AFFDL-TR-75-134

EXACT AND APPROXIMATE BAYESIAN SOLUTIONS FOR INFERENCE ABOUT VARIANCE COMPONENTS AND MULTIVARIATE INADMISSIBILITY

THE UNIVERSITY OF MICHIGAN
DEPARTMENT OF STATISTICS
ANN ARBOR, MICHIGAN 48109

FEBRUARY 1976

TECHNICAL REPORT AFFDL-TR-75-134
FINAL REPORT FOR PERIOD 1 MAY 1972 – 1 JANUARY 1976

Approved for public release; distribution unlimited

AIR FORCE FLIGHT DYNAMICS LABORATORY
AIR FORCE WRIGHT AERONAUTICAL LABORATORIES
Air Force Systems Command
Wright-Patterson Air Force Base, Ohio 45433

NOTICE

This technical report has been reviewed and is approved for publication.

H. Leon Harter

H. LEON HARTER
Project Engineer
Applied Mathematics Group
Aerospace Dynamics Branch

FOR THE COMMANDER

Gerald G. Leigh

GERALD G. LEIGH, Lt. Col, USAF
Chief, Structures Division

AIR FORCE - 17 MAY 76 - 400

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFFDL-TR-75-134	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): Exact and Approximate Bayesian Solutions for Inference about Variance Components and Multivariate Inadmissibility.	5. TYPE OF REPORT & PERIOD COVERED: Technical - Final 1 May 1972 - 1 Jan 1976
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Bruce M. Hill	8. CONTRACT OR GRANT NUMBER(s): F33615-72-C-1213
9. PERFORMING ORGANIZATION NAME AND ADDRESS: The University of Michigan, Department of Statistics, Ann Arbor, Michigan 48109	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: DOD Element 61102F 70710201
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Flight Dynamics Laboratory, Aerospace Dynamics Branch, FYS, Wright-Patterson AFB, Ohio 45433	12. REPORT DATE: February 1976
	13. NUMBER OF PAGES: 38
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Air Force Flight Dynamics Laboratory (Same as above)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Variance components, Bayesian inference, random model, admissible multivariate estimator.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Exact and approximate Bayesian solutions for inference about variance components and multivariate inadmissibility are obtained, using a class of prior distributions which allows realistic forms of prior knowledge to be incorporated. Results are based upon new mathematical representations for the Appell and ordinary hypergeometric functions.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

PREFACE

This report, prepared by Professor Bruce M. Hill, is one of two parts of the final report on Contract No. F33615-72-C-1213, which was funded by the Aerospace Research Laboratories and was technically monitored by Dr. David A. Harville (until 30 June 1975) and Dr. H. Leon Harter (since 1 July 1975). The other part, prepared by Professor William A. Ericson and entitled "A Bayesian Approach to Two-Stage Sampling", will appear as AFFDL-TR-75-145.

TABLE OF CONTENTS

1. INTRODUCTION AND SUMMARY

In this report exact and approximate Bayesian solutions are derived for the problem of inference about the ratio of variance components in the one-way balanced random model analysis of variance, and for the closely related problem of obtaining admissible Bayes estimators for the mean of a multivariate normal distribution, using a class of prior distributions which allows reasonably realistic forms of prior knowledge to be incorporated.

In the usual balanced one-way random model analysis of variance, with $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, $i = 1,\ldots,I$, $j = 1,\ldots,J$ (see Section 2 for notation and assumptions), and with τ^2 the ratio of between to within variance components, it was shown by Hill [1965] that inference about all other parameters in the model is quite simple, given τ^2. However, the posterior distribution of τ^2 is of a complicated form, and the posterior moments of τ^2 had to be evaluated numerically. It was then shown by Lindley and Smith [1972] that, given τ^2, the posterior expectation of $\mu_i = \mu + \alpha_i$ is simply $\theta\bar{y}_{i.} + (1 - \theta)\,\bar{y}_{..}$, and hence the Bayes estimate of μ_i for squared error loss is

(1.1) $$\bar{y}_{i.}\ E\{\theta|\text{data}\} + \bar{y}_{..}\ (1 - E\{\theta|\text{data}\}),$$

where $\theta = J\tau^2/(1 + J\tau^2)$. See also Hill [1975]. Since the one-way balanced ran model can be regarded as consisting of a sample of J independent observations from an I dimensional multivariate normal distribution, where the usual random model distribution of the α_i is incorporated into the prior distribution for the mean vector $\underline{\mu} = (\mu_1,\ldots,\mu_I)'$, it follows that (1.1) will yield admissible estimators for $\underline{\mu}$ if the prior distributions employed are not too bizarre. However, just as with τ^2, the posterior distribution of θ is quite complicated, and the required posterior expectation of θ had to be calculated either numerically, or alternatively modal estimates employed, as, for example, by Lindley and Smith [1972]. Thus previous work in this area had stopped short of a full Bayesian analysis due to intrinsic mathematical difficulties. In the present report some new results concerning the Appell hypergeometric function are obtained which lead to both exact and approximate evaluations of the posterior moments of τ^2 and θ, and thus to Bayes estimates for these parameters under various loss functions, as well as for $\underline{\mu}$.

Section 2 sets forth the basic model and form of prior and posterior distributions. In Section 3 the relationship to the Appell hypergeometric function is explored, while Sections 4, 5, and 6, provide probabilistic representations for this function, upon which are based simple formulas for the exact evaluation of the posterior moments of τ^2 and θ. In Section 7 a limit theorem is proved which yields simple approximations to these moments, while in Section 8 upper and lower bounds are derived for the posterior moments. Section 9 then studies the behavior of certain roots of a quadratic equation, which are of central importance in the formulas for the posterior moments. Finally, in Section 10 some examples illustrate the application of the results of this report, and in Section 11, some general comments are made concerning these results.

2. MODEL AND FORM OF PRIOR AND POSTERIOR DISTRIBUTIONS

The model considered in this report is the balanced one-way random model

$$(2.1)\qquad y_{ij} = \mu + \alpha_i + \epsilon_{ij},\ \text{for } i=1,\ldots,I,\ j=1,\ldots,J,$$

where $\alpha_i \sim N(0,\sigma_\alpha^2)$, $\epsilon_{ij} \sim N(0,\sigma^2)$, and these random variables are mutually independent. As derived by Hill [1967], the likelihood function for $(\mu,\sigma_\alpha^2,\sigma^2)$, is

$$L(\mu,\sigma_\alpha^2,\sigma^2) \propto (\sigma^2)^{-I(J-1)/2} \exp[-SSW/2\sigma^2]$$
$$\times (\sigma^2 + J\sigma_\alpha^2)^{-I/2} \exp[-SSB/2(\sigma^2 + J\sigma_\alpha^2)]$$
$$\times \exp[-IJ(\bar{y}_{..} - \mu)^2/2(\sigma^2 + J\sigma_\alpha^2)],$$

where $-\infty < \mu < \infty$, $\sigma_\alpha^2 > 0$, $\sigma^2 > 0$, and

$$\bar{y}_{i.} = \sum_{j=1}^{J} y_{ij}/J,\ \bar{y}_{..} = \sum_{i=1}^{I} \bar{y}_{i.}/I,$$

$$SSW = \sum_{i,j} (y_{ij} - \bar{y}_{i.})^2,\ \text{and } SSB = J \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2 .$$

