A Bayesian Approach to Ranking and Selection of Related Means with Alternatives to AOV Methodology

by

James O. Berger
Purdue University

John Deely
Purdue University and University of Canterbury

Technical Report #86-8

Department of Statistics
Purdue University

July 1987

AUTHORS' FOOTNOTE

James Berger is the Richard M. Brumfield Distinguished Professor of Statistics at the Department of Statistics, Purdue University, West Lafayette, IN 47907. John Deely is Professor of Statistics at the Department of Mathematics, University of Canterbury, Christchurch 1, New Zealand. This research was supported by U.S. National Science Foundation Grant DMS-8401996 and by an Erskine Fellowship at the University of Canterbury. The authors thank K. H. Fong for implementing the numerical algorithms and the referees and associate editor for helpful comments.

KEY WORDS AND PHRASES

Unbalanced Design; Hierarchical Bayes; Exchangeability; Bayes Factor; Monte-Carlo Integration.

A Bayesian Approach to Ranking and Selection of
Related Means with Alternatives to AOV Methodology

Abstract

A set of unknown normal means (treatment effects, say) $\{\theta_1, \theta_2, \ldots, \theta_k\}$ is to be investigated. Two common questions in AOV and ranking and selection are (1) What is the strength of evidence against the hypothesis H_0 of equality of means? (2) If H_0 is false, which mean is the largest (or smallest)? A Bayesian approach to the problem is taken, leading to calculation of the posterior probability of H_0 and the posterior probabilities that each mean is the largest, conditional on H_0 being false. A variety of exchangeable, nonexchangeable, informative, and noninformative prior assumptions are considered. Calculations involve, at worst, only low dimensional numerical integration, in spite of the fact that the dimension k can be arbitrarily large.

As an example, Table 1 in the introduction presents, for each baseball team in the National league in 1984, the highest batting average obtained by any player on the team with at least 150 at bats. The observed batting averages are treated as sample proportions from binomial distributions with parameters θ_i = "true probability of getting a hit for the given player", and it is desired to select the best hitter from the group, namely the player with the largest θ_i.

Calculated, using a Bayesian model of exchangeability for the θ_i, are quantities such as the posterior probabilities that each θ_i is the largest. Such posterior probabilities give very easy to understand and useful measures to assist in selection and ranking. Of substantial interest is that, in unbalanced examples such as the baseball example (the players all had different numbers of "at bats", and hence different variances), it need not be the case that the treatment with the largest sample mean is judged to have the largest true mean. Thus Player 1's observed batting average was higher than that of Player 2, but Player 2 had a substantially smaller variance and was determined (by the hierarchical Bayes method) to have a larger probability of being the best true hitter.

An interesting sidelight to the development is the presentation of a closed form solution for testing $H_0: \theta_1 = \theta_2$ vs. $H_1: \theta_1 < \theta_2$ vs. $H_2: \theta_1 > \theta_2$, when the treatments are judged to be apriori exchangeable.

1. Introduction

Selection and ranking procedures have been developed in modern statistical methodology over the past 30 years with fundamental papers beginning with Bechhofer (1954) and Gupta (1956). A discussion of their respective differences and the various modifications that have taken place since then can be found in many places, e.g., Gibbons et al (1977), Gupta and Panchapakesan (1979), Dudewicz and Koo (1982).

Even though various problems and models amenable to ranking and selection techniques have been discussed in the literature, the methods are less frequently used than is classical AOV and related methodology. Probably due to historical reasons, the AOV model has become the "standard approach". In fact it is often used without much consideration given to the basic problem which most experimenters ultimately face when using AOV; namely, after significant effects are indicated, which combination of the various factors is the most significant, i.e., "best". Ad hoc procedures which answer this question and which are compatible with the AOV model have found great acceptance; hence the development and application of multiple comparisons and simultaneous confidence intervals (cf. Miller (1977)). In fact, the properties of the procedures are not well known and various attempts to justify specific procedures have appeared only recently, e.g., Morely (1982), Keselman et al (1978), Hsu(1982).

A third general approach to these problems has been the Bayesian approach. To see some of the advantages of this approach, consider the specific situation in which a set of unknown means $\{\theta_1, \theta_2, \ldots, \theta_k\}$ is to be investigated. Independent normal experiments are conducted for each mean, yielding sample means $X_i \sim \mathcal{N}(\theta_i, \sigma_i^2)$. The following two questions are often considered in AOV and ranking and selection:

(1) What is the strength of evidence against the hypothesis H_0: $\theta_1 = \theta_2 = \ldots = \theta_k$?

(2) If H_0 is false, which mean is the largest? (Of course, one could similarly deal with the question of which mean is the smallest.)

