hoadley's quality measurement plan used to estimate a quality index. It can be used for removing noise from images and for small-area estimation. This paper presents the theory of BLUP, some examples of its application and its relevance to the foundations of statistics.

Understanding of procedures for estimating random effects should help people to understand some complicated and controversial issues about fixed and random effects models and also help to bridge the apparent gulf between the Bayesian and Classical schools of thought.

Key words and phrases: Best linear unbiased prediction (BLUP), estimation of random effects, fixed versus random effects, foundations of statistics, likelihood, selection index, Kalman filtering, parametric empirical Bayes methods, small-area estimation, credibility theory, ranking and selection.

1. INTRODUCTION

The acronym BLUP stands for "Best Linear Unbiased Prediction" and is in common usage in animal breeding. It is a method of estimating random effects.

The context of BLUP is the linear model

$$y = X\beta + Zu + e \tag{1.1}$$

where y is a vector of n observable random variables, β is a vector of p unknown parameters having fixed values (fixed effects), X and Z are known matrices, and u and e are vectors of q and n, respectively, unobservable random variables (random effects) such that $E(u) = 0$, $E(e) = 0$ and

$$\mathrm{Var}\begin{bmatrix} u \\ e \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}\sigma^2$$

where G and R are known positive definite matrices and σ^2 is a positive constant.

At times, we will discuss the estimation of dispersion parameters and will use θ to denote a vector of dispersion parameters on which the matrices G and R depend. Generally, it will be assumed that the variance-covariance structure is known except perhaps for the single parameter σ^2.

BLUP estimates of the realized values of the random variables u are *linear* in the sense that they are linear functions of the data, y; *unbiased* in the sense that the average value of the estimate is equal to the average value of the quantity being estimated; *best* in the sense that they have minimum mean squared error within the class of linear unbiased estimators; and *predictors* to distinguish them from estimators of fixed effects. A convention has somehow developed that estimators of random effects are called *predictors* while estimators of fixed effects are called *estimators*. As discussed in Section 7.1, I prefer to use the term "estimators" for both fixed and random effects.

Mathematically, the BLUP estimates $\hat{\beta}$ of β and $\hat{u}$ of u are defined as solutions to the following simultaneous equations which were given by Henderson (1950), although in summation rather than matrix form:

$$\begin{aligned} X^TR^{-1}X\hat{\beta} + X^TR^{-1}Z\hat{u} &= X^TR^{-1}y \\ Z^TR^{-1}X\hat{\beta} + (Z^TR^{-1}Z + G^{-1})\hat{u} &= Z^TR^{-1}y. \end{aligned} \tag{1.2}$$

These equations have sometimes been called

G. K. Robinson is Principal Research Scientist, CSIRO, Division of Mathematics and Statistics, Private Bag 10, Clayton Victoria 3168, Australia.

This content downloaded from 14.139.222.72 on Fri, 01 Nov 2019 07:01:24 UTC
All use subject to https://about.jstor.org/terms

"mixed model equations," and $\hat{\beta}$ and $\hat{u}$ referred to as "mixed model solutions." Note that as G^{-1} tends to the zero matrix these equations tend formally to the generalized least-squares equations for estimating β and u when the components of u are regarded as fixed effects.

Henderson (1975) showed that provided X is of full rank, p, the variance-covariance matrix of estimation errors is

$$E\left\{\begin{bmatrix}\hat{\beta}-\beta\\ \hat{u}-u\end{bmatrix}\begin{bmatrix}\hat{\beta}-\beta\\ \hat{u}-u\end{bmatrix}^T\right\} = \begin{bmatrix}X^TR^{-1}X & X^TR^{-1}Z\\ Z^TR^{-1}X & Z^TR^{-1}Z+G^{-1}\end{bmatrix}^{-1}\sigma^2.$$

That BLUP estimates generally differ from the generalized least squares estimates that would be used if u were regarded as fixed is illustrated by the following example.

EXAMPLE. A simple example of model (1.1) is that of first lactation yields of dairy cows with sire additive genetic merits being treated as random effects (u) and herd effects being treated as fixed effects (β). The matrix $R\sigma^2$ is the variance-covariance matrix of the vector e of departures from a model in which yield was explicable entirely by sire effects and herd effects. The matrix R will be taken to be the identity matrix. Assume that the matrix G is a known multiple of the identity matrix, say $0.1I$. This would be a reasonable assumption provided that the sires were unrelated and provided that the variance ratio had been estimated previously.

Suppose that we had records as follows.

Herd	Sire	Yield
1	A	110
1	D	100
2	B	110
2	D	100
2	D	100
3	C	110
3	C	110
3	D	100
3	D	100

Then the entities in equation (1.1) are

$$y = (110, 100, 110, 100, 100, 110, 110, 100, 100)^T,$$

$$\beta = (h_1, h_2, h_3)^T$$

where h_i is the environmental effect of the ith herd,

$$u = (s_A, s_B, s_C, s_D)^T,$$

where s_j is the effect of the jth sire on his daughters' lactation yields,

$$X = \begin{pmatrix}1&0&0\\1&0&0\\0&1&0\\0&1&0\\0&1&0\\0&0&1\\0&0&1\\0&0&1\\0&0&1\end{pmatrix}$$

and

$$Z = \begin{pmatrix}1&0&0&0\\0&0&0&1\\0&1&0&0\\0&0&0&1\\0&0&0&1\\0&0&1&0\\0&0&1&0\\0&0&0&1\\0&0&0&1\end{pmatrix}.$$

Equation (1.2) gives us

$$\begin{pmatrix}2&0&0&1&0&0&1\\0&3&0&0&1&0&2\\0&0&4&0&0&2&2\\1&0&0&11&0&0&0\\0&1&0&0&11&0&0\\0&0&2&0&0&12&0\\1&2&2&0&0&0&15\end{pmatrix}\begin{pmatrix}\hat{h}_1\\ \hat{h}_2\\ \hat{h}_3\\ \hat{s}_A\\ \hat{s}_B\\ \hat{s}_C\\ \hat{s}_D\end{pmatrix} \tag{1.3}$$

$$= \begin{pmatrix}210\\310\\420\\110\\110\\220\\500\end{pmatrix},$$

which has solution

$$\hat{\beta} = (105.64, 104.28, 105.46)^T,\quad \hat{u} = (0.40, 0.52, 0.76, -1.67)^T. \tag{1.4}$$

If sire effects were treated as fixed, then equation (1.2) would be changed by omission of G^{-1}. This means that the last four diagonal elements in the left-hand-side matrix of equation (1.3) would be reduced by 10. The matrix equation for $\hat{\beta}$ and $\hat{u}$ would no longer be of full rank, but a solution can

This content downloaded from 14.139.222.72 on Fri, 01 Nov 2019 07:01:24 UTC
All use subject to https://about.jstor.org/terms

be obtained by setting, arbitrarily, $s_D = 0$. This solution is

$$\hat{\beta} = (100, 100, 100)^T, \quad \hat{u} = (10, 10, 10, 0)^T. \tag{1.5}$$

The solution given by (1.5) is the least-squares solution with which most statisticians are well acquainted. Intuitively, each of the sires other than D has daughters that yield 10 units more on average than the daughters of sire D to which they can be directly compared.

The BLUP solution given by (1.4) takes into account the information that sire effects have less variation than the variance of lactation yields from daughters of a single sire. The extent to which a sire's estimated genetic merit is regressed toward the mean depends on the amount of information available concerning that sire. For instance, sire C is estimated to be better than sires A and B because more is known about him—the lactation yields of his daughters are the same (110) as those of sires A and B.

The variance-covariance matrix of the estimates from the mixed model is σ^2 times the inverse of the left-hand-side matrix in equation (1.3). The diagonal elements of the inverse matrix for s_A, s_B, s_C and s_D are 0.0954, 0.0941, 0.0916 and 0.0833, respectively. Since the merit of a sire about which nothing was known would have a variance of $0.1\sigma^2$, there has been little gain in precision of sire effect estimates due to data on lactations of daughters.

REMARKS. In this example, numbers with few significant digits have been used in order to make the example easier to follow. Consequently, variance parameters should not be estimated from the given data. In practical situations, the variance ratio that was taken to be 0.1 or the variance σ^2 may need to be estimated from the same data as is used to estimate sire genetic merits.

My introduction to the estimation of random effects was as statisician for the Australian Dairy Herd Improvement Scheme in mid-1980. This means that I think first of the estimation of genetic merits of dairy cattle when I think of estimating random effects. Readers might like to allow for this point of view.

2. OBJECTIVES

In a discussion at the Royal Statistical Society, Dawid (1976) remarked

> A constant theme in the development of statistics has been the search for justification for what statisticians do. To read the textbooks, one might get the distorted idea that 'Student' proposed his t-test because it was the Uniformly Most Powerful Unbiased test for a Normal mean, but it would be more accurate to say that the concept of UMPU gains much of its appeal because it produces the t-test, and everyone knows that the t-test is a good thing.

The words "a good thing" in the title of this paper are to be interpreted as coming from this quotation. I wish to argue that the BLUP method for estimating random effects is "a good thing" just as Student's t-test is "a good thing."

I believe that the Classical school of thought in statistical inference should accept estimation of random effects as a legitimate activity. This theme will be developed in Section 4.3, which gives a classical justification for BLUP, and in Section 6, which lists applications. If estimation of random effects were accepted as legitimate by the Classical school, then the Bayesian and Classical schools of thought in statistics would differ less than much current rhetoric suggests.

Another objective is to encourage communication between people who deal with the various applications where random effects are estimated. The 50th Anniversary Conference, Iowa State Statistical Laboratory, encouraged such communication. See Harville (1984). Much theory has been developed separately in each of several areas of application and further theoretical work in each area might be assisted by looking at other fields. The computing problems associated with estimating random effects might also be alleviated by learning about methods used in other areas of application. See also Kackar and Harville (1984) and Robinson and Jones (1987) on the computational problems of estimating standard errors.