We consider the implications, for Bayesian inference and design, of the family of prior distributions with densities

$$(2.2)\qquad \rho(\mu,\sigma_\alpha^2,\sigma^2) \propto (\sigma_\alpha^2)^{-\lambda_\alpha/2 -1} \exp[-C_\alpha/2\sigma_\alpha^2]$$
$$\times (\sigma^2)^{-\lambda/2 -1} \exp[-C_o/2\sigma^2],\ \lambda_\alpha,\ C_\alpha,\ \lambda,\ C_o > 0,$$

i.e., with σ_α^2 and σ^2 independent a priori, each having an inverted gamma distribution, and with μ having the Jeffreys improper uniform distribution. This family of prior distributions was introduced and studied by Hill [1965] for inference in the one-way random model, and that study is continued here both for inference and design purposes. It is believed that Bayesian analysis is particularly appropriate in the one-way random model because there is typically substantial prior knowledge about the variance components σ_α^2 and σ^2, and that knowledge can be very effectively used either in inference or in design. The above family of prior distributions is, of course, not the only one of interest. However, the inverted gamma family is rich and flexible, and there is often reason to view the variance components as approximately independent a priori. In regard to μ, our use of the Jeffreys prior distribution is motivated by the fact that neither inference nor design is very sensitive to the choice of prior distribution for μ, and the flat prior may typically be regarded as a reasonable approximation via the stable estimation argument of L.J. Savage. Alternatively, it is easy to verify that use of the Jeffreys prior for μ is tantamount to basing inference about the variance components solely upon SSW and SSB, ignoring $\bar{y}_{\cdot\cdot}$, and thus obtaining the posterior density

$$\begin{aligned} \rho''(\sigma_\alpha^2,\sigma^2) \propto\ & (\sigma^2)^{-[I(J-1)+\lambda]/2\ -\ 1} \exp[-(SSW+C_o)/2\sigma^2] \\ & \times (\sigma^2 + J\sigma_\alpha{}^2)^{-(I-1)/2} \exp[-SSB/2(\sigma^2 + J\sigma_\alpha^2)] \\ & \times (\sigma_\alpha{}^2)^{-\lambda_\alpha/2\ -1} \exp[-C_\alpha/2\sigma_\alpha^2]. \end{aligned} \tag{2.3}$$

We are particularly interested in the implications of (2.3) with respect to the parameter $\tau^2 = \sigma_\alpha^2/\sigma^2$, or equivalently, $\Theta = J\tau^2/(1+J\tau^2)$, since, as shown by Hill [1965] and by Lindley and Smith [1972], conditional upon either of these parameters, the analysis of the model is extremely simple. Thus the complexity and difficulty in the analysis of the one-way random model is really contained in the nature and quantity of information that the data provide about τ^2 or Θ. It turns out that Θ is a somewhat more convenient parameter, and it is easy to verify that the posterior density for Θ, from (2.3), is

$$(2.4) \qquad p''(\Theta) \propto \frac{\Theta^{N_1/2-1}(1-\Theta)^{N_2/2-1}}{\{Q(\Theta)\}^{N_3/2}}, \quad 0 \leq \Theta \leq 1 ,$$

where $N_1 = IJ + \lambda - 1$, $N_2 = I + \lambda_\alpha - 1$, $N_3 = IJ + \lambda + \lambda_\alpha - 1$, and $Q(\Theta) = (SSW + C_o)\Theta + JC_\alpha(1-\Theta) + SSB\,\Theta(1-\Theta)$. This density was obtained by Culver [1971] in his University of Michigan doctoral dissertation.

Much of the remainder of this article is concerned with the exact and approximate mathematical analysis of distributions of the form (2.4). Since $Q(0) = JC_\alpha > 0$, $Q(1) = SSW + C_o$, and the coefficient of Θ^2 is negative, it follows that the quadratic $Q(\Theta)$ always has two real roots, say x_1 and x_2, with $x_1 > 1$ and $x_2 < 0$. With $q_1 = 1/x_1$, $q_2 = 1/(1-x_2)$, the posterior density for Θ can be written as

$$(2.5) \qquad p''(\Theta) \propto \frac{\Theta^{N_1/2-1}(1-\Theta)^{N_2/2-1}}{[(1-q_1\Theta)(1-q_2(1-\Theta))]^{N_3/2}}$$

If we define

$$f(A,B,C;q_1,q_2) = \int_0^1 \frac{\Theta^A(1-\Theta)^B}{\{(1-q_1\Theta)(1-q_2(1-\Theta))\}^C} \, d\Theta ,$$

the posterior moments of Θ and τ^2 are then

$$E(\Theta^k|\text{data}) = \frac{f(N_1/2 + k - 1,\ N_2/2 - 1,\ N_3/2;\ q_1,\ q_2)}{f(N_1/2 - 1,\ N_2/2 - 1,\ N_3/2;\ q_1,\ q_2)}$$

(2.6)

$$E(\tau^{2k}|\text{data}) = J^{-k} \frac{f(N_1/2 + k - 1,\ N_2/2 - k - 1,\ N_3/2;\ q_1,\ q_2)}{f(N_1/2 - 1,\ N_2/2 - 1,\ N_3/2;\ q_1,\ q_2)} ,$$

for all k such that the integrals are finite.

Thus the posterior moments of Θ and τ^2, which are crucial for inference and design purposes, are all expressible in terms of the functions $f(A,B,C;q_1,q_2)$. The study of such integrals is the primary aim of this report. In the next section such integrals are represented in terms of Appell hypergeometric functions, of which the hypergeometric function is a special case.

3. HYPERGEOMETRIC REPRESENTATIONS

We first observe that $f(A,B,C;q_1,q_2)$ is a generalization of the hypergeometric function $F(a,b,c;z) = 1 + \frac{ab}{1\times c} z + \frac{a(a+1)b(b+1)}{1\times 2\times c\times(c+1)} z^2 + \ldots$ For $c > b > 0$, then [Whittaker and Watson, p. 293],

$$F(a,b,c;z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 \frac{x^{b-1}(1-x)^{c-b-1}}{\{1-xz\}^a} dx ,$$

where $\Gamma(\cdot)$ is the usual gamma function. Hence, if for example $q_2 = 0$, then

$$f(A,B,C;q_1,q_2) = \frac{\Gamma(A+1)\Gamma(B+1)}{\Gamma(A+B+2)} F(C,A+1,A+B+2;q_1) ;$$

while if $q_1 = 0$, then

$$f(A,B,C;q_1,q_2) = \frac{\Gamma(A+1)\Gamma(B+1)}{\Gamma(A+B+2)} F(C,B+1,A+B+2;q_2) .$$

Here, as is true in almost all of our applications, it is assumed that $A + 1 > 0$ and $B + 1 > 0$.

The condition $q_i = 0$ is equivalent to the root $|x_i| = \infty$, so that when one or more of the roots is sufficiently extreme, then the integrals we require can be approximated by multiples of the hypergeometric function. In fact, if both $q_i = 0$, then

$$f(A,B,C;q_1,q_2) = \Gamma(A+1)\Gamma(B+1)/\Gamma(A+B+2) ,$$

and the posterior distribution of Θ is simply the beta distribution with parameters $N_1/2$ and $N_2/2$. Needless to say this situation rarely arises in applications, although having one extreme root is quite common, as will be discussed later.