Because of conditionality concerns with inference after testing, it is difficult for frequentist methods to simultaneously answer both questions, but Bayesian methods encounter no difficulties. In answer to (1), one calculates the posterior probability of H_0 (or Bayes factor against H_0); in answer to (2), one can calculate p_j, the posterior probability that θ_j is the largest mean, for $j = 1, \ldots, k$, conditional on H_0 being false. Other advantages that can accrue from a Bayesian approach are:

(i) The vector $(p_1, \ldots, p_k)$ gives a relatively complete and easily interpretable answer to the selection of largest mean problem;

(ii) The Bayesian measures of evidence concerning H_0 are much easier to interpret; Berger and Sellke (1987) argue that P-values for precise hypotheses such as H_0 can be very misleading if interpreted quantitatively.

(iii) Unbalanced cases (i.e., cases where not all σ_i^2 are equal) can be handled with only slight additional difficulty.

(iv) Relationships among the θ_i, such as an apriori belief in exchangeability of the θ_i, can be incorporated into the analysis.

In elaboration of the importance of this last point, consider the data given in Table 1. Listed, for each baseball team in the National League in 1984, is the highest batting average (x_i) obtained by any player on the team with at least 150 at bats. (For convenience, the x_i have been ordered.) We will treat these x_i as sample proportions from binomial distributions with parameters $\theta_i =$ "true probability of getting a hit for the given player" and $n_i =$ "number of at bats". All n_i are large enough for the usual normal approximation to hold; the σ_i^2 are thus calculated as $x_i(1 - x_i)/n_i$. (Ignore the p_i rows for now.)

Table 1. Observed Batting Averages

i	1	2	3	4	5	6
x_i	.362	.351	.351	.346	.324	.321
n_i	185	606	342	214	262	474
$1000 \times \sigma_i^2$	1.25	0.38	0.67	1.06	0.84	0.46
p_i	.159	.222	.165	.125	.077	.061

i	7	8	9	10	11	12
x_i	.314	.312	.311	.303	.298	.290
n_i	636	600	550	535	181	607
$1000 \times \sigma_i^2$	0.34	0.36	0.39	0.39	1.16	0.34
p_i	.038	.035	.036	.024	.048	.010

It is desired, based on this data, to select the best hitter from the group, namely the player with the largest true θ_i. A rather naive approach to this would be to simply select the player with the largest x_i, in this case Player 1. But consider Player 2; $x_2 = .351$ is close to $x_1 = .362$, and the variance of X_2 is much smaller than that of X_1 (due to the many more at bats of Player 2). Many peoples' intuition would suggest choosing Player 2 over Player 1, because of this difference in variance. The basis of such intuition is perhaps a belief that the θ_i are apriori exchangeable, so that a large x_i associated with a large σ_i^2 is likely to have arisen from a substantially smaller θ_i. To put this another way, large X_i which have large σ_i^2 should be "shrunk" towards a central average. The most effective way of modelling exchangeability and effecting such a shrinkage pattern is through Bayesian analysis. Indeed, without modelling the θ_i, classical methods have a very difficult time with examples such as this. For instance, no matter how different the variances are, classical procedures which select according to the largest sample mean usually satisfy classical optimality criteria. (Several qualifications concerning the above example should be mentioned. We have ignored the fact that each of the x_i is the largest average on their team; through hierarchical modelling of all players, this information could be utilized. Also, the number of at bats is also relevant information — good players usually bat more — and more complicated Bayesian modelling would take this into account. Both of these additional inputs would only cause more "order reversal" of the means, however, and will be ignored for simplicity of exposition.)

The Bayesian analysis we consider is based on modelling exchangeability among the θ_i through use of a hierarchical Bayesian analysis. The option of testing H_0 is made possible by specification of γ, the prior probability of H_0. (Specification of γ is *not* necessary for those who place no credence in such tests; the ranking probabilities turn out to be independent of γ.) Among the many Bayesian papers dealing with AOV or hierarchical priors are Hill (1965), Box and Tiao (1968), Lindley and Smith (1972), and Ghosh and

Meeden (1984). (Further references can be found in Berger (1985).) Goel and Rubin (1977) actually give calculations similar to some of ours for a decision theoretic approach to these questions. Gupta and Yang (1985) consider a Bayesian selection problem but with a non-hierarchical prior.

We actually consider a substantially more general situation than that indicated above. The AOV type hypothesis can be of the form H_0: $\theta = y\,\beta + d$, where y is a matrix of covariates, β is arbitrary and d a vector of constants. (The hypothesis H_0: $\theta_1 = \theta_2 = \ldots = \theta_k$ is of this form with $y = (1, \ldots, 1)^t$, β an unknown constant and d the zero vector.) The probability calculations considered are for quite arbitrary quadrants and a rich class of hierarchical models.