Another objective of this paper is to ask people to question the meanings of some fundamental statistical ideas. These include unbiasedness, likelihood, and the distinction between fixed and random effects. They will be discussed further in Section 7.

3. STRUCTURE

Section 1 of this paper has introduced BLUP and the estimation of random effects without justifying the mathematical formulae used. Section 4 presents some basic theory on estimation of random effects assuming θ is known. This shows that BLUP can be derived in many different ways and is robust with respect to philosophy of statistics. Section 5 discusses the relationship between estimation of random effects and other theoretical ideas. Its

This content downloaded from 14.139.222.72 on Fri, 01 Nov 2019 07:01:24 UTC
All use subject to https://about.jstor.org/terms

purpose is to show that understanding the estimation of random effects can help with the understanding of other theory. Section 6 reviews applications involving estimation of random effects. It shows that many groups of people are estimating random effects and that it makes sense. Section 7 reviews some fundamental ideas about statistics, suggesting that an understanding of estimation of random effects should influence our approaches and attitudes.

4. DERIVATIONS OF BLUP

Four derivations of BLUP are given below. Those in Sections 4.1 and 4.2 require the assumption of normality. Those in Sections 4.3 and 4.4 do not require normality as they only use first and second moments.

4.1 Henderson's Justification

Henderson (1950) described the BLUP estimates (1.2) as being "joint maximum likelihood estimates." Henderson (1973, page 16) explained that his derivation had actually been to assume that u and e are normally distributed and to maximize the joint density of y and u with respect to β and u. He suggests that this should not be called "maximum likelihood" because the function being maximized is not a likelihood.

The joint density of y and u is

$$(2\pi\sigma^2)^{-\frac{1}{2}n-\frac{1}{2}q}\left(\det\begin{bmatrix} G & 0 \\ 0 & R\end{bmatrix}\right)^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2\sigma^2}\begin{pmatrix} u \\ y - X\beta - Zu\end{pmatrix}^T \cdot\begin{bmatrix} G & 0 \\ 0 & R\end{bmatrix}^{-1}\begin{pmatrix} u \\ y - X\beta - Zu\end{pmatrix}\right\}. \tag{4.1}$$

To maximize this with respect to β and u requires minimizing

$$\begin{pmatrix} u \\ y - X\beta - Zu\end{pmatrix}^T\begin{bmatrix} G & 0 \\ 0 & R\end{bmatrix}^{-1}\begin{pmatrix} u \\ y - X\beta - Zu\end{pmatrix} = u^TG^{-1}u + (y - X\beta - Zu)^TR^{-1}(y - X\beta - Zu).$$

Differentiating this with respect to β and u using the usual rules for vector differentiation of scalar functions and equating the derivatives to zero gives Henderson's mixed model equations (1.2).

4.2 Bayesian Derivation

A Bayesian derivation of BLUP is straightforward. Regard β as a parameter with a uniform, improper prior distribution and u as a parameter which has a prior distribution that has mean zero and variance $G\sigma^2$, independent of β. Given β and u, the density of y is

$$(2\pi\sigma^2)^{-\frac{1}{2}n}\det(R)^{-\frac{1}{2}} \cdot \exp\left\{-\frac{1}{2\sigma^2}(y - X\beta - Zu)^TR^{-1} \cdot (y - X\beta - Zu)\right\}.$$

The prior density is

$$(2\pi\sigma^2)^{-\frac{1}{2}q}\det(G)^{-\frac{1}{2}}\exp -\frac{1}{2\sigma^2}(u^TG^{-1}u).$$

Therefore the posterior density for β and u is proportional to expression (4.1), and so the posterior mode is given by the BLUP estimates.

Dempfle (1977) gave a Bayesian presentation along these lines. Lindley and Smith (1972) presented a derivation which is equivalent to this.

It is generally true that Bayesian procedures are not affected by a stopping rule, provided that the stopping rule depends only on the data included in the analysis. This can be of substantial consolation in some applications. In estimating genetic merits of animals, mating and culling decisions depend on available information. Henderson (1965) investigated the conditions under which BLUP estimates are unbiased despite selection. In geostatistics, decisions about where to drill are based on data available at the time.

4.3 Within the Classical School

The simplest case. The simplest case of estimation of random effects is in the estimation of residuals from a simple normal model.

Suppose that n observations are taken from a normal population which has mean μ and variance σ^2 known to be 1. If the observations are $X_1, X_2, \ldots, X_n$ with mean $\overline{X}$, then it is common to estimate the parameter μ by $\overline{X}$. Other estimates may be used if robustness in some sense is required, but we will here assume that $\overline{X}$ is the most desirable estimator.

The model could be written in the form

$$X_i = \mu + e_i,$$

where e_i is the error associated with the ith observation and comes from a standard normal distribution. These errors are also called residuals, being what is left of the observational data after the deterministic component is removed. Now, we might wish to ask: "What is the best estimate of e_i?" The obvious estimate of the residual e_i is $\hat{e}_i = X_i - \overline{X}$.

This content downloaded from 14.139.222.72 on Fri, 01 Nov 2019 07:01:24 UTC
All use subject to https://about.jstor.org/terms

Properties of the residuals as estimators of the unknown errors are the following.

1. They are linear in the data.
2. They are unbiased in the sense that

$$E[\hat{e}_i] = E[e_i].$$

   Note however that $E[\hat{e}_i \mid e_i]$ is nearer to zero than is e_i. In some circumstances, people tend to expect 'unbiased' to be interpretable as meaning that the expectation of an estimate of a random effect given the true value of the random effect is equal to that random effect. This is not the case.
3. They have minimum mean square error amongst the class of linear unbiased estimators.

In this very simple case, none of this is very interesting or enlightening; but it is notable that there is a situation where estimation of random effects is standard practice. We need to consider situations involving more than one source of variation before anything nontrivial happens. However, it does suggest that it is reasonable to ask that estimates of random effects be linear, unbiased and minimum mean square error.

The general case. Henderson (1963) showed using Lagrange multipliers that BLUP estimates of linear combinations of fixed and random effects are the estimates that satisfy the classical requirements of being linear, unbiased and minimum mean square error. Harville (1976) showed, further, that the Gauss–Markov theorem could be extended to cases when matrices G and R are of less than full rank. See also Ishii (1969, Example 2, pages 482–487).

A more intuitive approach to showing that the BLUP estimates have minimum mean squared error within the class of linear unbiased estimates was given by Harville (1990). First, note that

$$E[yy^T] = X\beta\beta^T X^T + ZGZ^T\sigma^2 + R\sigma^2,$$
$$E[uy^T] = GZ^T\sigma^2$$

and that, as Henderson, Kempthorne, Searle and von Krosigk (1959) showed, an alternative form for the BLUP estimates is

$$\hat{\beta} = \left\{X^T(R + ZGZ^T)^{-1}X\right\}^{-1}X^T(ZGZ^T + R)^{-1}y$$
$$\hat{u} = (Z^TR^{-1}Z + G^{-1})^{-1}\Big[Z^TR^{-1} - Z^TR^{-1}X \cdot \left\{X^T(R + ZGZ^T)^{-1}X\right\}^{-1} \cdot X^T(R + ZGZ^T)^{-1}\Big]y.$$

Linear unbiased estimates of zero are of the form a^Ty, where a satisfies $X^Ta = 0$. They are uncorrelated with the errors of BLUP estimates, since

$$\begin{aligned}
E[(\hat{\beta} - \beta)y^Ta] &= \left\{X^T(R + ZGZ^T)^{-1}X\right\}^{-1}X^T(ZGZ^T + R)^{-1} \cdot E[yy^T]a - \beta E[y^T]a \\
&= \beta\beta^TX^Ta + \left\{X^T(R + ZGZ^T)^{-1}X\right\}^{-1}X^T\sigma^2 a - \beta\beta^TX^Ta \\
&= 0
\end{aligned}$$

and

$$\begin{aligned}
E[(\hat{u} - u)y^Ta] &= (Z^TR^{-1}Z + G^{-1})^{-1}\Big[Z^TR^{-1} - Z^TR^{-1}X \cdot \left\{X^T(R + ZGZ^T)^{-1}X\right\}^{-1}X^T \cdot (R + ZGZ^T)^{-1}\Big]E[yy^T]a - E[uy^T]a \\
&= (Z^TR^{-1}Z + G^{-1})^{-1}Z^TR^{-1}(ZGZ^T + R)\sigma^2 a - GZ^T\sigma^2 a \\
&= (Z^TR^{-1}Z + G^{-1})^{-1}(Z^TR^{-1}Z + G^{-1}) \cdot GZ^T\sigma^2 a - GZ^T\sigma^2 a \\
&= 0.
\end{aligned}$$

Any linear unbiased estimator of a linear combination $b^T\beta + c^Tu$ of fixed and random effects must be of the form $b^T\hat{\beta} + c^T\hat{u} + a^Ty$, where $X^Ta = 0$. Its variance-covariance matrix of estimation errors is

$$\begin{aligned}
&E\Big[\{b^T(\hat{\beta} - \beta) + c^T(\hat{u} - u) + a^Ty\} \cdot \{b^T(\hat{\beta} - \beta) + c^T(\hat{u} - u) + a^Ty\}^T\Big] \\
&= E\Big[\{b^T(\hat{\beta} - \beta) + c^T(\hat{u} - u)\} \cdot \{b^T(\hat{\beta} - \beta) + c^T(\hat{u} - u)\}^T\Big] + E[a^Tyy^Ta] \\
&\quad + E[\{b^T(\hat{\beta} - \beta) + c^T(\hat{u} - u)\}y^Ta] \\
&\quad + E\Big[a^Ty\{b^T(\hat{\beta} - \beta) + c^T(\hat{u} - u)\}^T\Big] \\
&= E\Big[\{b^T(\hat{\beta} - \beta) + c^T(\hat{u} - u)\} \cdot \{b^T(\hat{\beta} - \beta) + c^T(\hat{u} - u)\}^T\Big] + E[a^Tyy^Ta].
\end{aligned}$$

Now $E[a^Tyy^Ta] = E[a^Ty\{a^Ty\}^T]$ is a symmetric positive semidefinite matrix, so this variance-covariance matrix of estimation errors exceeds that for the BLUP estimate $b^T\hat{\beta} + c^T\hat{u}$.