In the general case, where neither q_i can be approximated by 0, the function $f(A,B,C;q_1,q_2)$ is a generalization of the hypergeometric function studied by Appell [Bailey, p. 77]. Even in the case where one of the roots is extreme, however, so that we are dealing with the ordinary hypergeometric function, little is known about many of the situations that are pertinent for the analysis of the one-way model. For example, if x_1 is very large but x_2 is near 0, i.e. $q_1 \simeq 0$, $q_2 \simeq 1$, we must evaluate $F(C,B+1,A+B+2;q_2)$ for q_2 nearly 1. This, however, cannot be approximated by setting $q_2 = 1$, since the parameters of the hypergeometric function here are such that the series is divergent when $q_2 = 1$.

In the next sections we proceed to develop new methods for the exact and approximate analysis of $f(A,B,C;q_1,q_2)$.

4. $\rho''(\theta)$ AS MIXTURE OF BETA DISTRIBUTIONS

For a Bernoulli sequence with probability of success p, let S_r be the number of failures preceding the r^{th} success. Then $\Pr\{S_r = i\} = \binom{r+i-1}{i} p^r(1-p)^i$, $i=0,1,2,\ldots$, defines the Pascal distribution with parameters p and r. In fact this formula defines a distribution (negative binomial) even when $r > 0$ is not integral, and we shall use the above notation S_r whether or not r is integral.

Expanding the denominator in (2.5) yields

$$\rho''(\theta) \propto \theta^{N_1/2-1}(1-\theta)^{N_2/2-1} \sum_{i=0}^{\infty} \binom{r+i-1}{i} q_1^i \theta^i \sum_{j=0}^{\infty} \binom{r+j-1}{j} q_2^j (1-\theta)^j$$

(4.1)

$$\propto \theta^{N_1/2-1}(1-\theta)^{N_2/2-1} E\{\theta^{S_r^{(1)}} (1-\theta)^{S_r^{(2)}}\},$$

for $q_1\theta < 1$ and $q_2(1-\theta) < 1$, where $S_r^{(i)}$ has the negative binomial distribution with parameters $p_i = 1-q_i$ and $r = N_3/2$, $i = 1,2$, and $S_r^{(1)}$ is independent of $S_r^{(2)}$. Thus $\rho''(\theta)$ can be viewed as a mixture of beta distributions. Note that when both $p_i = 1$ the mixture again reduces to a beta distribution with parameters $N_1/2$ and $N_2/2$.

From (4.1) it follows that evaluation of $f(A,B,C;q_1,q_2)$ is equivalent to evaluation of a sum of the form

$$\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \binom{r+i-1}{i} q_1^i \binom{r+j-1}{j} q_2^j \frac{\Gamma(A+1+i)\Gamma(B+1+j)}{\Gamma(A+B+2+i+j)}$$

and

$$E\{\theta|\text{data}\} = \frac{E\{\int_0^1 \theta^{N_1/2 + S_r^{(1)}} (1-\theta)^{N_2/2 + S_r^{(2)} - 1} d\theta\}}{E\{\int_0^1 \theta^{N_1/2 + S_r^{(1)} - 1} (1-\theta)^{N_2/2 + S_r^{(2)} - 1}\}}$$

(4.2)

$$= \frac{E\{\Gamma(N_1/2 + 1 + S_r^{(1)})\Gamma(N_2/2 + S_r^{(2)})/\Gamma(N_1/2 + N_2/2 + S_r^{(1)} + S_r^{(2)} + 1)\}}{E\{\Gamma(N_1/2 + S_r^{(1)})\Gamma(N_2/2 + S_r^{(2)})/\Gamma(N_1/2 + N_2/2 + S_r^{(1)} + S_r^{(2)})\}}.$$

Although we have developed direct methods for evaluating such sums and expectations, we shall not pursue them in this report, since the methods of Section 5 based upon integrated generating functions are generally more useful for our purposes. The representation of the posterior distribution as an infinite mixture of beta distributions is, however, of interest in its own right, and aids in developing intuition into the mathematical nature of the problem.

5. INTEGRATED GENERATING FUNCTIONS

A great deal of insight into the nature of $f(A,B,C;q_1,q_2)$ and the equivalent summation is gained by a probabilistic interpretation as an integrated generating function.

Let X be any random variable taking on only non-negative integral values. The generating function for X is the function $M(t) = E\{t^X\}$ for $0 \leq t \leq 1$. Clearly

$$\int_0^1 M(t)\, dt = \int_0^1 E\{t^X\}\, dt = E\{(X+1)^{-1}\} ,$$

where the interchanging of the order of expectation and integration is justified by Fubini's Theorem. More generally, for any $a > 0$ and integral $b \geq 0$,

$$\int_0^1 t^{a-1}(1-t)^b M(t)\, dt = \Gamma(b+1)\, E\{(X+a)(X+a+1)\times\ldots\times(X+a+b)\}^{-1}. \tag{5.1}$$

It will now be shown that $f(A,B,C;q_1,q_2)$ can be represented as such an integrated generating function.

If, as in Section 4, S_r has the negative binomial distribution with parameters p and r, then it is easily verified that the generating function of S_r is $(\frac{p}{1-qt})^r$. Now let $S_r^{(1)}$ have a negative binomial distribution with parameters p_1 and r, and let Z have a negative binomial distribution with parameters p_2 and $S_r^{(1)}$, so that Z can be viewed as the number of failures preceding the $S_r^{(1)}$ th success in a Bernoulli sequence with probability p_2 of success on each trial. By the well-known formula for compound distributions (see Feller [1950, p. 223]), the generating function for Z is then $\left(\dfrac{p_1}{1-q_1 \dfrac{p_2}{1-q_2 t}}\right)^r$. Now let H = A+B-C+2. Making the substitution $t = \theta/(p_2 + q_2\theta)$ in the integral defining $f(A,B,C;q_1,q_2)$ then yields

$$(5.2) \qquad f(A,B,C;q_1,q_2) = p_1^{-C}\, p_2^{-(B+1)} \int_0^1 t^A (1-t)^B \left\{\frac{p_2}{1-q_2 t}\right\}^H \left(\frac{p_1}{1-q_1 \dfrac{p_2 t}{1-q_2 t}}\right)^C dt.$$

By the previous results for $S_r^{(1)}$ and Z, the generating function of $X = Z + S_r^{(1)} + S_H^{(2)}$ is $M(t) = E\{t^X\} = \{\frac{p_2}{1-q_2 t}\}^H E\{t^{S_r^{(1)} + Z}\}$

$$= \left\{\frac{p_2}{1-q_2 t}\right\}^H E\{t^{S_r^{(1)}} E[t^Z | S_r^{(1)}]\}$$

$$= \left\{\frac{p_2}{1-q_2 t}\right\}^H E\{t^{S_r^{(1)}} \left[\frac{p_2}{1-q_2 t}\right]^{S_r^{(1)}}\}$$

$$= \left\{\frac{p_2}{1-q_2 t}\right\}^H E\left[\frac{p_2 t}{1-q_2 t}\right]^{S_r^{(1)}}$$

$$= \left\{\frac{p_2}{1-q_2 t}\right\}^H \left(\frac{p_1}{1 - \dfrac{q_1\, p_2 t}{1-q_2 t}}\right)^C.$$

From (5.1) and (5.2) it follows that when B is integral,

$$f(A,B,C;q_1,q_2) = p_1^{-C}\, p_2^{-(B+1)} \Gamma(B+1)\, E\{(X+A+1)\times\ldots\times(X+A+1+B)\}^{-1}$$

(5.3) and

$$E(\theta|\text{data}) = \frac{E\{(X_1+A+2)\times\ldots\times(X_1+A+2+B)\}^{-1}}{E\{(X+A+1)\times\ldots\times(X+A+1+B)\}^{-1}}, \text{ where } X_1 \text{ is defined}$$

like X but with H replaced by H + 1.