The major impetus for this development was the realization that implementation of the Bayesian procedures requires only low dimensional numerical integration. What would at first sight appear to require at least k dimensional integration, can in most cases be reduced to just 3 dimensional integration. "Balanced" or "exchangeable" cases can be reduced to 2 dimensional integration. The case of $k = 2$ can be reduced to a one dimensional integral, and in special cases can actually be carried out in closed form. The $k = 2$ situation is of considerable interest in its own right, since it will be seen to correspond to testing H_0: $\theta_1 = \theta_2$ versus $H_1 : \theta_1 > \theta_2$ versus H_2: $\theta_1 < \theta_2$ (or generalizations thereof) under a rich class of dependent prior distributions for (θ_1, θ_2).

Section 2 develops the notation used, formally presents the problems to be addressed, and gives the class of prior distributions that will be considered. Important special cases are given for illustrative purposes, including cases of noninformative second stage priors (which yield automatic procedures not requiring subjective input of the user). Section 3 presents the most general results. Section 4 specializes these results to the balanced, exchangeable case, and Section 5 considers the situation for $k = 2$. Section 6 discusses the calculation of the required integrals, and presents some conclusions.

Section 2 - Notation and the Prior Distribution

As mentioned in the introduction, we assume $X_i \sim \mathcal{N}(\theta_i, \sigma_i^2)$, independently for $i = 1, \ldots, k$. The variances σ_i^2 are assumed to be known and to incorporate the sample size effect. (Typically X_i will be the sample mean from n_i observations on population i, and σ_i^2 will be the variance of the sample mean.) Comments about unknown σ_i^2 are given at the end of the paper.

2.1 Basic Selection Problem

To measure the evidence against H_0: $\theta_1 = \theta_2 = \ldots = \theta_k$ we will calculate

$$p_0 = \text{ Posterior probability of } H_0 \, . \tag{2.1}$$

This will, of course, depend on

$$\gamma = \text{ Prior probability of } H_0,$$

which some may find troublesome. It is common, therefore, to consider instead the *Bayes factor* against H_0, namely

$$B = \frac{p_0}{(1-p_0)} \cdot \frac{(1-\gamma)}{\gamma}. \tag{2.2}$$

This is the ratio of posterior odds to prior odds, and can often be interpreted as the "odds for H_0 provided by the data" (see Berger (1985)).

Also of interest, for $i = 1, \ldots, k$, is the posterior probability

$$p_j = \Pr(\theta_j \text{ is the largest mean} \mid \text{data}, H_0 \text{ is false}). \tag{2.3}$$

It is important to note that this will *not* depend on γ. The vector $(p_0, p_1, \ldots, p_k)$ (or the vector $(B, p_1, \ldots, p_k)$) provides a simple yet fairly complete answer to the questions posed in the introduction. Of course, other quantities can also be of interest such as estimates for the θ_i, together with standard errors. For our scenario, such estimates and standard errors are given in Berger (1985), and will not be repeated here.

2.2 General Selection Problem

We will actually formulate the problems of interest somewhat more generally. The AOV type hypothesis being tested will be generalized to (writing $\theta = (\theta_1, \theta_2, \ldots, \theta_k)^t$)

$$H_0\colon\ \theta = y\,\beta + d \tag{2.4}$$

where y is a $(k \times \ell)$ matrix of known covariates, $d = (d_1, \ldots, d_k)^t$ is a fixed vector, and β is unknown. The i^{th} row of y will be denoted $y_i = (y_{i1}, \ldots, y_{i\ell})$, and consists of the covariates corresponding to the i^{th} population (or θ_i). The unknown $\beta = (\beta_1, \ldots, \beta_\ell)^t$ could be constrained, but will usually be unconstrained.

Particularly interesting is the case $\ell = 1$ and β completely unconstrained, for then we can write $y = (y_1, \ldots, y_k)^t$ and (2.4) can be rewritten (assuming all y_i are nonzero)

$$H_0\colon\ \frac{\theta_1 - d_1}{y_1} = \frac{\theta_2 - d_2}{y_2} = \ldots = \frac{\theta_k - d_k}{y_k}. \tag{2.5}$$

Of interest then are the probabilities (which again do not depend on the prior probability of H_0)

$$p_j = \Pr\left(\frac{\theta_j - d_j}{y_j} \geq \frac{\theta_i - d_i}{y_i} \text{ for all } i \;\middle|\; \text{data}, H_0 \text{ is false}\right). \tag{2.6}$$

We will still use p_0 and B to denote the posterior probability of H_0 and the Bayes factor, respectively. Note that (2.5) and (2.6) reduce to (2.1) and (2.3) respectively when all y_i equal 1 and all d_i equal 0.