4.4 Goldberger's Derivation

Goldberger (1962) considered a linear model

$$y = X\beta + \varepsilon,$$

This content downloaded from 14.139.222.72 on Fri, 01 Nov 2019 07:01:24 UTC
All use subject to https://about.jstor.org/terms

where the disturbance ε satisfies $E(\varepsilon) = 0$ and $\operatorname{Var}(\varepsilon) = \Omega$. Given a new (observable) vector x_* of regressors and (unobservable) prediction disturbance ϵ_*, which is correlated with the disturbances for the data already obtained, satisfying

$$E(\varepsilon_*) = 0$$
$$E\left(\varepsilon_* \varepsilon^T\right) = w^T,$$

Goldberger's equation (3.12) tells us that the best linear unbiased predictor of the future observation $y_* = x_*^T\beta + \epsilon_*$ is

$$x_*^T\left(X^T\Omega^{-1}X\right)^{-1}X^T\Omega^{-1}y + w^T\Omega^{-1}y - w^T\Omega^{-1}X\left(X^T\Omega^{-1}X\right)^{-1}X^T\Omega^{-1}y.$$

For our model, $\epsilon = Zu + e$, so

$$\Omega = \left(ZGZ^T + R\right)\sigma^2.$$

To estimate $x_*^T\beta + z_*^T u$ take

$$\epsilon_* = z_*^T u$$

and hence

$$w^T = E\left[z_*^T u(Zu + e)^T\right] = z_*^T G Z^T \sigma^2.$$

So Goldberger's derivation tells us that the best linear unbiased predictor of $x_*^T\beta + z_*^T u$ is

$$\begin{aligned} & x_*^T\left[X^T\left(ZGZ^T + R\right)^{-1}X\right]^{-1}X^T\left(ZGZ^T + R\right)^{-1}y \\ & \quad + z_*^T G Z^T\left(ZGZ^T + R\right)^{-1}y \\ & \quad - z_*^T G Z^T\left(ZGZ^T + R\right)^{-1} \\ & \quad \cdot X\left[X^T\left(ZGZ^T + R\right)^{-1}X\right]^{-1} \\ & \quad \cdot X^T\left(ZGZ^T + R\right)^{-1}y, \end{aligned}$$

which is

$$x_*^T\hat{\beta} + z_*^T G Z^T\left(ZGZ^T + R\right)^{-1}\left(y - Z\hat{\beta}\right).$$

Using the matrix identity (5.2), below, and the second of the simultaneous equations (1.2), this is

$$\begin{aligned} & x_*^T\hat{\beta} + z_*^T\left(Z^T R^{-1} Z + G^{-1}\right)^{-1} Z^T R^{-1}\left(y - X\hat{\beta}\right) \\ & \quad = x_*^T\hat{\beta} + z_*^T\hat{u}. \end{aligned}$$

Thus Goldberger's predictor is the same as that given by (1.2).

To the best of my knowledge, Goldberger was the first to use the term "best linear unbiased predictor" and Henderson started using the acronym BLUP in 1973.

Goldberger's derivation seems unobjectionable from a Classical viewpoint. His emphasis is on prediction, but his formulae still apply generally to prediction of a future observation, y_*, which is perfectly correlated with a past disturbance, so they do apply to estimation of random effects.

5. LINKS WITH OTHER STATISTICAL THEORY

5.1 Recovery of Inter-Block Information

Henderson, Kempthorne, Searle and von Krosigk (1959, page 196) showed that the BLUP estimate $\hat{\beta}$ is identical with the generalized least-squares estimate of β that would be obtained after recovery of inter-block information if the random effects u were block effects. They eliminated $\hat{u}$ from equation (1.2), giving

$$\begin{aligned} & X^T R^{-1} X\hat{\beta} \\ & \quad - X^T R^{-1} Z\left(Z^T R^{-1} Z + G^{-1}\right)^{-1} Z^T R^{-1} X\hat{\beta} \\ & = X^T R^{-1} y - X^T R^{-1} Z\left(Z^T R^{-1} Z + G^{-1}\right)^{-1} Z^T R^{-1} y. \end{aligned}$$

Now using a matrix identity which is commonly used in this subject area

$$\begin{aligned} & \left(R + ZGZ^T\right)^{-1} \\ & \quad = R^{-1} - R^{-1} Z\left(Z^T R^{-1} Z + G^{-1}\right)^{-1} Z^T R^{-1}, \end{aligned} \tag{5.1}$$

gives

$$X^T\left(R + ZGZ^T\right)^{-1} X\hat{\beta} = X^T\left(R + ZGZ^T\right)^{-1} y.$$

Hence

$$\hat{\beta} = \left[X^T\left(R + ZGZ^T\right)^{-1} X\right]^{-1} X^T\left(R + ZGZ^T\right)^{-1} y,$$

which can be seen to be the generalized least squares estimate of β, since the variance-covariance matrix of the random effects is

$$E\left[(Zu + e)(Zu + e)^T\right] = R + ZGZ^T.$$

The matrix identity (5.1) is a particular case of

$$\begin{aligned} & (A + UBV)\left[A^{-1} - A^{-1}U \cdot \left(I + BVA^{-1}U\right)^{-1} BVA^{-1}\right] \\ & \quad = I + UBVA^{-1} - U\left(I + BVA^{-1}U\right)^{-1} BVA^{-1} \\ & \quad \quad - UBVA^{-1}U\left(I + BVA^{-1}U\right)^{-1} BVA^{-1} \\ & \quad = I \end{aligned}$$

of which the history and many variants, generalizations and special cases are discussed by Henderson

This content downloaded from 14.139.222.72 on Fri, 01 Nov 2019 07:01:24 UTC
All use subject to https://about.jstor.org/terms

and Searle (1981). In this paper, we will also use the matrix equality

$$(5.2)\qquad \begin{aligned}&(Z^TR^{-1}Z + G^{-1})^{-1}Z^TR^{-1}\\&\quad = GZ^T(R + ZGZ^T)^{-1},\end{aligned}$$

which can be derived from (5.1).

Recovery of interblock information is most commonly discussed for experimental data from incomplete block designs. Recovery of the interblock information improves the efficiency of the estimates of the fixed effects, but the estimates based only on intra-block information are often considered to be satisfactory.

For unbalanced data, estimates of fixed effects based only on intrablock information can sometimes be quite unsatisfactory. An example of this is presented below. It was discussed in Henderson, Kempthorne, Searle and von Krosigk (1959).

EXAMPLE. Fifty cows produce an average of 100 kilograms of butterfat in their first lactations. Forty of these cows survive to complete their second lactations. The average first lactation butterfat yield of these 40 cows is 110 kilograms and the average second lactation butterfat yield is 140 kilograms. Estimate the average difference between lactations in the absence of culling!

One answer is (140 − 100) kg = 40 kg, using all cows. Another answer is (140 − 110) kg = 30 kg, using only the cows which completed second lactations. This is the estimate that uses only intra-cow information. If the true correlation between first and second lactations can be taken to be 1/2, then recovery of inter-cow information gives the estimate 35 kg.

Intuitively, the cows culled are likely to be worse than average. Therefore the cows completing second lactations are likely to be better than average; so the first answer is likely to be too large. The cows not culled are unlikely to be as good as they appear to be because the data has been selected. An extreme case to illustrate this is that if first lactation yield and second lactation yield were uncorrelated, then culling on first lactation would not increase average second lactation yield but it would increase the first lactation yield of cows completing a second lactation by rejecting some data. The second answer is likely to be too small because the 110 is an overestimate.

In this example, the effect of lactation parity is being regarded as a fixed effect (treatment) and cow effects are being regarded as random effects (blocks). Thinking of the BLUP estimate of the lactation parity effect as being a Bayesian estimate, the likelihood principle tells us that the estimate does not need to be modified if culling has taken place, provided that culling decisions were based only on the data included in the analysis. Within the Classical framework, Henderson (1975) showed that selection and culling which is based on linear combinations (L^Ty) of the data y do not affect the optimality of the BLUP estimates provided that $L^TX = 0$. An aspect of this formalization that I cannot understand is the meaning of selection based on the L matrix. In examples given in Henderson (1973, 1975, 1984), the numbering of the random effects is always such that best is first but ranking is not a linear function of the data. The problem has been of interest to Henderson throughout his career, but his work involving the L matrix is not widely understood or accepted.

Estimates of environmental and genetic trends from dairying data tend to suffer from biases similar to this, unless the inter-cow information is recovered. Of course, the resulting estimates are sensitive to the values used for dispersion parameters, such as the correlation in the example above, but to not recover information is equivalent to using extreme values for dispersion parameters, and is worse.

5.2 Random Effects Models

Mathematically, it is easy to see from equations (1.2) and (5.2) that when there are no fixed effects the BLUP estimates of the random effects are given by

$$(5.3)\qquad (Z^TR^{-1}Z + G^{-1})\hat{u} = Z^TR^{-1}y$$

and have variance-covariance matrix

$$(5.4)\qquad (Z^TR^{-1}Z + G^{-1})^{-1}\sigma^2.$$

In animal breeding, BLUP for random effects models is known as the selection index. See Smith (1936) and Hazel (1943). Lush (1949) referred to it as "most probable producing ability."