By symmetry of (2.5), $f(A,B,C;q_1,q_2) = f(B,A,C;q_2,q_1)$, so that if A is an integer, then also

$$f(A,B,C;q_1,q_2) = f(B,A,C;q_2,q_1)$$

(5.4) $$= p_2^{-C}\, p_1^{-(A+1)}\Gamma(A+1)\; E\{(X^*+B+1)\times\ldots\times(X^*+A+B+1)\}^{-1}, \text{ and}$$

$$E(\theta|\text{data}) = \frac{E\{(X_1^*+B+2)\times\ldots\times(X_1^*+A+B+2)\}^{-1}}{E\{(X^*+B+1)\times\ldots\times(X^*+A+B+1)\}^{-1}},$$

where $X^* = Z^* + S_r^{(2)} + S_H^{(1)}$, and Z^* is defined like Z except that p_1 and p_2 are interchanged.

Now let $q = q_1+q_2-q_1q_2$ and $p = 1-q = p_1p_2$. From (5.2),

$$f(A,B,C;q_1,q_2) = p_2^{-\lambda}\alpha/2 \int_0^1 t^A(1-t)^B(1-q_2t)^{C-H}(1-qt)^{-C}\, dt,$$

(5.5)

$$= p_2^{-\lambda}\alpha/2 \sum_{i=0}^{\infty} (-1)^i\binom{C-H}{i}q_2^i \int_0^1 t^{A+i}(1-t)^B(1-qt)^{-C}\, dt,$$

where the summation terminates at C-H if C-H is integral. Alternatively, expanding $(\frac{1-qt}{1-q_2t})^{-C} = (1 - \frac{tq_1p_2}{1-q_2t})^{-C}$ in (5.5) yields

(5.6) $$f(A,B,C;q_1,q_2) = p_2^{-\lambda}\alpha/2 \sum_{j=0}^{\infty} \binom{C+j-1}{j} q_1^j\, p_2^j \int_0^1 t^{A+j}(1-t)^B(1-q_2t)^{-H-j}\, dt.$$

Both of the expansions will prove useful in studying $f(A,B,C;q_1,q_2)$. First suppose that one q_i is 0. As was noted earlier, in this case $f(A,B,C;q_1,q_2)$ reduces to an ordinary hypergeometric function. Suppose

$q_1 = 0$. From (5.2), expansion of $(1-t)^B$ yields

$$(5.7) \qquad f(A,B,C;0,q_2) = p_2^{-(B+1)} \sum_{i=0}^{\infty} (-1)^i \binom{B}{i} \int_0^1 t^{A+i} \left\{\frac{p_2}{1-q_2 t}\right\}^H dt ,$$

where the summation terminates at B if B is an integer. From (5.1) and the fact that $\{\frac{p_2}{1-q_2 t}\}^H$ is the generating function of the negative binomial distribution with parameters p_2 and H, we have

$$(5.8) \qquad f(A,B,C;0,q_2) = p_2^{-(B+1)} \sum_{i=0}^{\infty} (-1)^i \binom{B}{i} E(S_H^{(2)}+A+i+1)^{-1} ,$$

which is a representation for the ordinary hypergeometric function in terms of reciprocal moments of a translated negative binomial random variable. Now define $\Psi(q;k,r) = \int_0^q x^k(1-x)^{-r-1} dx$, where $0 \leq q < 1$, $k > -1$, and r is arbitrary. Then if S_r has a negative binomial distribution with parameters p and r,

$$E(S_r+k+1)^{-1} = \int_0^1 t^k \left\{\frac{p}{1-qt}\right\}^r dt$$

$$= p^r q^{-(k+1)} \int_0^q x^k (1-x)^{-r} dx = p^r q^{-(k+1)} \Psi(q;k,r-1).$$

Applying this result in (5.8) yields

$$(5.9) \qquad f(A,B,C;0,q_2) = p_2^{-\lambda_\alpha/2} q_2^{-(A+1)} \sum_{i=0}^{\infty} (-1)^i \binom{B}{i} q_2^{-i} \Psi(q_2;A+i,H-1).$$

Note also, from (5.7), that, when B is an integer,

$$f(A,B,C;0,q_2) = p_2^{-(B+1)} \int_0^1 (1-t)^B t^A \left\{\frac{p_2}{1-q_2 t}\right\}^H dt$$

(5.10)
$$= p_2^{(H-1)} q_2^{-(A+B+1)} (-1)^B \int_0^{q_2} \left\{1 - \frac{(1-x)}{p_2}\right\}^B x^A (1-x)^{-H} dx$$

$$= p_2^{(H-1)} q_2^{-(A+B+1)} (-1)^B \sum_{i=0}^{B} (-1)^i \binom{B}{i} p_2^{-i} \Psi(q_2;A,H-1-i),$$

with

$$\sum_{i=0}^{B} (-1)^i \binom{B}{i} q_2^{-i} \Psi(q_2;A+i,H-1)$$

$$= (-1)^B (p_2/q_2)^B \sum_{i=0}^{B} (-1)^i \binom{B}{i} p_2^{-i} \Psi(q_2;A, H-1-i) .$$

Now return to the general case. From (5.6) and (5.10), if B is an integer

$$f(A,B,C;q_1,q_2) = p_2^{H-B-1} \sum_{j=0}^{\infty} \binom{C+j-1}{j} q_1^j p_2^j \sum_{i=0}^{B} (-1)^i \binom{B}{i} q_2^{-(A+i+j+1)} \Psi(q_2;A+i+j,H+j-1)$$

(5.11)
$$= p_2^{H-1} q_2^{-(A+B+1)} (-1)^B \sum_{j=0}^{\infty} \binom{C+j-1}{j} q_1^j (p_2/q_2)^j \sum_{i=0}^{B} (-1)^i \binom{B}{i} p_2^{-i} \Psi(q_2;A+j, H+j-1-i).$$

On the other hand, from (5.5) with B and C-H integers, we obtain

$$f(A,B,C;q_1,q_2) = p_2^{H-B-1} \sum_{j=0}^{C-H} (-1)^j \binom{C-H}{j} q_2^j \sum_{i=0}^{B} (-1)^i \binom{B}{i} q^{-(A+i+j+1)} \times \Psi(q;A+i+j,C-1). \quad (5.12)$$

These formulas reduce $f(A,B,C;q_1,q_2)$ to a (sometimes finite) sum of terms involving the function $\Psi(q;s,t)$ for various s and t. In the next section we study properties of the function $\Psi(q;s,t)$.

