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

$$\mathbf{y} = LF\beta + L\mathbf{u} + \mathbf{e}, \tag{1}$$

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.)

The following examples make the preceding discussion more concrete.

1. The term "superpopulation model" is associated with model-based inference in finite populations. A general discussion and many references are found in Cassel, Sarndal and Wretman (1977). $\mathbf{T}$ is here the finite set of population unit labels. $F\beta$ is a linear model for the mean of the outcome variable $\mathbf{m}$ as a function of p auxiliary variables $\mathbf{f}_i$ that are known in advance for each population unit. L is the $n \times |\mathbf{T}|$ sampling matrix whose ith row contains a one in the tth column if the ith sampled unit is the tth population unit, and zeros elsewhere. Commonly P is modeled as a function of the auxiliary variables $\mathbf{f}$ with some parameters that are estimated from the data. Very often it is assumed that the outcome variable $\mathbf{m}$, can be measured without error once the tth unit is sampled, and so $R = 0$.

2. In image analysis, $\mathbf{T}$ indexes the pixels. Frequently the only fixed effect is a constant, i.e., $p = 1$ and $\mathbf{f}_1 = 1$. L is generally convolution with a known or assumed point-spread function, followed by exhaustive sampling. Although there may be $|\mathbf{T}|$ observations, the effective dimension of $\mathbf{R}(L)$ is much less than $|\mathbf{T}|$ because of the limited resolution of the point-spread function, and the problem is ill-conditioned. Generally, in addition, R is not zero. With large numbers of observations at their disposal, image analysts have been quite adventurous in their modeling of the random effects. A popular choice for the probabilistic structure of $\mathbf{u}$ in the recent literature is a Markov random field (see Marroquin, Mitter and Poggio, 1987).

3. The parallels between kriging and BLUP have been described by Cressie (1990). In geostatistical applications, $\mathbf{T}$ indexes a subset of two- or three-dimensional Euclidean space, and the functions $\mathbf{f}_i$ are low-order monomials in the spatial coordinates. L is again a sampling operator. R includes analytical error. Sampling error, resulting from imperfect surveying, small sample volumes and/or local heterogeneity, can be assigned to R or to a discontinuous component of P called the "nugget effect." The continuous part of P is estimated as a function of the spatial coordinates.

4. The use of superpopulation models and BLUP for approximating the output of large computer codes as a function of many input parameters is discussed by Sacks, Welch, Mitchell and Wynn (1989). Their use in function approximation is illustrated by Blight and Ott (1975) and O'Hagan (1978). In these applications, $\mathbf{T}$ indexes a collection of r-vectors $\mathbf{x}_t = (x_{t,1}, \ldots, x_{t,r})$. The components of $\mathbf{x}_t$ are r input parameters for a computer code, or the r independent variables in the case of function approximation. Again the functions $\mathbf{f}_i$ are frequently monomials in these r variables, and L is a sampling operator. For computer codes R is zero, while curve-fitting applications may or may not include measurement error. The almost universal choice for P is a product of the form

$$P_{s,t} = \prod_{i=1}^{r} \exp(-\rho_i |x_{s,i} - x_{t,i}|^{\theta_i}).$$

The choice of parameter values reflects assumptions about the continuity and differentiability of realizations of $\mathbf{M}$ but is seldom the result of estimation based on the data. Spline interpolation implies another choice, namely a generalized covariance (see below) of the form

$$K_{s,t} = \rho |\mathbf{x}_s - \mathbf{x}_t|^2 \log(\theta |\mathbf{x}_s - \mathbf{x}_t|)$$

(see Dubrule, 1983).

COVARIANCE MODELS

The greatest difficulty in practical application of BLUP is the specification of G. (Specification of R is seldom problematic.) The brief discussion of Section 5.4 suggests some of the difficulties encountered; estimation of covariances is notoriously a more difficult problem than estimation of means. The covariance matrices $G \approx LPL^T$ required by superpopulation models are nontrivial. Many areas, notably geostatistics, have developed their own methods, some rather *ad hoc*, to estimate this parameter.