Henderson (1963) describes BLUP as a form of selection index. In our notation, if β were known, then $y - X\beta = Zu + e$ follows a random effects model, so the best estimate of u is

$$(Z^TR^{-1}Z)^{-1}Z^TR^{-1}(y - X\beta).$$

If we replace β by the estimate

$$\hat{\beta} = \left[X^T(R + ZGZ^T)^{-1}X\right]^{-1}X^T(R + ZGZ^T)^{-1}y,$$

then the resulting estimate of u is the BLUP estimate, as shown by Henderson (1963).

This content downloaded from 14.139.222.72 on Fri, 01 Nov 2019 07:01:24 UTC
All use subject to https://about.jstor.org/terms

Box and Tiao (1968) presented a detailed derivation of BLUP estimates for random effects models within a Bayesian framework. Dempfle (1977) gave the following Bayesian derivation, which I find intuitively helpful. The idea is due to Robertson (1955) and is only applicable when the matrix Z is of full rank.

The generalized least squares estimate of u is

$$\hat{u}_1 = (Z^T R^{-1} Z)^{-1} Z^T R^{-1} y$$

and has precision

$$\operatorname{Var}(\hat{u}_1 - u) = (Z^T R^{-1} Z)^{-1}.$$

The prior estimate of u is

$$\hat{u}_2 = 0$$

and has precision

$$\operatorname{Var}(\hat{u}_2 - u) = G.$$

The best estimate of u gives these two estimates weight in inverse proportion to their precision and is

$$\begin{aligned}\hat{u} &= \left[\operatorname{Var}(\hat{u}_1 - u)^{-1} + \operatorname{Var}(\hat{u}_2 - u)^{-1}\right]^{-1} \\ &\quad \cdot \left[\operatorname{Var}(\hat{u}_1 - u)^{-1}\hat{u}_1 + \operatorname{Var}(\hat{u}_2 - u)^{-1}\hat{u}_2\right] \\ &= \left[Z^T R^{-1} Z + G^{-1}\right]^{-1} Z^T R^{-1} y.\end{aligned}$$

A straightforward classical derivation is to use standard results on the multivariate normal distribution (e.g., Searle, 1971, page 47). Since u and y have zero means and variance-covariance matrix

$$\begin{pmatrix} G & GZ^T \\ ZG & ZGZ^T + R \end{pmatrix} \sigma^2,$$

the distribution of u given y has mean

$$\begin{aligned} &GZ^T(ZGZ^T + R)^{-1} y \\ &\quad = (Z^T R^{-1} Z + G^{-1})^{-1} Z^T R^{-1} y = \hat{u} \end{aligned} \tag{5.5}$$

and variance

$$\begin{aligned} &\left[G - GZ^T(ZGZ^T + R)^{-1} ZG\right]\sigma^2 \\ &\quad = (Z^T R^{-1} Z + G^{-1})^{-1}\sigma^2 \end{aligned}$$

in agreement with (5.3) and (5.4). This derivation shows that, when there are no fixed effects to be estimated simultaneously, the theory of estimating random effects follows the theory of correlation very closely.

Ideas about correlation are quite old. Pearson (1896, page 261) wrote

> The fundamental theorems of correlation were for the first time and almost exhaustively discussed by BRAVAIS ('Analyse Mathématique sur les probabilités des erreurs de situation d'un point'. Mémoires par divers Savans, T. IX., Paris, 1846, pp 255–332) nearly half a century ago. He deals with the correlation of two and three variables ... GALTON ... introduced an improved notation ...

Random effects models are also related to the idea of *regression to the mean* attributed by Davis (1986) to Galton. The best estimate of a characteristic of an offspring given the characteristics of the parents is regressed towards the population mean from the parental average.

Example. Suppose that true intelligence quotient (IQ) is normally distributed with mean 100 and standard deviation 15. Two tests are available. Both tests give scores that are normally distributed with mean the true IQ. The first test score has standard deviation 10 given true IQ, while the second test score has standard deviation 5. A person scoring 130 on the first test would be estimated to have a true IQ of 120.8 and a person scoring 130 on the second test would be estimated to have a true IQ of 127. Features of these estimates worth noting are as follows.

- They are shrunk towards the overall mean (100) from the data. The amount of shrinkage is greater when the data point is less informative.
- They are biased given true IQ. This is obvious since the raw scores are unbiased and the estimates are nontrivial linear functions of the raw scores.
- They have zero average bias when averaged over the distribution of possible true IQs.
- The expected value of true IQ given the data is equal to the BLUP estimate of IQ, by (5.5).

This example is far from new. In the discussion to Lindley and Smith (1972), Novick suggested that Kelley (1927) was familiar with the basic ideas of shrinkage estimators. Henderson (1973, page 15) explained that considering this example was crucial in his development of BLUP in 1949.

5.3 Fixed Effects Models and Admissibility

Fixed effects models are, of course, a particular case of mixed models. Stein's (1956) demonstration that the sample mean is inadmissible for the mean of a multidimensional normal population of known variance when the dimensionality is at least three has led to some theoretical work that I believe to be

This content downloaded from 14.139.222.72 on Fri, 01 Nov 2019 07:01:24 UTC
All use subject to https://about.jstor.org/terms

of little practical value. This work is characterized by a tendency to combine unrelated estimation problems. BLUP helps us to know when to combine estimation problems. Situations where estimation problems ought to be combined are when the parameters to be estimated can be regarded as coming from some distribution. Equivalently, they are "exchangeable," or are "random effects." I agree with the view expressed by E. F. Harding in discussion of Lindley and Smith (1972) that estimates of the characteristics of butterflies in Brazil, ball bearings in Birmingham, and brussels sprouts in Belgium ought not to be related to each other.

5.4 Estimation of Variance Parameters

The estimation of variance parameters is a very extensive topic. See, for instance, Khuri and Sahai (1985). The comments below concentrate on one method of estimating variance parameters, REML, which can be interpreted as either Classical or Bayesian.

For balanced experimental data, the analysis of variance provides estimates which are often considered acceptable. Sometimes estimates of variance components are negative, in which case they are taken to be zero.

For unbalanced data, REML is the method of estimating variance components that seems to have the best credentials from a Classical viewpoint. See Robinson (1987) for a recent discussion with examples. It was expounded by Patterson and Thompson (1971). They called it "modified maximum likelihood." Some people now refer to it as "restricted maximum likelihood," while others use the term "residual maximum likelihood."

Thompson (1973) generalized REML to the multivariate case and showed that it may be used even when the data available has been selected in certain ways.

Consider the problem of estimating θ for the linear model given by equation (1.1). Bayesian statisticians would, in principle, start with a joint prior distribution for θ, β and u. If a uniform prior distribution is used for β, then the posterior mode gives a point estimate of θ and the BLUP estimates $\hat{\beta}$ and $\hat{u}$ given that θ. These are not ideal estimators. Bayesian statisticians would prefer to estimate θ by integrating over β and u rather than merely looking at the posterior for $\hat{\beta}$ and $\hat{u}$.

Harville (1974) showed that REML is equivalent to marginalizing the likelihood over the fixed effect parameters, so practical approximate Bayesian procedures for estimating dispersion parameters can use REML to approximately integrate over the fixed and random effects.

The Classical concept of modified profile likelihood due to Barndorff-Nielsen (1983) can also be regarded as an approximate Bayesian technique in which second derivatives of the posterior density are used to approximately integrate out nuisance parameters by assuming normal distributions with the given second derivatives for the nuisance parameters. Such approximate integration gives a multiplicative factor that is the exponential of the $-\frac{1}{2}$ power of the determinant of the observed information matrix for the nuisance parameters.

5.5 Estimation of Outliers

When fitting models with only one variance component, it is common practice to compute residuals in order to look for outliers. If it appears that some data points are outliers, then they may be ignored in some circumstances, they may be highlighted in other circumstances, or appropriately robust methods of analysis may be used for estimating parameters of interest.

When fitting models with two or more variance components, BLUP estimates of the realized values of random effects are the natural generalization of the concept of residuals. Outlying values of some random effects will mean that groups of data points are likely to fail to fit the model, but looking at the estimates of the random effects provides a more sensitive test for such outliers than merely looking at residuals. Fellner (1986) discusses this topic with some examples.

5.6 Estimation of Fixed and Random Effects when the Dispersion Parameters Must be Estimated

For most of this paper, estimation of random effects is considered under the assumption that θ is known. All approaches seem to agree as to the best estimates of random effects in this case. There is less consensus about what to do when θ must be estimated from the data.

Student's t-test differs from simple use of the normal distribution in that it takes the uncertainty of estimating the variance of a normal distribution into account when considering hypotheses about the mean. It is natural to suspect that BLUP estimates or their estimated precision would need to be modified when θ must be estimated.

Kackar and Harville (1981) showed that estimates $\hat{\beta}$ and $\hat{u}$ remain unbiased when θ must be estimated provided that the estimates of components of θ are translation-invariant and are even functions of the data vector. This suggests that the BLUP point estimates will generally not need to be modified. We do in principle, however, need to modify the estimated precision of the BLUP esti-

This content downloaded from 14.139.222.72 on Fri, 01 Nov 2019 07:01:24 UTC
All use subject to https://about.jstor.org/terms

mates. In practice, this difficulty is sometimes ignored or handled by conservative interpretation of calculations based on the best point estimate of the dispersion parameters.

Estimating the precision of BLUP estimates using the Classical approach seems mathematically complicated when the uncertainty in dispersion parameters is considered. For some recent work on the topic, see Kackar and Harville (1984), Harville (1985), Prasad and Rao (1986), Fuller and Harter (1987) and Jeske and Harville (1988).