6. THE FUNCTION $\psi(q;s,t)$

The results of the previous section show that the study of $f(A,B,C;q_1,q_2)$ can be reduced to that of the function

$$\psi(q;s,t) = \int_0^q y^s (1-y)^{-t-1} dy,$$

where $s > -1$, $0 \leq q < 1$, and t is any real number. When t is negative clearly $\psi(q;s,t)$ is a multiple of the incomplete beta function; while for $t \geq 0$ we have an interpretation in terms of the expectation of the reciprocal of a translated negative binomial random variable, as given by (5.9). We now derive further properties of this function which will be useful either for insight or for computational purposes.

Expanding $y^s = (1+y-1)^s$ by the binomial formula yields

$$\begin{aligned} \psi(q;s,t) &= \sum_{i=0}^{\infty} (-1)^i \binom{s}{i} \int_0^q (1-y)^{i-t-1} dy \\ &= \sum_{i=0}^{\infty} (-1)^i \binom{s}{i} \left[\frac{1-p^{i-t}}{i-t}\right] , \end{aligned} \tag{6.1}$$

where $p = 1-q$, and $(1-p^0)/0$ is defined as $\ell n(1/p)$, if it should occur in the sum. When t is positive it follows from (6.1) that $\psi(q;s,t) \sim t^{-1} p^{-t}$ as $q \to 1$, where, here and throughout, the symbol "$\sim$" means that the ratio tends to one. When $t = 0$, $\psi(q;s,t) \sim \ell n(1/p)$ as $q \to 1$; while for negative t, $\psi(q;s,t)$ tends to $\Gamma(s+1)\Gamma(-t)/\Gamma(s+1-t)$, as follows immediately from its relationship with the incomplete beta function. We note that the summation in (6.1) is finite if s is an integer; while, if $w = t-s-1$ is a non-negative integer,

$$\psi(q;s,t) = \int_0^{q/p} y^s (1+y)^w \, dy = \sum_{i=0}^{w} \binom{w}{i} [q/p]^{s+i+1} (s+i+1)^{-1}$$

is again a finite sum. In these cases computation of $\psi(q;s,t)$ is particularly easy.

The function $\psi(q;s,t)$ also satisfies certain recursion formulas which are of use either for computation or for insight. Integration by parts yields

(6.2) $$\psi(q;s,t) = t^{-1} [q^s p^{-t} - s\, \psi(q;s-1,t-1)] \text{ if } t \neq 0.$$

Also

(6.3) $$\psi(q;s+1,t) - \psi(q;s,t) = -\psi(q;s,t-1),$$

from which

(6.4) $$\psi(q;s,0) = \psi(q;s-[s],0) - \sum_{i=0}^{[s]-1} q^{s-i}/(s-i),$$

where [s] denotes the largest integer less than or equal to s. Other formulas of use are

(6.5) $$\psi(q;s,-1) = q^{s+1} (s+1)^{-1} \text{ for all } s > -1,$$

(6.6) $$\psi(q;s,-\tfrac{1}{2}) = (1+2s)^{-1} [2s\psi(q;s-1,-\tfrac{1}{2}) - 2\, q^s p^{\frac{1}{2}}],$$

for $s \neq -\frac{1}{2}$, and

(6.7) $$\psi(q;s,t) = \sum_{i=0}^{\infty} \binom{t+i}{i} q^{s+i+1} (s+i+1)^{-1}, \text{ for } t \geq 0.$$

Some special values that can be used to start off the recursions are

$$\psi(q;0,0) = \ln(1/p), \quad \psi(q;0,-\tfrac{1}{2}) = 2[p^{\frac{1}{2}} - 1],$$

$$\text{(6.8)} \quad \psi(q;-\tfrac{1}{2},-\tfrac{1}{2}) = 2 \arcsin (q^{\frac{1}{2}}),$$

$$\psi(q;-\tfrac{1}{2},0) = 2 \sum_{j=0}^{\infty} q^{j+\frac{1}{2}} (1+2j)^{-1} = \ln[(1+q^{\frac{1}{2}})/(1-q^{\frac{1}{2}})].$$

In most applications it will suffice to choose λ and λ_α so that the requisite s and t are either integral or half of an odd integer, in which case the starting values given by (6.8) are particularly useful.

7. $E(\Theta|\text{data})$ for Small p_1, p_2

For applications to the random model the most interesting, and also the most difficult, cases are those in which one or both of the p_i are small. We have already seen in Section 4 that when both p_i are nearly 1, the mixture of beta distributions representing the posterior density of Θ can be approximated by a single beta distribution, since the $S_r^{(i)}$ will then be 0 with high probability. However, when one or both of the p_i are small, one or both of the $S_r^{(i)}$ will be large with high probability, and the mixture cannot be adequately approximated by a single beta distribution.

Since $N_3 > N_1$ and $N_3 > N_2$, it is apparent from (2.5) that the posterior distribution of Θ becomes degenerate at 1 if $p_1 \to 0$ with p_2 fixed, and becomes degenerate at 0 if $p_2 \to 0$ with p_1 fixed. Since $0 \leq \Theta \leq 1$, also $E(\Theta|\text{data})$ goes to 1 or 0, respectively, in these cases. If both p_i tend to 0, however, it is not clear from (2.5) how the posterior distribution of Θ will behave. We now obtain asymptotic expressions for $E(\Theta|\text{data})$ when one or both of the p_i tend to 0. These expressions lead to simple approximations that will be suitable in most applications to the random model analysis of variance.

Let $D = C-H = 2C - A - B - 2$, and define the functions

$$(7.1) \quad G(A,B,C,D;q_1,q_2) = \int_0^1 t^A (1-t)^B (1-qt)^{-C} (1-q_2t)^D \, dt.$$

From (5.5) and (6.1),

$$G(A,B,C,D;q_1,q_2) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} (-1)^{i+j} \binom{B}{i}\binom{D}{j} q_2^j \int_0^1 t^{A+i+j} (1-qt)^{-C} dt$$

(7.2)
$$= q^{-(A+1)} \sum_{\ell=0}^{\infty} (-1)^{\ell} \left[\frac{p^{\ell+1-C}-1}{C-\ell-1}\right] V(A,B,D,\ell;1/q,q_2/q),$$

where $V(A,B,D,\ell;x,y)$ is formally defined as the series

$$\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} (-1)^{i+j} \binom{B}{i} \binom{D}{j} \binom{A+i+j}{\ell} x^i y^j ,$$

with ℓ a non-negative integer.

It is straightforward, however, to show that in fact

$$\text{(7.3)} \quad V(A,B,D,\ell;x,y) = (1-x)^B (1-y)^D \sum_{i=0}^{\ell-j} \sum_{j=0}^{\ell} (-1)^{i+j} \binom{B}{i}\binom{D}{j}\binom{A}{\ell-i-j} (x/1-x)^i (y/1-y)^j$$

which is a finite sum and always convergent for $x \neq 1$, $y \neq 1$.

From (2.6), we have

$$E(\Theta|\text{data}) = p_2 \frac{G(A+1,B,C,D-1;q_1,q_2)}{G(A,B,C,D;q_1,q_2)} , \text{ and}$$

(7.4)
$$E(1-\Theta|\text{data}) = \frac{G(A,B+1,C,D-1;q_1,q_2)}{G(A,B,C,D;q_1,q_2)} .$$

The asymptotic behavior of these posterior expectations as one or both p_i tend to 0 is now given by Theorem 1. In the theorem we write $E(\Theta)$ for $E(\Theta|\text{data})$.