Flexibility in choices for y and d also permit various contrasts amongst the components of θ to be tested, namely those that can be expressed as

$$\Omega = \{\theta\colon\ a_i\,\theta_i + b_i \leq \theta_j, \quad \text{for all } i \neq j\}.$$

(Setting $b_i = -\infty$ for some i will remove that θ_i from the specification.) Our results do *not* apply to general contrasts such as $\sum_{i=1}^{k} a_i\, \theta_i \le \sum_{i=1}^{k} b_i\, \theta_i$, in the sense that there is not necessarily any reduction in the dimensionality of the integral that must then be calculated.

2.3 The Prior Distribution

A quite general form for the prior distribution is given below. Various special cases are given following the general description, and indicate the flexibility allowed. Note that noninformative choices are given, to allow for automatic Bayesian procedures which do not require specific prior inputs.

The prescription for the prior distribution on $\theta = (\theta_1, \theta_2, \ldots, \theta_k)^t$ will follow the hierarchical approach as given in Berger (1985) (Sections 3.6 and 4.6) and thus consists of two stages: a distribution of θ given 'hyperparameters' (β, σ_π^2) and a distribution for (β, σ_π^2), written $\pi_1(\theta|\beta, \sigma_\pi^2)$ and $\pi_2(\beta, \sigma_\pi^2)$, respectively. Specifically, it will be assumed that $\pi_1(\theta|\beta, \sigma_\pi^2)$ is $\mathcal{N}_k(y\, \beta + d, \sigma_\pi^2 I)$ and that

$$\pi_2(\beta, \sigma_\pi^2) = \pi_{2,1}(\beta) \;\cdot\; \pi_{2,2}(\sigma_\pi^2), \tag{2.7}$$

where $\pi_{2,1}(\beta)$ is $\mathcal{N}_\ell(\beta^0, A)$ (or $\pi_{2,1}(\beta) \equiv 1$, corresponding to $A = \infty$),

$$\pi_{2,2}(\sigma_\pi^2) = \gamma I_{\{0\}}(\sigma_\pi^2) + (1-\gamma)\pi_{2,2}^*(\sigma_\pi^2)\ , \tag{2.8}$$

and $\pi_{2,2}^*(\sigma_\pi^2)$ is arbitrary (though specific noninformative choices will be suggested). The known y and d and unknown β have already been described. The ℓ-vector β^0, the $(\ell \times \ell)$ positive definite matrix A, and γ are all subjectively chosen constants, and $I_{\{0\}}(\sigma_\pi^2)$ denotes the degenerate distribution which gives unit mass to the point $\sigma_\pi^2 = 0$. The motivation and interpretation of this prior can best be seen by looking at special cases.

Case 1. Exchangeable Means

Set $\ell = 1$, $y = (1, \ldots, 1)^t$, and $d = 0$. Then $\pi_1(\theta|\beta, \sigma_\pi^2)$ specifies that, given (β, σ_π^2), the θ_i are i.i.d. $\mathcal{N}(\beta, \sigma_\pi^2)$. The hyperparameters (β, σ_π^2) are given second stage prior π_2; it is easy to see that the marginal prior of θ is then exchangeable. Noninformative and informative choices of π_2 are discussed in Cases 4 and 5.

Case 2. Means Following a Regression Structure

Suppose that, for population i, there is available a vector of known covariates $y_i = (y_{i1}, \ldots, y_{i\ell})$, and that θ_i is assumed to be related to the covariates by the linear model

$$\theta_i = y_i\, \beta + \varepsilon_i,$$

where β is an unknown vector of regression coefficients and the ε_i are i. i. d. $\mathcal{N}(0, \sigma_\pi^2)$. As an example, suppose θ_i is the mean of a process at time t_i, and it is believed that

$$\theta_i = \beta_1 + t_i\beta_2 + \varepsilon_i$$

(so $\ell = 2, y_i = (1, t_i), \beta = (\beta_1, \beta_2)^t$, and $d = 0$).

Case 3. Testing H_0: $\theta = \mathbf{y}\ \beta + \mathbf{d}$

The prior parameter γ in (2.8) can be interpreted as the prior probability that $\theta = y\ \beta + d$ (for some β), since γ is the probability that $\sigma_\pi^2 = 0$ and hence the probability that $\pi_1(\theta|\beta, \sigma_\pi^2)$ is degenerate at $y\ \beta + d$. The most common use for this option will be in allowing for a test of H_0: $\theta_1 = \theta_2 = \ldots = \theta_k$ (corresponding to $\ell = 1$, $y = (1, \ldots, 1)^t$, and $d = 0$), though the generalization in (2.5) may sometimes be of interest.