From a Bayesian point of view, the estimation of dispersion parameters taking the uncertainty in the fixed and random effects into account and the estimation of fixed and random effects taking uncertainty in the dispersion parameters into account are not different in principle, but different types of approximation are likely to be appropriate. A practical approximate Bayesian procedure for estimating the precision of fixed and random effects is to approximate the Bayesian posterior distribution for θ by a discrete distribution over a small number of possibilities and to use BLUP at each of these possibilities thereby calculating a mixture of distributions as posterior for the fixed and random effects.

5.7 Empirical Bayes Methods

Empirical Bayes methods are concerned first with estimating distributions from which random effects have been generated. Once a distribution of random effects has been estimated, this distribution is used to estimate realized values of random effects using Bayes' Theorem. If the distribution of random effects is Gaussian, then empirical Bayes methods would use BLUP estimates to estimate fixed and random effects.

When parametric assumptions are made about the distribution of the random effects then the statistical methods employed are described as *parametric empirical Bayes*. BLUP is equivalent to one of the techniques of parametric empirical Bayes methodology.

5.8 Ranking and Selection

BLUP was originally developed for ranking and selection in the contexts of animal breeding and genetics. It is an appropriate technique when the ideal ranking or selection criteria involve unobservable characteristics that may be regarded as random effects.

Much theoretical work on ranking and selection seems to have been done in ignorance of BLUP. This work tries to control the probability that ranking or selection decisions will be made correctly, when the probability is to be calculated before the data are available. In contrast, BLUP is concerned with correct selection between random effects given the data, as in Berger and Deely (1988).

Stein (1945) presented a procedure that obtains an interval estimate of preassigned length for the mean of a normal distribution when the population variance is unknown. It achieves this by sampling in two stages and ignoring the information about the population variance contained in the sample variance for the second sample. It is not a satisfactory technique for practical application because it concentrates on initial precision rather than final precision, it does not achieve its nominal coverage probability conditional on the ancillary statistic s_1^2/s_2^2 where s_i^2 is the sample variance in the ith stage of sampling, and it violates the likelihood principle. Stein's procedure is a foundation stone for some theoretical work on ranking and selection that I believe to be misguided. BLUP would be a better foundation stone.

6. APPLICATIONS

There are a number of situations in which estimation of random effects is precisely what is required. Morris (1983) has discussed several of them. A brief discussion of some of them should help to establish the point that estimation of random effects is a legitimate activity, even if not very common.

There are four features which are common to most of the applications.

1. The data and the random effects to be estimated are often multivariate.
2. Computational issues are often extremely important.
3. Sparsity or approximate sparsity of partial correlation matrices is more important than sparsity or approximate sparsity of correlation matrices.
4. Random effects selected on the basis of the estimates made are often of particular interest.

People having difficulty with one of these features for any of the applications should consider how the feature is handled for the other applications.

6.1 Estimation of Merit of Individuals

Efron and Morris (1977) discussed estimating the batting abilities of 18 baseball players. The true batting abilities can be regarded as random effects drawn from a distribution. To estimate the mean and variance of the distribution of batting abilities

This content downloaded from 14.139.222.72 on Fri, 01 Nov 2019 07:01:24 UTC
All use subject to https://about.jstor.org/terms

is of some interest, but the main problem is estimating the random effects.

In plant variety trials, it is sometimes realistic to regard the varieties as random effects since they have been generated by processes that are random at the chromosome level. The objective of variety trials is generally to find the best varieties or to estimate the yield (or some other characteristic) of all varieties, not to estimate the parameters of the distribution from which the varieties are a sample.

6.2 Selection Index

In quantitative genetics, the selection index provides a way of ranking plants or animals given measurements on several traits on the individuals to be ranked and their relatives.

The selection index can be seen to be a particular case of the BLUP estimates of random effects

$$\begin{aligned} \hat{u} &= \left(Z^T R^{-1} Z + G^{-1}\right)^{-1} Z^T R^{-1} y \\ &= G Z^T \left(R + Z G Z^T\right)^{-1} y \end{aligned}$$

using (5.2). The model being fitted is

$$y = Zu + e,$$

where u is a vector of additive genetic merits of animals with variance-covariance matrix G and e is a vector of other influences including nonadditive genetic merits, permanent environmental effects and measurement errors. In this context u is often referred to as "genotype," although it only gives the so-called "additive" part of genetic merit, which is the average merit of all potential offspring for a population of potential mates. The matrix G is called the genetic variance-covariance matrix. The measurable quantities, y, are referred to as "phenotype," and $\mathrm{Var}(y) = R + ZGZ^T$ is called the phenotypic variance-covariance matrix.

Consider a situation where a trait is measured with equal precision on an animal and on one of its parents. Suppose that the variance of genotype is σ^2 and the variance from other sources is $\sigma^2(1 - h^2)/h^2$ so that the heritability (ratio of genotypic variance to total variance) is h^2. Taking

$$y = \begin{bmatrix} \text{measurement on animal} \\ \text{measurement on parent} \end{bmatrix},$$

$$Z = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

$$G = \sigma^2 \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix},$$

the genotypes of the two animals having correlation $\frac{1}{2}$, and

$$R = \frac{1 - h^2}{h^2} \sigma^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

gives

$$\hat{u} = \frac{h^2}{4 - h^4} \begin{bmatrix} 4 - h^2 & 2 - 2h^2 \\ 2 - 2h^2 & 4 - h^2 \end{bmatrix}$$

in agreement with a formula from Turner and Young (1969, page 148). This tells us how to weight the two pieces of data in order to best estimate the genetic merit of the animal. It provides the best index combining the two pieces of information on which to base selection decisions.

It is often easier to work with the BLUP form of the selection index because the matrix G^{-1} is generally sparser than G, as shown by Henderson (1976). Essentially, this sparsity arises because G^{-1} is a matrix of partial correlations, and the partial correlations of genetic merits of pairs of animals that are not mates or parent and offspring given the genetic merits of all their mates, parents and offspring is zero, as argued by Robinson (1986).

6.3 Estimation of Quality

In his many seminars on quality and management, Deming often refers to the distinction between "analytic" and "enumerative" studies as explained in Deming (1950, Chapter 7). One way of explaining this distinction is that analytic studies are concerned with estimating fixed effects while enumerative studies are concerned with estimating random effects.

Average quality over a particular time period is a random effect. The quality measurement plan (QMP) of Hoadley (1986) regards the true quality index for a period of production as having come from an probability distribution. The QMP estimates the true quality index. It is an example of BLUP where distributions other than normal are assumed.

An example of quality estimation that illustrates the distinction between estimation of fixed effects and estimation of random effects is sampling and testing of a shipload of coal or iron ore. The average grade of a large number of samples taken from a conveyor belt as the ship is loaded provides a fairly precise estimate of the average grade of the shipload. However, it may not provide a precise estimate of the average grade of all shiploads, particularly if there is substantial long-term variation in grade.

If the grade of ore as a function of cumulative tonnage, $X(t)$, is regarded as a stochastic process,

This content downloaded from 14.139.222.72 on Fri, 01 Nov 2019 07:01:24 UTC
All use subject to https://about.jstor.org/terms

then the average grade of the shipload is a random effect. It is $\int_A^B X(t)\,dt/(B-A)$ for some constants A and B that delimit the shipload. The average grade of all shiploads is a fixed effect. It is the mean of a stochastic process.

Deming would regard the sampling and testing as being an "analytic" study insofar as it is studying the mean of all shiploads and as being an "enumerative" study in so far as it is studying a single shipload.

Further details of this application appear in Saunders, Robinson, Lwin and Holmes (1989). See also Cochran (1946), Yates (1949), Jowett (1952), Duncan (1962), Matheron (1965) and Gy (1982).

It should be noted that many sampling standards are designed to estimate the population mean of a process (an analytic study—a fixed effect), not the mean of a lot (an enumerative study—a random effect), but they fail to make the distinction clear. Often, the mean of a lot can be estimated much more precisely than the process mean with given data.

6.4 Time Series and Kalman Filtering

Some time series problems are concerned with estimating fixed parameters associated with time series. Many other time series problems may usefully be regarded as problems of estimating random effects.

The observed value of a time series is the sum of signal and noise that may be regarded as random effects that differ in their spectra. Smoothing, filtering and prediction problems all involve estimation of the random effects that form the signal.

To illustrate the link between estimation of random effects and time series, we consider the Kalman filter, which is used for estimating the current value of the signal in a time series. The estimates obtained by Kalman filtering and by BLUP must be the same because of their optimality properties. However, it seems worthwhile to consider an approach to Kalman filtering directly from BLUP. See also Broemeling and Diaz (1985), Harrison and Stevens (1976) and Sallas and Harville (1981). Incidentally, Sorenson (1970) suggests that the Kalman filter is not entirely due to Kalman.

We follow the terminology of Meinhold and Singpurwalla (1983) who explained the Kalman filter in a Bayesian framework.

Suppose that unobservable vector-valued random variables u_t are related by

$$u_t = G_t u_{t-1} + w_t$$

for $t = 1, 2, \dots, n$ with $u_0 = 0$. An observable vector-valued random variable y_t is related to the u_t by

$$y_t = F_t u_t + v_t.$$

Assume that G_t and F_t are known. Further, suppose that the w_t and v_t are independent and normally distributed with zero means and known variances W_t and V_t.