Theorem 1. If $p_1^{C-B-2}\ p_2^{-(C-A-1)} \to 0$, then $E(1-\Theta) \sim p_1(B+1)/(C-B-2)$; if $p_2^{C-A-2}\ p_1^{-(C-B-1)} \to 0$, then $E(\Theta) \sim p_2(A+1)/(C-A-2)$.

Proof. First note that by symmetry the assertion regarding $E(\Theta)$ follows from that for $E(1-\Theta)$, so we consider only the latter. There are a number of special cases to consider in proving the Theorem, depending upon which, if any, of A,B,C,D, are integers, and whether one or both p_i tend to 0. However, the proofs for other cases are minor variants of the proof for the case in which A, B and D are integers and both p_i tend to 0. This case will now be proved.

From (7.2) and (7.3),

$$G(A,B,C,D;q_1,q_2) = (-1)^B\ p^B\ (q_1p_2)^D\ q^{-(A+1+B+D)}$$

$$\times \sum_{\ell=0}^{A+B+D} (-1)^\ell \left[\frac{p^{\ell+1-C}-1}{C-\ell-1}\right] \sum_{i=0}^{B} \sum_{j=0}^{D} (-1)^j \binom{B}{i}\binom{D}{j}\binom{A}{\ell-i-j} p^{-i} (q_2/q_1p_2)^j$$

Now for $\ell \leq B$, the dominant term in the sum on the right-hand side is obtained by taking $i = \ell$ and $j = 0$, so that

$$\sum_{\ell=0}^{B} (-1)^\ell \left[\frac{p^{\ell+1-C}-1}{C-\ell-1}\right] \sum_{i=0}^{B} \sum_{j=0}^{D} (-1)^j \binom{B}{i}\binom{D}{j}\binom{A}{\ell-i-j} p^{-i} (q_2/q_1p_2)^j$$

$$\sim \sum_{\ell=0}^{B} (-1)^\ell \binom{B}{\ell} \ (C-\ell-1)^{-1}\ p^{1-C} \text{ as } p_1,p_2 \to 0.$$

On the other hand, for $\ell \geq B + D$, the corresponding dominant term is obtained by taking $i = B$ and $j = D$, and

$$\sum_{\ell=B+D}^{A+B+D} (-1)^\ell \left[\frac{p^{\ell+1-C}-1}{C-\ell-1}\right] \sum_{i=0}^{B} \sum_{j=0}^{D} (-1)^j \binom{B}{i}\binom{D}{j}\binom{A}{\ell-i-j} p^{-i} (q_2/q_1p_2)^j$$

$$\sim \sum_{\ell=B+D}^{A+B+D} \left[\frac{(-1)^{\ell+1+D}}{C-\ell-1}\right] \binom{A}{\ell-B-D}\ p^{-B}\ p_2^{-D}.$$

Since $p^{C-1} \; p^{-B} \; p_2^{-D} = p_1^{C-B-1} \; p_2^{-(C-A-1)}$ goes to 0 by assumption, it follows that the terms for $\ell \leq B$ dominate those for $\ell \geq B + D$; and similarly, it can be shown that they dominate the terms with $B < \ell < B + D$. Hence

$$G(A,B,C,D;q_1,q_2) \sim (-1)^B \; p^{B+1-C} \; p_2^D \sum_{\ell=0}^{B} (-1)^{\ell} \binom{B}{\ell} (C-\ell-1)^{-1} ,$$

if $p_1, p_2 \to 0$ in such a way that $p_1^{C-B-1} \; p_2^{-(C-A-1)} \to 0$. Applying the same argument to $G(A,B+1,C,D-1;q_1,q_2)$, $G(A,B+1,C,D-1;q_1,q_2)$

$$\sim (-1)^{B+1} \; p^{B+2-C} \; p_2^{D-1} \sum_{\ell=0}^{B+1} (-1)^{\ell} \binom{B+1}{\ell} (C-\ell-1)^{-1}, \text{ if } p_1^{C-B-2} \; p_2^{-(C-A-1)} \to 0.$$

Since

$$\sum_{\ell=0}^{B} (-1)^{\ell} \binom{B}{\ell} (C-\ell-1)^{-1} = \int_0^1 (1-t^{-1})^B \; t^{C-2} \, dt$$

$$= (-1)^B \Gamma(B+1) \Gamma(C-B-1)/\Gamma(C),$$

(7.4) then yields $E(1-\Theta) \sim p_1 (B+1)/(C-B-2)$, as claimed.

It may be noted that in applications the case in which $p_1^{C-B-2} \times p_2^{-(C-A-1)} \to 0$ is typically the most relevant, since B is typically much smaller than A. In terms of the parameters of the prior distribution, the asymptotic values in the two cases covered by the Theorem are

$$E(\Theta) \sim p_2 \;\; (IJ+\lambda-1)/(\lambda_{\alpha}-2) \; ,$$

$$E(1-\Theta) \sim p_1 \;\; (I-1+\lambda_{\alpha})/[I(J-1)+\lambda-2].$$

An approximation to $E(\theta)$ which combines both of these asymptotic values, and is appropriate when one p_i is small and the other large, is

$$(7.5) \qquad E(\theta) \simeq \frac{A+1+(C-B-2)q_1/p_1}{A+B+2+(C-B-2)q_1/p_1+(C-A-2)q_2/p_2}$$

$$= \frac{N_1+[I(J-1)+\lambda-2]q_1/p_1}{N_1+N_2+[I(J-1)+\lambda-2]q_1/p_1+[\lambda_\alpha-2]q_2/p_2}$$

This approximation yields the exact value $N_1/(N_1+N_2)$ when $p_1=p_2=1$, and has the same asymptotic values given by Theorem 1 when one of the p_i goes to 0 and the other to 1. Of course (7.5) only applies when $\lambda_\alpha > 2$. In general it yields a good approximation whenever p_2 is not extremely small.

Finally, we note that when A, B, C, and D are integers, from (7.2), (7.3), and (7.4) we obtain the exact formula

$$(7.6) \qquad E(1-\theta) = (-1)p_1/q_1 \frac{\sum_{\ell=0}^{A+B+D} (-1)^\ell \left[\frac{p^{\ell+1-C}-1}{C-\ell-1}\right] \sum_{i=0}^{B+1} \sum_{j=0}^{D-1} (-1)^j \binom{B+1}{i}\binom{D-1}{j}\binom{A}{\ell-i-j} p^{-i}\delta^j}{\sum_{\ell=0}^{A+B+D} (-1)^\ell \left[\frac{p^{\ell+1-C}-1}{C-\ell-1}\right] \sum_{i=0}^{B} \sum_{j=0}^{D} (-1)^j \binom{B}{i}\binom{D}{j}\binom{A}{\ell-i-j} p^{-i}\delta^j},$$

where $\delta = (q_2/p_2q_1)$.

This formula makes computation very simple in many applications. Using linear interpolation, we can also employ (7.6) for non-integral values of A,B,C, and D.

8. BOUNDS FOR $E(\Theta|\text{data})$

It is possible to obtain simple and useful bounds for the posterior expectation of Θ. These bounds are in fact closely related to the asymptotic values of Theorem 1. Throughout this section we write $E(\Theta)$ for $E(\Theta|\text{data})$.

