Bayesian Inference for Variance Components Using Only Error Contrasts
Author(s): David A. Harville
Source: *Biometrika*, Vol. 61, No. 2 (Aug., 1974), pp. 383-385
Published by: Oxford University Press on behalf of Biometrika Trust
Stable URL: http://www.jstor.org/stable/2334370

REFERENCES

DAVIS, A. W. (1970). Exact distribution of Hotelling's T_0^2. *Biometrika* **57**, 187–91.

HOTELLING, H. (1951). A generalized T test and measure of multivariate dispersion. *Proc. 2nd Berkeley Symp.*, 23–41.

HUGHES, D. T. & SAW, J. G. (1972). Approximating percentage points of Hotelling's T_0^2 statistic. *Biometrika* **59**, 224–6.

PILLAI, K. C. S. & SAMPSON, P. (1959). On Hotelling's generalization of T^2. *Biometrika* **46**, 160–8.

[*Received February* 1974]

Bayesian inference for variance components using only error contrasts

BY DAVID A. HARVILLE

Aerospace Research Laboratories, Wright-Patterson A.F.B., Ohio

SUMMARY

Patterson & Thompson (1971) proposed estimating the variance components of a mixed analysis of variance model by maximizing the likelihood of a set of error contrasts. In the present paper, a convenient representation is obtained for that likelihood. It is shown that, from a Bayesian viewpoint, using only error contrasts to make inferences on variance components is equivalent to ignoring any prior information on the fixed effects and using all the data.

Some key words: Bayesian inference; Error contrasts; Maximum likelihood; Mixed linear models; Variance components.

1. INTRODUCTION

Suppose that an appropriate model for an $n \times 1$ vector y of observations is

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

where X is a known $n \times p$ matrix of rank p, β is a $p \times 1$ vector of unobservable parameters, and e is a $n \times 1$ vector of normally distributed random errors having zero means. Suppose also that

$$\operatorname{var}(e) = H(\theta),$$

an $n \times n$ matrix whose entries are known functions of a vector θ of unobservable parameters. We will often write H for $H(\theta)$. It will be assumed that $H(\theta)$ is nonsingular for all θ in the parameter space. The model includes the general mixed analysis of variance model as a special case.

By an error contrast, we shall mean a linear combination $b'y$ of the observations such that $E(b'y) \equiv 0$, that is, such that $b'X = 0$.

Suppose that we want to use the data vector y to estimate θ. The maximum likelihood estimator of θ takes no account of the loss in degrees of freedom resulting from estimating β. In an important paper, Patterson & Thompson (1971) proposed a modified maximum likelihood technique for estimating θ which does not suffer from that defect. Their technique consists essentially of maximizing the likelihood function associated with a specified set of $n-p$ linearly independent error contrasts rather than the full likelihood function.

A Bayesian who reasons like Stone (1963) might, in making inferences on θ alone, also find attractive the idea of using error contrasts rather than all the data. If he uses only error contrasts, he need not specify a prior distribution for β, nor does he have to carry out analytical or numerical integrations in determining his posterior distribution for θ as he would otherwise. Moreover, the theoretical advantage in precision of using all the data may in practice be negated by the added approximations needed to specify a complete prior distribution for θ and β and to actually calculate the marginal posterior distribution of θ. Also, if the prior distribution for β is bad in the sense that it is concentrated away from the true value of β, the posterior distribution of θ based on all the data will be affected adversely, but that based on only error contrasts will not.

In § 2, we derive an expression for the likelihood function associated with any $n-p$ linearly independent error contrasts which appears to be more convenient than the one obtained by Patterson & Thompson. As related in § 3, this derivation reveals an interesting relationship between the posterior distribution of θ based on only the error contrasts and that based on all the data.

2. Likelihood function for error contrasts

The maximum possible number of linearly independent contrasts in any set of error contrasts is $n-p$. Following Patterson & Thompson, define the $n \times n$ matrix S by $S = I - X(X'X)^{-1}X'$ and the $n \times (n-p)$ matrix A by the conditions $S = AA'$ and $A'A = I$. The vector $w = A'y$ provides a particular set of $n-p$ linearly independent error contrasts. Moreover, the likelihood function associated with any other set of $n-p$ linearly independent error contrasts is proportional to that associated with w.

The likelihood function associated with w is $f_w(A'y|\theta)$, where $f_w(\cdot|\theta)$ is the probability density function of w. We now turn to the problem of deriving a convenient expression for $f_w(A'y|\theta)$. Take

$$G = H^{-1}X(X'H^{-1}X)^{-1}.$$

Using well-known results on determinants, we find

$$\begin{aligned} |\det(A, G)| &= [\det\{(A, G)'(A, G)\}]^{\frac{1}{2}} \\ &= \{\det(I)\}^{\frac{1}{2}}\{\det(G'G - G'AI^{-1}A'G)\}^{\frac{1}{2}} = \{\det(X'X)\}^{-\frac{1}{2}}. \end{aligned} \tag{1}$$