This is an example of a random effects model. The variance-covariance matrix for the random effects is not simple. Denoting it by

$$G = (g_{ij})$$

where

$$g_{ij} = \mathrm{Cov}(u_i, u_j) = E[u_i u_j^T],$$

it is defined recursively by the following equations:

$$\begin{aligned} g_{11} &= W_1 \\ g_{ij} &= G_i g_{i-1j} \quad \text{for } i > j \\ g_{ij} &= g_{ij-1} G_j^T \quad \text{for } i < j \\ g_{ii} &= G_i g_{i-1i} + W_i \\ &= g_{ii-1} G_i^T + W_i. \end{aligned}$$

The matrix G^{-1} is simpler than G. It is block tridiagonal essentially because the partial correlations of pairs of nonadjacent u_t given an intermediate u_t are always zero. It is

$$G^{-1} = (g^{ij})$$

where

$$g^{nn} = W_{n+1}^{-1},$$

for $i < n$,

$$\begin{aligned} g^{ii} &= W_i^{-1} + G_{i+1}^T W_{i+1}^{-1} G_{i+1} \\ g^{ii+1} &= -G_{i+1}^T W_{i+1}^{-1} \\ g^{i+1i} &= -W_{i+1}^{-1} G_{i+1} \end{aligned}$$

and all other g^{ij} are zero matrices.

The matrix algebra required to show equivalence of BLUP estimates with the Kalman–Bucy algorithm is not trivial, but the statistical theory is simple. The Kalman-Bucy algorithm is computationally efficient because computations for time t make use of results obtained for time $t-1$. This use of the immediate past depends crucially on the simplicity of the matrix G^{-1}, which is due to the Markovian nature of the process generating the u_t. A general principle illustrated here is that partial correlation matrices are more important computationally than correlation matrices.

Note that, being equivalent to BLUP estimation

This content downloaded from 14.139.222.72 on Fri, 01 Nov 2019 07:01:24 UTC
All use subject to https://about.jstor.org/terms

of random effects, Kalman filtering is not maximum likelihood.

6.5 Removing Noise from Images

Besag (1986) reviewed many methods of attempting to restore images that have been corrupted by noise. The simplest models for continuously variable intensity over a grid are auto-normal models in which the point intensities are normally distributed and have nonzero partial correlations with intensities at neighboring points on the grid. Within our framework, the true image is a random effect that we wish to estimate. What Besag calls the "maximum *a posteriori*" estimate of the true image is equivalent to BLUP for these models.

As in other applications being discussed, the objective of primary interest is to estimate random effects.

6.6 Geostatistics

The method of geological reserve estimation known as Kriging is essentially the same as BLUP. It was developed independently from BLUP. A difference from the estimation of genetic merits is that variance parameters are estimated more frequently than in genetic applications.

The random effect that is estimated is the pattern of grade of ore as a function of position in two- or three-dimensional space. Estimation of the random effect is of much more immediate interest to a company that is considering a mining operation than is estimation of the parameters of the distribution from which the pattern of grade is a sample.

There is much more to ore reserve estimation than Kriging. A concise summary of some of the difficulties written by Matheron appears in the Foreword to Journel and Huijbregts (1978). One very substantial difficulty is that the economic feasibility of mining a deposit is based on a small number of drill-holes, but decisions about which blocks of ore will be processed and which will be dumped as waste will be made using data from blast-holes that are much more numerous than the drill-holes. The average grade of the ore that will be selected for processing is much harder to predict than is the overall average grade. This difficulty is most acute when only a small proportion of the rock to be mined will be processed.

6.7 Credibility Theory

Credibility theory is a collection of ideas used by actuaries to work out insurance premiums. As an example, consider setting workers' compensation premiums for industrial companies that have different numbers of employees and different safety records. This example was discussed by Mowbray (1914).

In terminology like that of the present paper, let u_i denote the true fair premium for company number i. The u_i may be regarded as random effects and the BLUP estimates of them will be useful. If a company has an extensive claims record, then its estimated premium will depend almost entirely on its own record. If a company has little claims record, then the estimated premium for that company will differ little from the average premium for all companies.

The classical approach to credibility is to assume that there is a credibility factor, denoted $Z(t)$, where t is the size of the risk class, such that the estimated fair premium is

$$Z(t)y + \{1 - Z(t)\}m',$$

where y is the fair premium as estimated using data for the single risk class only, and m' is the average fair premium over all risk classes. The function $Z(t)$ is chosen in a somewhat arbitrary manner which need not concern us here. (See Hickman, 1975.)

The Bayesian approach to credibility is to use conjugate prior distributions for the u_i. The development of the Bayesian approach is attributed to Bailey (1945, 1950) and Mayerson (1965). See Kahn (1982) for a brief review.

6.8 Small-Area Estimation

Small-area estimation involves using direct survey information from areas of individual interest together with information on similar or related areas. It has been found that more precise estimates can be made than if information on the other areas had been ignored. One approach is to regard the quantities of interest in the individual areas as being random effects which are to be estimated. See Battese, Harter and Fuller (1988), Prasad and Rao (1986), Fuller and Harter (1987) and the other papers in the same volume from the May 1985 International Symposium on Small Area Statistics for further information about this field of application.

7. DISCUSSION

7.1 Prediction or Estimation?

Henderson originally used the term "predictor" rather than "estimator" in order to evade criticism of BLUP. Henderson (1984, page 37) expressed

This content downloaded from 14.139.222.72 on Fri, 01 Nov 2019 07:01:24 UTC
All use subject to https://about.jstor.org/terms

doubts about the appropriateness of the terminology:

> Which is the more logical concept, prediction of a random variable or estimation of the realized value of a random variable? If we have an animal already born, it seems reasonable to describe the evaluation of its breeding value as an estimation problem. On the other hand, if we are interested in evaluating the potential breeding value of a mating between two potential parents, this would be a problem in prediction. If we are interested in future records, the problem is clearly one of prediction.

From conversations with him, I believe that he accepts the weaknesses of the terminology:

- in some applications, the thing being estimated has already occurred, and
- BLUP is a predictor only in the same way as most estimates are predictors—if they were not relevant to something which might happen in the future then they would not be of interest.

It has become common practice to "estimate" fixed effects and to "predict" random effects. I believe that Henderson would be content to use terms such as "estimate" and "estimate of realized value." I prefer simple terminology—to use "estimation" of both fixed and random effects.

7.2 Unbiasedness and Shrinkage

The BLUP estimator $\hat{u}$ of u is sometimes said to be "unbiased" because it satisfies

$$E[\hat{u}] = E[u].$$

It is also described as a "shrinkage" estimator because its components are less spread than the generalized least-squares estimates that would be used if the components of u were regarded as fixed effects. These two descriptions seem to be in conflict with one another.

The difficulty is that the use of the term "unbiased" is very different from the condition

$$E[\hat{u} \mid u] = u \qquad \text{for all } u,$$

which is what many people intuitively expect, particularly in circumstances where the random processes generating the random effects (genetic merits of dairy bulls or ore grades in a deposit, for instance) occur prior to other processes generating noise in the data. Whenever I mean "unbiased" in the sense that $E[\hat{u}] = E[u]$ I try to make this very clear. Perhaps some new terminology would be a good idea.

7.3 Fixed and Random Effects

Eisenhart (1947) distinguished between two uses for analysis of variance: (1) detection and estimation of fixed (constant) relations among the means of subsets of the universe of objects concerned; and (2) detection and estimation of components of (random) variation associated with a composite population. He suggested two parallel sets of questions to help distinguish fixed and random effects, and these tend to imply that estimation of random effects is not sensible because effects of interest must be treated as fixed.

1. "Are the conclusions to be confined to the things actually studied (the animals, or the plots); to the immediate sources of these things (the herds, or the fields); or expanded to apply to more general populations (the species, or the farmland of the state)?"
2. "In complete repetitions of the experiment would the same things be studied again (the same animals, or the same plots); would new samples be drawn from the identical sources (new samples of animal from the same herds, or new experimental arrangements on the same fields); or would new samples be drawn from more general populations (new samples of animals from new herds, or new experimental arrangements on new fields)?"

The discussion of the difference between fixed and random effects given in Searle (1971) is similarly misleading in my view. On page 383, Searle writes

> ... when inferences are going to be confined to the effects in the model the effects are considered fixed; and when inferences will be made about a population of effects from which those in the data are considered to be a random sample then the effects are considered as random.

Both of these people have defined their terms in such a way that estimation of random effects is not possible.

In my view, there are two questions which might need to be answered in order to decide whether the effects in a class are to be treated as fixed or random. The first of these is: Do these effects come from a probability distribution? If the effects do not come from a probability distribution then the effects should be treated as fixed.

If the effects do come from a probability distribution, then if the effects are themselves being estimated they should be treated as random. For classes

This content downloaded from 14.139.222.72 on Fri, 01 Nov 2019 07:01:24 UTC
All use subject to https://about.jstor.org/terms

of effects which are nuisance parameters, there is a second question to be answered: Is interclass information to be recovered for this class of effects? If interclass information is to be recovered then treat the class of effects as random, otherwise fixed.

EXAMPLE. When estimating the genetic merit of dairy bulls, the herd-year-season effects that are used to model environmental variation do come from a probability distribution which describes variation: that between herds, years and calving seasons. When estimating herd-year-season effects I believe that they should be treated as random. However, when estimating genetic merits of dairy bulls I believe that herd-year-season effects should be treated as fixed because the inter-herd information should not be recovered—Farmers' choices of artificial insemination bulls and their choice of feeding regimes are somewhat related, and to recover the inter-herd information would favor the bulls that were more often chosen by farmers providing high feed intakes. Similarly, bull choices vary between years, and yearly climatic variation might bias estimated breeding values. The decision not to recover inter-herd-year-season information is based on a willingness to sacrifice some efficiency for greater robustness relative to departures from the assumptions of the model.

I agree with Tukey's remark made in discussion of Nelder (1977): "... our focus must be on questions, not models." The choice of whether a class of effects is to treated as fixed or random may vary with the question which we are trying to answer.