This option may not be desired. In other words, there may be no real belief that the means could be equal (or at least approximately equal). Our analysis still applies in this case, since the p_i do not depend on γ. If one does want to provide the evidence against this hypothesis, yet is concerned with making a subjective choice of γ, two options are available. The first is to make the "noninformative" choice $\gamma = \frac{1}{2}$. The second, which we shall follow, is to report the Bayes factor, since this does not depend on γ and is a reflection of the evidence provided by the data.

Note that our prior assumption, that either $\theta = y\ \beta + d$ or that the θ_i vary continuously over R^k, is not completely natural from a Bayesian perspective. More natural would be a prior which allowed various subsets of the θ_i to (say) be equal with positive probabilities. While appealing, this would considerably complicate the issue, and would leave us with something far more than a Bayesian analogue of AOV and ranking and selection.

Case 4. Noninformative Second Stage

For either Case 1 or Case 2, one could make a noninformative choice of the second stage prior π_2 in (2.7). The noninformative choice for $\pi_{2,1}(\beta)$ is traditionally

$$\pi_{2,1}\ (\beta) \equiv 1. \tag{2.9}$$

There is less consensus on a noninformative choice for $\pi_{2,2}^*(\sigma_\pi^2)$. Some (e.g. Morris (1983) and Berger (1985)) recommend also

$$\pi_{2,2}^*(\sigma_\pi^2) \equiv 1. \tag{2.10}$$

Another possible choice, which has advantages for small k is

$$\pi_{2,2}^*(\sigma_\pi^2) = \prod_{i=1}^{k} \frac{1}{(\sigma_i^2 + \sigma_\pi^2)^{1/k}}\ . \tag{2.11}$$

When the σ_i^2 equal a common value σ^2, this reduces to

$$\pi_{2,2}^*(\sigma_\pi^2) = \frac{1}{\sigma^2 + \sigma_\pi^2}\ , \tag{2.12}$$

a common noninformative prior for variance component problems. Again, the noninformative choice for γ would traditionally be $\gamma = \frac{1}{2}$.

Some might be leery of noninformative priors, especially since their definition here is clearly somewhat arbitrary. Our view is indeed that subjective proper priors are generally preferable, but that

(i) If there is sufficient data for it to be possible to claim that objective answers are attainable, then any sensible noninformative prior will typically yield such answers;

(ii) The Bayesian answers with any of these noninformative priors will tend to be much more sensible than classical answers.

For testing H_0: $\theta = y\ \beta + d$, improper noninformative priors for σ_π^2 cannot be used. A partly subjective analysis is thus really necessary. To avoid the misleading answers that can result from a classical analysis (see Berger and Sellke (1987)), it is often argued (cf. Jeffreys (1961) and Zellner and Siow (1980)) that in situations where a subjective analysis cannot be performed, analysis with "conventional" proper priors should be undertaken. (Conventional priors are generally selected from a given family using such notions as overall invariance to the scale of the problem.) Conventional proper versions of (2.11) and (2.12) are

$$\pi_{2,2}^*(\sigma_\pi^2) = \left[\prod_{i=1}^{k} \frac{K}{(\sigma_i^2 + \sigma_\pi^2)}\right]^{3/(2k)} \tag{2.13}$$

(K being the normalizing constant) and

$$\pi_{2,2}^*(\sigma_\pi^2) = \frac{\sigma}{2(\sigma^2 + \sigma_\pi^2)^{3/2}} \,. \tag{2.14}$$

We shall use these for hypothesis testing.

<u>Case 5. Informative Second Stage</u>

The general prior allows for the informative choice of a $\mathcal{N}_\ell(\beta^0, A)$ prior as the second stage prior π_2 for β. Thus β^0 can be considered to be a "best guess" for β, and A the covariance matrix for this guess. In the exchangeable Case 1 scenario, β would be the common prior mean for the θ_i, so that β^0 would be a guess for this common mean, with A being the variance of this guess. The assumed normal form for this density is actually needed only in certain special cases (essentially cases involving exchangeability or (2.6) and $\sigma_i^2 = \sigma^2$ for all i). In these specific cases the integral over β can be carried out in closed form; for the general case, any density for β could be used, since numerical integration is required.

For $\pi_{2,2}^*(\sigma_\pi^2)$, any density can be used, since numerical integration is required (except when $k = 2$ and $\sigma_1^2 = \sigma_2^2$). A commonly used class of priors for a variance is the gamma class, any proper member of which could be used for $\pi_{2,2}^*$. For variance components, however, decreasing densities are often plausible; a useful class of such priors, generalizing the noninformative prior in (2.12), is given (for $m > 1$ and $C > 0$) by

$$\pi^*_{2,2}(\sigma^2_\pi) = \frac{(m-1)C}{(1 + C\sigma^2_\pi)^m} \,. \tag{2.15}$$