Consider first the case $p_1 = 1$. From (2.6) and (5.2),

$$E(\Theta)/p_2 = W(q_2)\ [1 - q_2\, W(q_2)]^{-1}\ ,\quad \text{where}$$

$$W(q_2) = \frac{\int_0^1 t^{A+1}(1-t)^B(1-q_2t)^{-H-1}\,dt}{\int_0^1 t^A\,(1-t)^B(1-q_2t)^{-H-1}\,dt}\ .$$

Since $H = A+B+2-C = (I-1)/2 \geq 0$, it is easily seen that $W(q_2)$ is a non-decreasing function of q_2, for $0 \leq q_2 \leq 1$. But $W(0) = (A+1)/(A+B+2)$, and if $B-H = C-A-2>0$, then $W(1) = (A+1)/(C-1)$. Thus for $p_1 = 1$ and $\lambda_\alpha > 2$,

(8.1) $$p_2(A+1)/(A+B+2) \leq E(\Theta) \leq p_2(A+1)/(C-A-2).$$

Since $E(1-\Theta) = \dfrac{f(A,B+1,C;q_1,q_2)}{f(A,B,C;q_1,q_2)} = \dfrac{f(B+1,A,C;q_2,q_1)}{f(B,A,C;q_2,q_1)}$, the above bounds in terms of p_2 can be converted into bounds in terms of p_1, yielding, for $p_2 = 1$,

(8.2) $$1-p_1(B+1)/(C-B-2) \leq E(\Theta) \leq 1-p_1(B+1)/(A+B+2).$$

Clearly $E(\Theta)$ is an increasing function of q_1, for any q_2, and is a decreasing function of q_2, for any q_1, so we have proved

Theorem 2. For any q_1 and q_2 satisfying $0 \leq q_1 \leq 1$, $0 \leq q_2 \leq 1$,

$$p_2(A+1)/(A+B+2) \leq E(\Theta) \leq 1 - p_1(B+1)/(A+B+2).$$

It is worth noting that the upper bound for $E(1-\Theta)$ given by (8.2), and the upper bound for $E(\Theta)$ given by (8.1), are precisely the asymptotic values given in Theorem 1.

Sometimes the universal upper and lower bounds given in Theorem 2 are very nearly equal, in which case $E(\Theta)$ is very nearly determined. The case in which the universal upper and lower bounds provide the least information is when both p_i are nearly 0, since the bounds then become 0 and 1. However, in this case the asymptotic values given in Theorem 2 will typically be appropriate.

9. THE ROOTS AND PRIOR PARAMETERS

According to (2.5), for fixed I, J, λ, λ_α, the posterior distribution of θ is completely determined by the q_i, or equivalently, by the roots x_i of the quadratic equation $Q(\theta) = 0$. These roots are functions of the data SSW and SSB, and of the parameters C_0 and C_α of the prior distribution, and are explicitly given as

$$(9.1) \qquad x_i = \frac{SSW + SSB + C_0 - JC_\alpha \pm (4JC_\alpha SSB + [SSW + SSB + C_0 - JC_\alpha]^2)^{\frac{1}{2}}}{2(SSB)},$$

where always $x_1 > 1$ and $x_2 < 0$. Since $q_1 = x_1^{-1}$, $q_2 = (1 - x_2)^{-1}$, the interesting and delicate cases in which the p_i are small occur when x_1 is nearly 1 or when x_2 is nearly 0.

Noting that the sum of roots is $x_1 + x_2 = 1 + (SSW + C_0 - JC_\alpha)/SSB$, and that the product of roots is $x_1 x_2 = -JC_\alpha/SSB$, we have

$$(9.2) \qquad \begin{aligned} p_2/q_2 &= p_1/q_1 - (SSW + C_0 - JC_\alpha)/SSB, \\ p_2/q_2 &= q_1 JC_\alpha/SSB, \text{ and} \\ p_1/p_2 &= (SSW + C_0)/JC_\alpha. \end{aligned}$$

From (9.2) it follows that both p_i will be small if and only if both JC_α/SSB and $(SSW + C_0)/SSB$ are small. Since $Q(0) = JC_\alpha$, $Q(1) = SSW + C_0$, and $Q(1/2) = [Q(0) + Q(1)]/2 + SSB/4$, when SSB is small $Q(\theta)$ becomes nearly linear, with p_1 nearly 1 if $Q(1) \geq Q(0)$, and p_2 nearly 1 if $Q(0) \geq Q(1)$. It is important here to note the sensitivity of p_2 to the choice of C_α. Thus, $p_2 \to 0$ as $C_\alpha \to 0$, and $p_2 \to 1$ as $C_\alpha \to \infty$.

Let us now consider ways in which values of the parameters λ, C_0, λ_α, and C_α, of the prior distribution of the variance components, can be chosen. Clearly all these parameters must be positive in order that the prior distribution be proper. When $\lambda > 2$ the prior expectation of σ^2 is $C_0/(\lambda - 2)$, when $\lambda > 4$ the prior variance of σ^2 is $2C_0^2/(\lambda - 2)^2(\lambda - 4)$, and in any case the mode of the prior distribution of σ^2 is $C_0/(\lambda + 2)$. The same relationships hold for the prior distribution of σ_α^2 also, and since σ^2 and σ_α^2 are independent a priori, it follows that the prior distribution of σ_α^2/σ^2 is that of $[\lambda C_\alpha/(\lambda_\alpha C_0)]F_{\lambda,\lambda_\alpha}$, where $F_{\lambda,\lambda_\alpha}$ denotes a random variable having the (generalized) F distribution with λ and λ_α degrees of freedom (λ and λ_α are not necessarily integral). Hence the prior expectation of σ_α^2/σ^2 is $[\lambda C_\alpha/(\lambda_\alpha - 2)C_0]$, and the prior mode is $[C_\alpha(\lambda - 2)/(\lambda_\alpha + 2)C_0]$ if $\lambda>2$. These relationships can be used to choose values for λ, C_0, λ_α and C_α, which are in accord with one's prior knowledge about σ^2 and σ_α^2.

An important difference between the prior parameters λ and C_0 for σ^2, and the parameters λ_α and C_α for σ_α^2, is that ordinarily the posterior distribution of θ has only a slight dependence upon the choice of λ and C_0, but is very dependent upon the choice of λ_α and C_α. Thus there is usually substantial robustness in regard to the choice of λ and C_0, but not in regard to the choice of λ_α and C_α. The insensitivity to λ and C_0 may be seen from the fact that they affect the posterior distribution of θ only by virtue of the equations $Q(1) = SSW + C_0$, $N_1 = IJ + \lambda - 1$, and $N_3 = IJ + \lambda + \lambda_\alpha - 1$. Since the prior mode of SSW is $\simeq I(J - 1)C_0/(\lambda + 2)$, it follows that whenever λ is small compared to $I(J - 1)$, we anticipate that $Q(1) \simeq SSW$, $N_1 \simeq IJ - 1$, $N_3 \simeq IJ + \lambda_\alpha - 1$, and it is very nearly as though $\lambda = C_0 = 0$.