7.4 Generalizing Likelihood

When Fisher (1922, page 326) introduced the idea of likelihood as the probability of data given a hypothesis regarded as a function of the hypothesis, I doubt that he had considered problems of estimating random effects.

For estimation of fixed effects and dispersion parameters for the random effects model (1.1) the usual definition of likelihood seems adequate. However, if we wish to estimate the random effects, u, then the usual concept of likelihood seems to oblige us to regard the value of u as part of the hypothesis. In effect, the idea of likelihood tends to force us to regard things that we wish to estimate as fixed rather than random effects.

In my view, when we specify a mixed model the dispersion of the random effects, u, and the dispersion of the error, e, ought to be given similar logical status. How can this be achieved?

One possible attempt to resolve the problem based on Bayesian concepts would be to regard the assumed distribution of the random effects as an *objective* prior distribution. It is different in logical status from a *subjective* prior distribution from which the fixed effects might be considered to have come. One distinguishing feature is that the *objective* distribution of the random effects is being used to describe *variation* while the *subjective* prior distribution of the fixed effects is being used to describe *uncertainty*. A second distinguishing feature is that assumptions about the distribution of the random effects can be tested.

To use such an *objective* prior distribution for random effects should not be considered to make one a Bayesian. Good (1965, page 8) wrote

> the essential defining property of a Bayesian is that he regards it as meaningful to talk about the probability $P(H \mid E)$ of hypothesis H, given evidence E.

To be a Bayesian, you would need to be willing to put a prior distribution on fixed effects as well as random effects. Kempthorne in the discussion of Lindley and Smith (1972, page 37) described people who wish to estimate random effects as "legitimate" Bayesians; I agree with the ideas behind this designation, but I prefer to adhere to Good's use of the term "Bayesian."

This attempt to define likelihood in a way which allows estimation of random effects separates the distribution of e (the likelihood), the distribution of u (the *objective* prior distribution), and the uncertainty about β and θ (the *subjective* prior distribution for Bayesians) into three separate boxes. It does not achieve the stated goal of giving similar logical status to the dispersion of e and the dispersion of u.

Most likely unobservables. An alternative resolution of the problem based on Classical concepts is to formalize the principle behind Henderson's derivation of BLUP. This principle does achieve the goal of giving similar logical status to the dispersion of e and the dispersion of u. A colleague, T. Lwin, and I would like to suggest the name *method of most likely unobservables* for it.

Given the mixed model (1.1) and having observed the data y, the method of most likely unobservables is to say that the best estimates of β, u and e are the ones that maximize the density of the unobservables u and e subject to the constraint (1.1) of having observed the data. Because there are n constraints and n components to the vector e, the easiest way to solve for the maximum is to set

$$e = y - X\beta - Zu$$

and maximize the joint density of u and e with

This content downloaded from 14.139.222.72 on Fri, 01 Nov 2019 07:01:24 UTC
All use subject to https://about.jstor.org/terms

respect to β and u. This is precisely Henderson's derivation of BLUP discussed in Section 4.1.

For a fixed effects model, the likelihood of the data and the likelihood of the errors are equal, so the method of most likely unobservables gives estimates of the fixed effects that are the maximum likelihood estimates and gives estimates of the errors, e, that are the usual residuals.

7.5 On Schools of Thought

I believe that the distinction between fixed and random effects should be clarified before differences between schools of thought are considered.

Statistics is concerned with both variation and uncertainty. Classical statistics can be distinguished from Bayesian statistics by its refusal to use probability distributions to describe uncertainty. However, it is quite willing to use probability distributions to describe variation, and this should include variation between random effects. Once this is clarified, several situations in which the two schools of thought appeared to give different answers instead demonstrate close agreement between the schools.

There are a number of reasons why the estimation of random effects has been neglected to some extent by the Classical (Neyman–Pearson–Wald, Berkeley) school of thought.

1. The distinction between *fixed* and *random* effects has been often taken to be that effects are random when you are not interested in their individual values. Searle (1971), for instance, seems to suggest this. This implies that you should never be interested in estimating random effects.
2. The idea of estimating random effects seems suspiciously Bayesian to some Classical statisticians. The gulf between Bayesian and Classical statisticians seems to be like many other gulfs between schools of thought in that the adherents of each school emphasize the differences between schools rather than the similarities.

Within the Bayesian paradigm, there is little reason for distinguishing between fixed and random effects. All effects are treated as random in the sense that probability distributions used to describe uncertainty are not treated any differently from probability distributions used to describe variation.

ACKNOWLEDGMENT

I am thankful for the helpful criticism provided by the editors, C. N. Morris and the late M. H. DeGroot, two referees, colleagues too numerous to list individually—colleagues from the Herd Improvement Section of the Victorian Department of Agriculture, the Animal Genetics and Breeding Unit of the University of New England, and the Division of Mathematics and Statistics of the Commonwealth Scientific and Industrial Research Organization—and many other people with whom I have discussed these issues.

REFERENCES

BAILEY, A. L. (1945). A generalized theory of credibility. *Proceedings of the Casualty Actuarial Society* **32**.

BAILEY, A. L. (1950). Credibility procedures. *Proceedings of the Casualty Actuarial Society* **37**.

BARNDORFF-NIELSEN, O. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70** 343–365.

BATTESE, G. E., HARTER, R. M. and FULLER, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *J. Amer. Statist. Assoc.* **83** 28–36.

BERGER, J. O. and DEELY, J. (1988). A Bayesian approach to ranking and selection of related means with alternatives to analysis-of-variance methodology. *J. Amer. Statist. Assoc.* **83** 364–373.

BESAG, J. (1986). On the statistical analysis of dirty pictures (with discussion). *J. Roy. Statist. Soc. Ser. B* **48** 259–302.

BOX, G. E. P. and TIAO, G. C. (1968). Bayesian estimation of means for the random effects model. *J. Amer. Statist. Assoc.* **63** 174–181.

BROEMELING, L., YUSOFF, A. and DIAZ, J. (1985). Some Bayesian solutions for problems of adaptive estimation in linear dynamic systems. *Comm. Statist. Theory Methods* **14** 401–418.

COCHRAN, W. G. (1946). The relative accuracy of systematic and stratified random samples from a certain class of populations. *Ann. Math. Statist.* **17** 164–177.

DAVIS, C. E. (1986). Regression to the mean. In *Encyclopedia of Statistical Sciences* **7** 706–708. Wiley, New York.

DAWID, A. P. (1976). Comment on "Plausibility inference" by O. Barndorff-Nielsen. *J. Roy. Statist. Soc. Ser. B* **38** 123–125.

DEMING, W. E. (1950). *Some Theory of Sampling*. Wiley, New York.

DEMPFLE, L. (1977). Comparison of several sire evaluation methods in dairy cattle breeding. *Livestock Production Science* **4** 129–139.

DUNCAN, A. J. (1962). Bulk sampling: Problems and lines of attack. *Technometrics* **4** 319–344.

EFRON, B. and MORRIS, C. (1977). Stein's paradox in statistics. *Scientific Amer.* **236** (5) 119–127.

EISENHART, C. (1947). The assumptions underlying the analysis of variance. *Biometrics* **3** 1–21.

FAY, R. E. and HERRIOT, R. A. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *J. Amer. Statist. Assoc.* **74** 269–277.

FELLNER, W. H. (1986). Robust estimation of variance components. *Technometrics* **28** 51–60.

FISHER, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc. London Ser. A* **222** 309–368.

FULLER, W. A. and HARTER, R. M. (1987). The multivariate components of variance model for small area estimation. In *Small Area Statistics* (R. Platek, J. N. K. Rao, C. E. Sarndal and M. P. Singh, eds.) 103–123. Wiley, New York.

This content downloaded from 14.139.222.72 on Fri, 01 Nov 2019 07:01:24 UTC
All use subject to https://about.jstor.org/terms

Goldberger, A. S. (1962). Best linear unbiased prediction in the generalized linear regression model. *J. Amer. Statist. Assoc.* **57** 369–375.

Good, I. J. (1965). *The Estimation of Probabilities.* MIT Press, Cambridge, Mass.

Gy, P. (1982). *Sampling of Particulate Materials.* North-Holland, Amsterdam.

Harrison, P. J. and Stevens, C. F. (1976). Bayesian forecasting (with discussion). *J. Roy. Statist. Soc. Ser. B* **38** 205–247.

Harville, D. A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika* **61** 383–385.

Harville, D. A. (1976). Extension of the Gauss-Markov theorem to include the estimation of random effects. *Ann. Statist.* **4** 384–395.

Harville, D. A. (1984). Discussion of papers by Wahba, Journel and Henderson. In *Statistics: An Appraisal. Proceedings 50th Anniversary Conference Iowa State Statistical Laboratory* (H. A. David and H. T. David, eds.) 281–286. Iowa State Univ. Press, Ames, Ia.

Harville, D. A. (1985). Decomposition of prediction error. *J. Amer. Statist. Assoc.* **80** 132–138.

Harville, D. A. (1990). BLUP (Best Linear Unbiased Prediction) and Beyond. In *Advances in Statistical Methods for Genetic Improvement of Livestock* (D. Gianola and K. Hammond, eds.) 239–276. Springer, New York.

Hazel, L. N. (1943). The genetic basis for constructing selection indexes. *Genetics* **28** 476–490.

Henderson, C. R. (1950). Estimation of genetic parameters (abstract). *Ann. Math. Statist.* **21** 309–310.

Henderson, C. R. (1963). Selection index and expected genetic advance. In *Statistical Genetics and Plant Breeding* 141–163. Nat. Acad. Sci., Nat. Res. Council, Publication 982, Washington, DC.

Henderson, C. R. (1965). A method of determining if selection will bias estimates. *J. Animal Sci.* **24** 849.