A simple calculation shows that the c.d.f. corresponding to this prior is

$$F(t) = 1 - (1 + Ct)^{-(m-1)},$$

so that the median, $\rho_{.5}$, is given by

$$\rho_{.5} = (2^{1/(m-1)} - 1)\; C^{-1},$$

and the third quartile, $\rho_{.75}$, is

$$\rho_{.75} = (4^{1/(m-1)} - 1)\; C^{-1}.$$

Thus one could subjectively select $\rho_{.5}$ and $\rho_{.75}$, and use (solving for m and C),

$$C = \frac{\rho_{.75} - 2\rho_{.5}}{(\rho_{.5})^2} \;, m = 1 + \frac{\log 2}{\log\left([\rho_{.75} - \rho_{.5}]/\rho_{.5}\right)} \,.$$

(It is assumed that the elicited quartiles satisfy $\rho_{.5} < \rho_{.75}/2$; if not, a different functional form should be used.)

An interesting special case of this prior, for the situation where $\sigma^2_1 = \sigma^2_2 = \ldots = \sigma^2_k \equiv \sigma^2$, arises from the choice $C = 1/\sigma^2$. Then

$$\pi^*_{2,2}(\sigma^2_\pi) = \frac{(m-1)}{\sigma^2(1 + \sigma^2_\pi/\sigma^2)^m} \,, \tag{2.16}$$

and the noninformative prior in (2.12) is the renormalized limit of these as $m \to 1$. Also, the "conventional" prior in (2.14) is of this form with $m = 3/2$. To determine an element of this class, one need only specify $\rho_{.5}$ (or equivalently the median of the variance ratio σ^2_π/σ^2), and choose

$$m = 1 + \frac{\log 2}{\log(1 + \rho_{.5}/\sigma^2)} \,. \tag{2.17}$$

Section 3 - General Results

3.1 Basic Formulas

We shall here develop formulas for the calculation, with general prior as in Section 2.3, of the conditional posterior probability of any region of the form

$$\Omega = \{\theta\colon\ \theta_j \epsilon I_j, \text{ and } v_i(\theta_j) \le \theta_i \le w_i(\theta_j) \text{ for all } i \ne j\}, \tag{3.1}$$

where j is given, $v_i(\cdot)$ and $w_i(\cdot)$ are arbitrary (possibly infinite) functions, and I_j is a specified interval. The choice $I_j = (-\infty, \infty)$, $v_i(\theta_j) = -\infty$ and

$$w_i(\theta_j) = (y_i/y_j)(\theta_j - d_j) + d_i$$

yields the general selection probability in (2.6). The formula for the posterior probability of (3.1) is

$$\begin{aligned} p^* &= \text{Pr } (\Omega | \text{ data, } H_0 \text{ is false}) \\ &= \int_0^\infty \int_{R^\ell} \int_{I_j} \left\{ \prod_{i \neq j} \left[\Phi\left(\frac{w_i(\theta_j) - u_i}{\sqrt{V_i}}\right) - \Phi\left(\frac{v_i(\theta_j) - u_i}{\sqrt{V_i}}\right) \right] \right\} \\ &\quad \times \pi_j^*(\theta_j)\ \pi_{2,1}^*\ (\beta)\ \pi_{2,2}^*\ (\sigma_\pi^2 | x) d\theta_j\ d\beta\ d\sigma_\pi^2, \end{aligned} \tag{3.2}$$

where Φ is the standard normal c.d.f.;

$$u_i = x_i - \frac{\sigma_i^2}{(\sigma_i^2 + \sigma_\pi^2)} [x_i - (y_i\ \beta + d_i)], \quad V_i = \frac{\sigma_i^2\ \sigma_\pi^2}{(\sigma_i^2 + \sigma_\pi^2)}\ ;$$

$$\pi_j^*(\theta_j) \text{ is a } \mathcal{N}(u_j, V_j) \text{ density ;}$$

$$\pi_{2,1}^*(\beta) \text{ is a } \mathcal{N}_\ell(u^*, V^*) \text{ density, where}$$

$$u^* = \hat{\beta} - [I + A(y^t W y)]^{-1}\ (\hat{\beta} - \beta^0),$$

$$V^* = [A^{-1} + (y^t W y)]^{-1}, \qquad \hat{\beta} = (y^t W y)^{-1} y^t W (x - d),$$

$$\text{and } W = \text{ diag.}\{(\sigma_1^2 + \sigma_\pi^2)^{-1}, \cdots, (\sigma_k^2 + \sigma_\pi^2)^{-1}\}; \text{ and}$$

$$\pi_{2,2}^*(\sigma_\pi^2 | x) = K^{-1} L(\sigma_\pi^2)\ \pi_{2,2}^*(\sigma_\pi^2), \text{ where}$$