A geostatistical innovation in this area arises from the observation that $\mathbf{M}$ need not possess a covariance for the BLUP to exist. Recall that in a vector-space approach to multivariate analysis, $\operatorname{Cov} \mathbf{M} = P$ means that

$$\operatorname{Cov}\{\langle \boldsymbol{\lambda}, \mathbf{M} \rangle, \langle \boldsymbol{\nu}, \mathbf{M} \rangle\} = \langle \boldsymbol{\lambda}, P\boldsymbol{\nu} \rangle \qquad (2)$$

for all vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$. (Eaton, 1983, treats the case where $|\mathbf{T}|$ is finite, for which functions defined on $\mathbf{T}$ are just vectors in $\mathbf{R}^{|\mathbf{T}|}$, which can be supplied with an inner product $\langle \cdot, \cdot \rangle$ in the usual manner. Equation (2) in an appropriate Hilbert space of functions serves as a rigorous definition of $\operatorname{Cov} \mathbf{M}$ when $|\mathbf{T}|$ is infinite.) In geostatistics, the BLUP is usually computed using a "generalized" covariance K, for which it is sufficient that

$$\operatorname{Cov}\{\langle \boldsymbol{\lambda}, \mathbf{M} \rangle, \langle \boldsymbol{\nu}, \mathbf{M} \rangle\} = \langle \boldsymbol{\lambda}, K\boldsymbol{\nu} \rangle$$

for all $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$ orthogonal to $\mathbf{R}(F)$. This enlarges somewhat the set of models available for this component of the superpopulation model, and moreover K, unlike P, can be estimated without correcting for the unknown fixed effects β. A popular choice for K, when the fixed effects include an unknown constant, is based on the "semivariogram" function

$$\gamma(s, t) = \frac{1}{2}E(\mathbf{M}_s - \mathbf{M}_t)^2.$$

In practice, the BLUP is fairly robust to most aspects of the choice of G. However, a critical parameter is the relative size of G and R. If G is of the form $\alpha^2\Gamma$, the resulting BLUP is moderately insensitive to the choice of Γ but very sensitive to the parameter α, which controls the degree of smoothing or "shrinkage" in $\hat{\mathbf{u}}$ (Section 7.2). Cross-validation is the most common data-based method for estimating this parameter; see O'Sullivan (1986) and Woodbury (1989).

SUPERPOPULATIONS AND RANDOM EFFECTS

The reader may feel that the introduction of superpopulation models takes us quite far from the spirit of Robinson's paper, wherein pains have been taken to use only the classical interpretation of probability as a description of ontological variability. In the paper, the superpopulation generating the random effects is a real population (e.g., the population of potential sires), while in the context of ill-posed inverse problems it appears that a superpopulation is introduced merely as a mechanism for imposing additional constraints on the problem so that a unique solution can be defined. In this connection several observations are in order.

First of all, whether or not the superpopulation is real does not appear to be a central philosophical problem in the acceptance of random effect estimation by the classical school. Although another realization of a real superpopulation is feasible (we could repeat the experiment with another set of sires), BLUP estimates only that realization (the set of four sires) that was actually represented in the given data. In the case of inverse problems, nature, rather than an animal breeder, provides the realization that was observed. The sticking point with respect to BLUP seems to be that, despite the fact that the realization is now fixed, we continue to model its effects as random, with different numerical results, as Robinson's introductory example shows, than if we were to treat them as fixed.

What the superpopulation (real or imaginary) point of view makes clear is that the difference between a fixed effect and a random effect is that one belongs to the mean of the statistical model, while the other is a deviation from the mean, i.e., is described by the variance component of the statistical model. Classical statistics has no problem with this distinction in ordinary regression models, where (as in Section 4.3) such deviations are called "residuals" and may be individually estimated for diagnostic purposes, among others. No one would propose that therefore residuals should be treated as fixed effects! Similarly, BLUP preserves the distinction throughout the analysis, even after the realization of the real or implied superpopulation has been fixed and the data collected.

As used in the solution of ill-posed problems, superpopulation models often attempt, perhaps unhelpfully, to blur the distinction between the epistemological (Bayesian) and ontological (Classical) interpretations of probability. (In spirit, and also in form, superpopulation models thus come close to empirical Bayes ideas, although formal empirical Bayes techniques remain largely unexploited.) In particular, the empirical approach to the estimation of P adopted in many superpopulation applications implies that this parameter reflects ontological variability, not subjective uncertainty. Empirical superpopulations for geostatistics are the subject of "deterministic geostatistics" (cf. Isaaks and Srivastava, 1988).

INVERSE PROBLEMS AND STATISTICS

Although ill-posed inverse problems are certainly inference problems, this area has been neglected by Classical statisticians, apparently because probability in this context is generally thought to describe uncertainty rather than variability. Nevertheless, probabilistic regularization methods are widely used, have excellent track records and can often be given empirical interpretations. Statisticians ignore these developments at the risk of being found irrelevant by many of their colleagues in the physical sciences, where inverse problems are ubiquitous.

In particular, linear regularization methods for ill-posed inverse problems can be interpreted as BLUP under appropriate superpopulation models. I therefore welcome Robinson's article, not only for its wide-ranging survey of BLUP history and applications, but also for its examination of the philosophical questions raised by the estimation of random effects. Better understanding of these philosophical problems may induce statisticians to reevaluate this important class of problems as an appropriate subject for statistical research.