Henderson, C. R. (1973). Sire evaluation and genetic trends. In *Proceedings of the Animal Breeding and Genetics Symposium in Honour of Dr. Jay L. Lush* 10–41. Amer. Soc. Animal Sci.—Amer. Dairy Sci. Assn.—Poultry Sci. Assn., Champaign, Ill.

Henderson, C. R. (1975a). Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31** 423–447.

Henderson, C. R. (1976). A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* **32** 69–83.

Henderson, C. R. (1984). *Applications of Linear Models in Animal Breeding.* University of Guelph.

Henderson, C. R., Kempthorne, O., Searle, S. R. and von Krosigk, C. N. (1959). Estimation of environmental and genetic trends from records subject to culling. *Biometrics* **13** 192–218.

Henderson, H. V. and Searle, S. R. (1981). On deriving the inverse of a sum of matrices. *SIAM Rev.* **23** 53–60.

Hickman, J. C. (1975). Introduction and historical overview of credibility. In *Credibility Theory and Applications* (P. M. Kahn, ed.) 181–192. Academic, New York.

Hoadley, B. (1986). Quality Measurement Plan (QMP). In *Encyclopedia of Statistical Sciences* **7** 373–398. Wiley, New York.

Ishii, G. (1969). Optimality of unbiased predictors. *Ann. Inst. Statist. Math.* **21** 471–488.

Jeske, D. R. and Harville, D. A. (1988). Prediction-interval procedures and (fixed-effects) confidence-interval procedures for mixed linear models. *Comm. Statist. Theory Methods* **17** 1053–1087.

Journel, A. G. and Huijbregts, Ch. J. (1978). *Mining Geostatistics.* Academic, New York.

Jowett, G. H. (1952). The accuracy of systematic sampling from conveyor belts. *J. Roy. Statist. Soc. Ser. C* **1** 50–59.

Kackar, R. N. and Harville, D. A. (1981). Unbiasedness of two-stage estimation and prediction procedures for mixed linear models. *Comm. Statist. A—Theory Methods* **10** 1249–1261.

Kackar, R. N. and Harville, D. A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *J. Amer. Statist. Assoc.* **79** 853–862.

Kahn, P. M. (1982). Credibility. In *Encyclopedia of Statistical Science* **2** 222–224. Wiley, New York.

Kelley, T. L. (1927). *The Interpretation of Educational Measurements.* World Books, New York.

Khuri, A. I. and Sahai, H. (1985). Variance components analysis: A selective literature survey. *Internat. Statist. Rev.* **53** 279–300.

Lindley D. V. and Smith, A. F. M. (1972). Bayes estimates for the linear model (with discussion). *J. Roy. Statist. Soc. Ser. B* **34** 1–41.

Lush, J. L. (1949). *Animal Breeding Plans*, 3rd ed. Iowa State College Press, Ames, Ia.

Matheron, G. (1965). *Les Variables Regionalisées et Leur Application.* Masson, Paris.

Mayerson, A. L. (1965). A Bayesian view of credibility. *Proceedings of the Casualty Actuarial Society* **51**.

Meinhold, R. J. and Singpurwalla, N. D. (1983). Understanding the Kalman filter. *Amer. Statist.* **37** 123–127.

Morris, C. N. (1983). Parametric empirical Bayes inference: Theory and applications. *J. Amer. Statist. Assoc.* **78** 47–55.

Mowbray, A. H. (1914). How extensive a payroll exposure is necessary to give a dependable pure premium? *Proceedings of the Casualty Actuarial Society* **1**.

Nelder, J. A. (1977). A reformulation of linear models (with discussion). *J. Roy. Statist. Soc. Ser. A* **140** 48–76.

Patterson, H. D. and Thompson, R. (1971). Recovery of interblock information when block sizes are unequal. *Biometrika* **58** 545–554.

Pearson, K. (1896). Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia. *Philos. Trans. Roy. Soc. London Ser. A* **187** 253–318.

Prasad, N. G. N. and Rao, J. N. K. (1986). On the estimation of mean square error of small area estimators. *1986 Proceedings of the Section on Survey Research Methods, American Statistical Association* 108–116.

Robertson, A. (1955). Prediction equations in quantitative genetics. *Biometrics* **11** 95–98.

Robinson, D. L. (1987). Estimation and use of variance components. *The Statistician* **36** 3–14.

Robinson, G. K. (1986). Group effects and computing strategies for models for estimating breeding values. *J. Dairy Sci.* **69** 3106–3111.

Robinson, G. K. and Jones, L. P. (1987). Approximations for prediction error variances. *J. Dairy Sci.* **70** 1623–1632.

Sallas, W. M. and Harville, D. A. (1981). Best linear recursive estimation for mixed linear models. *J. Amer. Statist. Assoc.* **76** 860–869.

Saunders, I. W., Robinson, G. K., Lwin, T. and Holmes, R. J. (1989). A simplified variogram method for determining the estimation error variance in sampling from a continuous stream. *Internat. J. Mineral Processing* **25** 175–198.

Searle, S. R. (1971). *Linear Models.* Wiley, New York.

Smith, H. F. (1936). A discriminant function for plant selection. *Ann. Eugenics* **7** 240–250.

This content downloaded from 14.139.222.72 on Fri, 01 Nov 2019 07:01:24 UTC
All use subject to https://about.jstor.org/terms

SORENSON, H. W. (1970). Least-squares estimation from Gauss to Kalman. *IEEE Spectrum* **7** 63–68.

STEIN, C. (1945). A two-sample test for a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.* **16** 243–258.

STEIN, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1** 197–206. Univ. California Press, Berkeley.

THOMPSON, R. (1973). The estimation of variance and covariance components with an application when records are subject to culling. *Biometrics* **39** 527–550.

TURNER, H. N. and YOUNG, S. S. Y. (1973). *Quantitative Genetics in Sheep Breeding*. Macmillan, New York.

YATES, F. (1949). Systematic sampling. *Philos. Trans. Roy. Soc. London Ser. A* **241** 345–377.

Comment

Katherine Campbell

It has been a pleasure to read about the long history of Best Linear Unbiased Prediction, and especially about its uses in traditional statistical areas of application such as agriculture. My own experience with BLUP is in the context of ill-posed inverse problems, and I would like to discuss this paper from this point of view, where the random effects are generated by hypothesized superpopulations, in contrast with the identifiable populations considered by Robinson.

MODEL-BASED ESTIMATION FOR ILL-POSED INVERSE PROBLEMS

The author mentions two examples of superpopulation approaches to estimation: image restoration and geostatistics. The same ideas are also used in model-based estimation for finite populations, function approximation and many other inference problems. These problems concern inference about a reality that is in principle completely determined, but whose observation is limited by the number and/or resolution of the feasible measurements, as well as by noise. In geophysics, x-ray imaging and many other areas of science and engineering these are known as inverse problems (O'Sullivan, 1986; Tarantola, 1987).

The unknown reality we may consider to be a function $\mathbf{m}$ defined on some domain $\mathbf{T}$. The data typically consist of noisy observations on a finite number n of functionals of $\mathbf{m}$. We can write the data vector $\mathbf{y}$ in terms of a transformation L mapping $\mathbf{m}$ into an n-dimensional vector:

$$\mathbf{y} = L\mathbf{m} + \mathbf{e}.$$

Katherine Campbell is a staff member of the Statistics Group, Los Alamos National Laboratory, Los Alamos, New Mexico 87545.

In the sequel, we will assume that L is a linear transformation, i.e., that the observed functionals are linear. In particular, if the cardinality $|\mathbf{T}|$ of $\mathbf{T}$ is finite, L can be represented by an $n \times |\mathbf{T}|$ matrix.

BLUP arises when we embed this problem in a superpopulation model, under which $\mathbf{m}$ is one realization (albeit the only one of interest) of a stochastic process $\mathbf{M}$ indexed by $\mathbf{T}$. This superpopulation model has two components, corresponding to the "fixed" and "random" effects in Robinson's discussion. The fixed effects define the mean of the superpopulation, which is here assumed to lie in a finite-dimensional subspace of functions on $\mathbf{T}$. We denote this subspace by $\mathbf{R}(F)$, the range of the linear operator F that maps a p-vector $\mathbf{b}$ into the function

$$F\mathbf{b} = \sum b_i \mathbf{f}_i.$$

where $\{\mathbf{f}_1, \ldots, \mathbf{f}_p\}$ is a basis for the subspace.

Any realization of $\mathbf{M}$ can then be written as a sum $F\beta + \mathbf{u}$, where β is an unknown vector of p fixed effects and $\mathbf{u}$ is a realization of a stochastic "random effects" process with mean zero and covariance P. As we are interested in the realized $\mathbf{m}$, we need to estimate both the fixed and random effects. Among estimates that are linear functions of the data vector

$$\text{(1)} \qquad \mathbf{y} = LF\beta + L\mathbf{u} + \mathbf{e},$$

the BLUP $\hat{\mathbf{m}} = F\hat{\beta} + \hat{\mathbf{u}}$ is the optimal choice: under the assumed superpopulation model $\hat{\mathbf{m}}$ is unbiased in the sense of Section 7.2 (i.e., $E\hat{\mathbf{m}} = E\mathbf{m}$) and it minimizes the variance of any linear functional of $\hat{\mathbf{m}} - \mathbf{m}$. (To make the correspondence between equation 1 and Robinson's equation (1.1) explicit, $X \approx LF$, $Z \approx L$, $G \approx LPL^T$, and $\mathbf{e}$ is a realization of a random n-vector with mean zero and covariance R. $\hat{\beta}$ and $\hat{\mathbf{u}}$ are then provided by the BLUP formulas.)

This content downloaded from 14.139.222.72 on Fri, 01 Nov 2019 07:01:24 UTC
All use subject to https://about.jstor.org/terms