Let $\hat{\beta} = G'y$, and denote by $f_{\hat{\beta}}(\cdot|\theta, \beta)$ and $f_y(\cdot|\theta, \beta)$ the probability density functions of $\hat{\beta}$ and y, respectively. Using the statistical independence of w and $\hat{\beta}$, the result (1), and the well known-relationship

$$(y - X\beta)'H^{-1}(y - X\beta) = (y - X\hat{\beta})'H^{-1}(y - X\hat{\beta}) + (\beta - \hat{\beta})'(X'H^{-1}X)(\beta - \hat{\beta}),$$

we have

$$\begin{aligned} f_w(A'y|\theta) &= \int f_w(A'y|\theta) f_{\hat{\beta}}(G'y)|\theta, \beta)\, d\beta \\ &= \{\det(X'X)\}^{\frac{1}{2}} \int f_y(y|\theta, \beta)\, d\beta \qquad (2) \\ &= (2\pi)^{-\frac{1}{2}(n-p)}\{\det(X'X)\}^{\frac{1}{2}}\{\det(H)\}^{-\frac{1}{2}}\{\det(X'H^{-1}X)\}^{-\frac{1}{2}} \\ &\qquad \exp\{-\tfrac{1}{2}(y - X\hat{\beta})'H^{-1}(y - X\hat{\beta})\}. \qquad (3) \end{aligned}$$

The nonexponential part of the expression obtained by Patterson & Thompson for the likelihood function associated with w involves the nonzero characteristic values of HS. The representation (3) would appear to be more convenient than theirs for analytical work involving this likelihood function, such as deriving the likelihood equations, and to be definitely preferable for numerical computation of $f_w(A'y|\theta)$.

3. Posterior distributions

Denote by $g(\theta)$ the prior probability density function for θ. When only the error contrasts are used, the posterior probability density function for θ is proportional to $g(\theta) f_w(A'y|\theta)$. Now, suppose that the joint prior distribution of θ and β is such that θ and β are statistically independent, and the components of β are independently and uniformly distributed over the real line, i.e. that the joint prior density for θ and β is improper and is proportional to $g(\theta)$. Then, from (2), we have that, even when all the data are used, the posterior density for θ is proportional to $g(\theta) f(A'y|\theta)$. Thus, from a Bayesian viewpoint, using only error contrasts to make inferences on θ is equivalent to ignoring any prior information on β and using all the data to make those inferences.

The above Bayesian framework gives added insight into the difference between using full maximum likelihood to estimate θ and using the modified maximum likelihood technique of Patterson & Thompson to estimate that vector. Suppose that the data are sufficiently informative that the prior density for θ and β is flat relative to the likelihood function $f_y(y|\theta, \beta)$, so that for all practical purposes the joint posterior density for θ and β is proportional to the likelihood function. Then the maximum likelihood estimate of θ corresponds to the θ component of the mode of the posterior density for θ and β. In contrast, the modified maximum likelihood estimate is identical to the mode of the marginal posterior density for θ.

The author is indebted to a referee for a suggestion which simplified the development.

REFERENCES

PATTERSON, H. D. & THOMPSON, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**, 545–54.

STONE, M. (1963). Robustness of non-ideal decision procedures. *J. Am. Statist. Assoc.* **58**, 480–6.

[*Received September* 1973. *Revised October* 1973]

The variance of the inverse binomial estimator

BY D. J. BEST

C.S.I.R.O., North Ryde, New South Wales

SUMMARY

A closed form for the variance of the unbiased estimator of the proportion defective in a population, derived by inverse binomial sampling, is given. Comparisons are made with the maximum likelihood estimator which indicate that the unbiased estimator has greater efficiency and smaller mean squared error.

Some key words: Comparison of sampling variances; Estimation of proportion defective; Inverse binomial; Mean squared error of estimates; Sequential sampling.

1. INTRODUCTION

Suppose that a large population containing a proportion p of defectives is sampled until k defectives are obtained and, further, that X is the number of acceptable items sampled up to the kth defective. Then

$$\mathrm{pr}\,(X = x) = \binom{k+x-1}{x} p^k q^x,$$

where $0 < k < \infty$, $0 < p < 1$, $p+q = 1$, $x = 0, 1, \ldots$ and X is said to be distributed as a negative binomial or Pascal variate. The sequential sampling procedure is called an inverse binomial scheme.

2. VARIANCE OF THE ESTIMATOR

Various estimators of p have been proposed including (Haldane, 1945)

$$\hat{p}_{u,k} = (k-1)/(k+X-1).$$

It is easily seen that $\hat{p}_{u,k}$ is the unbiased estimator of p. Further, following Haldane (1945) or Kendall & Stuart (1967, p. 594) it can be shown that

$$E\{(\hat{p}_{u,k})^2\} = (k-1)\,p^k q^{1-k} \int_0^q t^{k-2}(1-t)^{1-k}\,dt,$$

$$\mathrm{var}\,(\hat{p}_{u,k}) = p^2 \sum_{r=1}^{\infty} \binom{k+r-1}{r}^{-1} q^r.$$

When $k = 1$, $\mathrm{var}\,(\hat{p}_{u,k})$ is thus equal to pq. A closed form, however, can also easily be obtained for $k > 1$.

First consider the maximum likelihood estimator of p, namely $\hat{p}_{ML,k} = k/(k+X)$. Feller (1968, Problem 33, pp. 241, 493) has given

$$E(\hat{p}_{ML,k}) = k\left[\sum_{r=1}^{k-1} \{(-p)^r q^{-r}\}/(r-k) + (-p)^k q^{-k} \log p\right]. \tag{1}$$