$$L(\sigma_\pi^2) = \frac{\exp\left\{-\frac{1}{2}\left[||x - (y\hat{\beta} + d)||_*^2 + ||\hat{\beta} - \beta^0||_{**}^2\right]\right\}}{(\det W)^{-1/2}[\det(y^t W y + A^{-1})]^{\frac{1}{2}}},$$

$$||x - (y\hat{\beta} + d)||_*^2 = [x - (y\hat{\beta} + d)]^t\ W[x - (y\hat{\beta} + d)],$$

$$||\hat{\beta} - \beta^0||_{**}^2 = (\hat{\beta} - \beta^0)^t([y^t W y]^{-1} + A)^{-1}(\hat{\beta} - \beta^0),$$

and $K = \int_0^\infty L(\sigma_\pi^2)\ \pi_{2,2}^*(\sigma_\pi^2) d\sigma_\pi^2$. Note that (3.2) does not involve γ, the prior probability of H_0.

Testing H_0 for the General Prior

The null hypothesis H_0: $\theta = y\beta + d$ has posterior probability

$$\begin{aligned} p_0 &= \text{ Pr } (H_0 | \text{ data}) \\ &= \left[1 + \frac{(1-\gamma)}{\gamma} \cdot \frac{K}{L(0)}\right]^{-1}, \end{aligned} \tag{3.3}$$

and the Bayes factor is

$$B = L(0)/K. \tag{3.4}$$

Note that specification of γ, the prior probability that $\sigma_\pi^2 = 0$ (and hence that H_0 holds), is needed only in the testing situation.

Noninformative Second Stage Prior

For the noninformative second stage prior defined by (2.9) and (2.11), the formulas are as above, with the changes

$$u^* = \hat{\beta} = (y^tWy)^{-1}y^tW(x-d), \quad V^* = (y^tWy)^{-1},$$

$$\pi_{2,2}^*(\sigma_\pi^2|x) = K^{-1}L(\sigma_\pi^2)\prod_{i=1}^{k}\frac{1}{(\sigma_i^2+\sigma_\pi^2)^{1/k}}, \tag{3.5}$$

$$L(\sigma_\pi^2) = \frac{\exp\{-\frac{1}{2}[x-(y\hat{\beta}+d)]^tW[x-(y\hat{\beta}+d)]\}}{(\det W)^{-1}[\det(y^tWy)]^{\frac{1}{2}}},$$

where K is the appropriate normalizing constant.

For testing, it turns out to be impossible to be completely noninformative. One can choose $\gamma = \frac{1}{2}$, although use of the Bayes factor obviates the necessity for choosing γ. And one can use the noninformative (2.9) for β, essentially arguing by sending A in $\pi_{2,1}(\beta)$ to infinity. But a noninformative choice for $\pi_{2,2}(\sigma_\pi^2)$ cannot be made. Use of (2.15), with specified median and third quartile, or some other proper $\pi_{2,2}$, is necessary. An alternative to subjective specification, is to use the "conventional" prior in (2.13) or (2.14).

Dimensionality Reduction

The dimensionality of the integral in (3.2) can be reduced to two dimensions in the special case in which all $\sigma_i^2 = \sigma^2$, $\ell = 1$, and Ω can be written (for given j)

$$\Omega = \{\theta\colon\ a_i \le (\theta_j - cy_i^{-1}\theta_i) \le b_i \qquad \text{for all } i \ne j\},$$

where $c \ne 0$, the $y_i \ne 0$, and $a_i < b_i$ for all i. (Infinite a_i and b_i are allowed.) Then

$$p^* = \int_0^\infty E^Z\left[\prod_{i\ne j}\left\{\left|\Phi\left(\psi - \frac{c^{-1}y_ia_i}{\sqrt{V}}\right) - \Phi\left(\psi - \frac{c^{-1}y_ib_i}{\sqrt{V}}\right)\right|\right\}\right]\pi_{2,2}^*(\sigma_\pi^2|x)d\sigma_\pi^2, \tag{3.6}$$

where Z is a standard normal random variable and

$$\psi = \frac{\sqrt{V}}{\sigma_\pi^2}\left\{(y_j-c)u^* + c^{-1}y_id_j - d_i\right\} + \frac{\sqrt{V}}{\sigma^2}(c^{-1}y_ix_j - x_i)$$
$$+ Z|c^{-1}y_i|\left[1 + \frac{\sigma^2(y_j-c)^2}{(|y|^2+(\sigma^2+\sigma_\pi^2)A^{-1})}\right]^{1/2};$$

here $V = \sigma^2\sigma_\pi^2/(\sigma^2+\sigma_\pi^2)$,

$$u^* = \hat{\beta} - \frac{(\sigma^2+\sigma_\pi^2)}{[(\sigma^2+\sigma_\pi^2)+A|y|^2]}(\hat{\beta}-\beta^0), \text{ and } \hat{\beta} = |y|^{-2}\sum_{i=1}^{p} y_i(x_i-d_i).$$