On the other hand there is a great deal of sensitivity to the choice of λ_α and C_α, since, as observed earlier, p_2 and hence also $E\{\theta|\text{data}\}$ go to 0 as $C_\alpha \to 0$, while in the equation $N_2 = I + \lambda_\alpha - 1$ often λ_α is not small compared to I. In particular, it would be absurd to take $\lambda_\alpha = C_\alpha = 0$, since this would imply $E\{\theta|\text{data}\} = 0$. Thus in applications it is necessary to exercise great care in the choice of λ_α and C_α.

10. EXAMPLES

Some examples illustrating how the previous results can be applied are now given.

Example 1. $I = 3$, $J = 3$, $\lambda = C_0 = 0$, $\lambda_\alpha = 2$, $C_\alpha = 4$, MSB = 5, and MSW = 1. Then $p_1 = .2387$, $p_2 = .4774$, and using (7.6) we easily obtain $E(\Theta|\text{data}) = .84$. The universal lower and upper bounds given in Theorem 2 are .32 and .92, respectively, while the pertinent asymptotic value is $1 - p_1 \times (B + 1)/(C - B - 2) = 1 - p_1 = .76$. Note, however, that $p_1^{C-B-2} p_2^{-(C-A-1)} = p_1^2/p_2 = .12$, which is not particularly small, so that the asymptotic value could not be expected to yield a very good approximation. Formula (7.5) does not apply in this example because $\lambda_\alpha = 2$.

Example 2. $I = 20$, $J = 10$, $\lambda = C_0 = 0$, $\lambda_\alpha = 8$, $C_\alpha = 10$, MSB = 10, and MSW = 1. Then $p_1 = .4206$, $p_2 = .2337$, and by numerical integration we obtain $E(\Theta|\text{data}) = .91$. The universal lower and upper bounds are .21 and .95, respectively. In this example $p_1^{C-B-2} p_2^{-(C-A-1)} = p_1^{89}/p_2^{4}$ is negligible, so we anticipate that the asymptotic value $1 - p_1(B+1)/(C-B-2)$ will yield a good approximation, which it does, namely, .94. Formula (7.5) does even better, giving .90.

Example 3. $I = 5$, $J = 2$, $\lambda = C_0 = 0$, $\lambda_\alpha = 8$, $C_\alpha = 1$, MSB = 10, and MSW = 1. Then $p_1 = .1069$, $p_2 = .0427$, and $E(\Theta|\text{data}) = .08$. The universal lower and upper bounds are .02 and .94, respectively. Now $p_2^{C-A-2} p_1^{-(C-B-1)} = p_2^{3}/p_1^{2.5} = .02$ is small, so we anticipate that the asymptotic value

$p_2(A+1)/(C-A-2)$ will yield a good approximation, which it does, namely, .06. Formula (7.5) gives .19.

11. COMMENTS

We conclude with some general comments. First, it appears that whenever either $p_1^{C-B-2} \; p_2^{-(C-A-1)}$ or $p_2^{C-A-2} \; p_1^{-(C-B-1)}$ is very small, then the appropriate asymptotic value as given by Theorem 1 yields a good approximation to $E(\Theta|\text{data})$. However, except when p_2 is very small, formula (7.5) often does as well as the asymptotic value, and has in addition the virtue of being a good approximation when both p_i are large. (Of course, as both p_i tend to 1, the universal lower and upper bounds both tend to $(A+1)/(A+B+2)$, so that the case of large p_i is generally quite easy to deal with.) Although it is difficult to give a completely general rule as to which approximation to use, ordinarily the asymptotic value $1 - p_1(B+1)/(C-B-2)$ will be appropriate whenever $C-B-2 = [I(J-1)+\lambda-2]/2$ is much larger than $C-A-2 = (\lambda_\alpha-2)/2$, as is usually the case. Note in this connection that Example 3 was unusual in so far as C-B-2 was 1.5, while C-A-2 was 3, and it was only in this example that $1-p_1(B+1)/(C-B-2) = .57$ did poorly. In general our attitude is that since the exact formulas derived in this article allow $E(\Theta|\text{data})$ to be calculated with only modest effort, the primary purpose of the approximations is first to provide insight into the nature and behavior of the Bayes estimates, and secondly, to use such insight to aid in the design of the experiment, i.e., the choice of I and J. Such questions are being explored by E. Bangura in his doctoral dissertation at the University of Michigan.

Our second general comment concerns the robustness of Bayesian inference to the choice of λ_α and C_α. Ordinarily one anticipates that small changes in prior parameters such as λ_α and C_α will have only a slight effect upon Bayesian inference, particularly, when there is substantial data. But this is not the case here. For example, taking C_α very small forces $E(\Theta|\text{data})$ to be nearly 0, while the behavior of the Bayes estimates depends sensitively upon whether $\lambda_\alpha \leq 2$ or $\lambda_\alpha > 2$. Indeed, as $p_2 \to 0$ with $\lambda_\alpha \leq 2$, we find an entirely different form of asymptotic behavior for $E(\Theta|\text{data})$ from that given by Theorem 1. We do not include these results here because $\lambda_\alpha \leq 2$ implies that the prior expectation of σ_α^2 is infinite, which is hardly realistic. Nonetheless, the lack of robustness here is important in and of itself as a general warning for those of us who take a Bayesian approach.

Our final comment is that the approximation $E(\Theta|\text{data}) = 1-p_1(B+1)/(C-B-2)$ is closely related to Stein-type estimators [1966], so that our results concerning when this approximation is not appropriate may be of some value even for non-Bayesian approaches. This situation arises, in particular, when MSW is substantially larger than MSB, in which case, from the Bayesian viewpoint presented by Hill [1965, 1967, 1975], the data carry negligible information about μ, the μ_i, and σ_α^2. This is clarified by the Theorem of Hill [1975, p.570], where it is shown that essentially only two possible forms of limiting distribution are possible for extreme data, and that ordinarily as SSW grows large the posterior distribution of the above parameters converges to the prior distribution, whereas when SSB grows large, the stable estimation argument of L. J. Savage can be employed to yield (approximately) the usual least squares estimates.

BIBLIOGRAPHY

Bailey, W. N. (1935). Generalized Hypergeometric Series. Cambridge University Press, London.

Culver, D. H. (1971). A Bayesian Analysis of the Balanced One-way Variance Components Model. University of Michigan, Ph.D. dissertation.

Feller, W. (1950). An Introduction to Probability Theory and its Applications. Wiley, New York.

Hill, B. M. (1965). Inference about variance components in the one-way model. Journal of the American Statistical Association, 60, 806-825.

Hill, B. M. (1967). Correlated errors in the random model. Journal of the American Statistical Association, 62, 1387-1400.

Hill, B. M. (1975). On coherence, inadmissibility and inference about many parameters in the theory of least squares. In S. E. Fienberg and A. Zellner, eds., Studies in Bayesian Econometrics and Statistics in Honor of L. J. Savage, 555-584. North-Holland Publishing Co., Amsterdam.

Lindley, D. and Smith, A. (1972). Bayes estimates for the linear model. Journal of the Royal Statistical Society B, 34, 1-41.

Stein, C. (1966). An approach to the recovery of inter-block information in balanced incomplete block designs. In F. N. David, ed., Festchrift for J. Neyman, 351-366. Wiley, New York.

Whittaker, E. T., and Watson, G. N. (1973). A Course of Modern Analysis, Fourth Edition. Cambridge University Press.

☆ U. S. GOVERNMENT PRINTING OFFICE: 1976 — 657-630/675