3.2 Generalized Selection Problem

For the generalized selection problem specified by (2.4) and (2.6), and when $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$, the dimensionality reduction of (3.6) can be applied to yield

$$\begin{aligned} p_j &= \Pr\left(\frac{\theta_j-d_j}{y_j} \geq \frac{\theta_i-d_i}{y_i} \quad \text{for all } i\neq j|\text{ data, } H_0 \text{ is false}\right) \\ &= \int_0^\infty E^Z\ [\prod_{i\neq j}\Phi(\psi_i(Z,\sigma_\pi^2))]\pi_{2,2}^*(\sigma_\pi^2|x)d\sigma_\pi^2, \end{aligned} \tag{3.7}$$

where Z is a standard normal random variable and

$$\psi_i(Z,\sigma_\pi^2) = \frac{\sigma\sigma_\pi}{(\sigma^2+\sigma_\pi^2)^{1/2}}\left\{\frac{1}{\sigma_\pi^2}\left(\frac{y_i}{y_j}d_j - d_i\right) + \frac{1}{\sigma^2}\left(\frac{y_i}{y_j}x_j - x_i\right)\right\} + Z\left|\frac{y_i}{y_j}\right|.$$

Section 4 Exchangeable Case

Consider Case 1 of Subsection 2.3, where $\ell = 1$, $y = (1,\cdots,1)^t$. Assume, in addition, that $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 = \sigma^2$. Then H_0 becomes H_0: $\theta_1 = \theta_2 = \cdots = \theta_k$. Also

$$\begin{aligned} p_j &= \Pr\ (\theta_j \text{ is the largest mean } |\text{ data, } H_0 \text{ being false}) \\ &= \int_0^\infty E^Z\Big[\prod_{i\neq j}\Phi\left(\frac{\sigma_\pi/\sigma}{\sqrt{\sigma^2+\sigma_\pi^2}}(x_j-x_i)\ +\ Z\right)\Big]\ \pi_{2,2}^*\ (\sigma_\pi^2|x)\ d\sigma_\pi^2, \end{aligned}$$

where Φ is the standard normal c.d.f.; Z is $\mathcal{N}(0,1)$; and

$$\pi_{2,2}^*(\sigma_\pi^2|x) = K^{-1}L(\sigma_\pi^2)\pi_{2,2}^*(\sigma_\pi^2), \quad \text{where}$$

$$L(\sigma_\pi^2) = \frac{\exp\left\{-\frac{1}{2}\left[\frac{S^2}{(\sigma^2+\sigma_\pi^2)} + \frac{k(\bar{x}-\beta^0)^2}{(\sigma^2+\sigma_\pi^2+kA)}\right]\right\}}{(\sigma^2+\sigma_\pi^2)^{(k-1)/2}(\sigma^2+\sigma_\pi^2+kA)^{1/2}A^{-1/2}},$$

$S^2 = \sum_{i=1}^{k}(x_i-\bar{x})^2$, and K is the appropriate normalizing constant. The relevant formulas for hypothesis testing are given by (3.3) and (3.4).

Noninformative Second Stage

For the noninformative second stage prior defined by (2.9) and (2.12) the above formulas hold with the change

$$\pi^*_{2,2}(\sigma^2_\pi|x) = K^{-1}L(\sigma^2_\pi)/(\sigma^2+\sigma^2_\pi), \tag{4.1}$$

$$L(\sigma^2_\pi) = \frac{\exp\left\{-S^2/[2(\sigma^2+\sigma^2_\pi)]\right\}}{(\sigma^2+\sigma^2_\pi)^{(k-1)/2}} .$$

If testing is to be done, one would need to either use a subjective choice for $\pi^*_{2,2}(\sigma^2_\pi)$, or use the "conventional" choice in (2.14). The latter would result in use of

$$\pi^*_{2,2}(\sigma^2_\pi|x) = \frac{K^{-1}L(\sigma^2_\pi)\sigma}{2(\sigma^2+\sigma^2_\pi)^{3/2}} \tag{4.2}$$

instead of (4.1).

It is of some interest that, for the proper priors in (2.16) (including the "conventional" choice of $m = 3/2$), the posterior probability of H_0 and the Bayes factor can be calculated in closed form, providing

$$m^* = m + (k-5)/2 \tag{4.3}$$

is an integer. Indeed, then

$$\begin{aligned} K &= \int_0^\infty L(\sigma^2_\pi)\ \pi^*_{2,2}(\sigma^2_\pi)\ d\sigma^2_\pi \\ &= \frac{(m-1)(m^*!)}{\sigma^{(k-1)}v^{(m^*+1)}} \left(1 - e^{-v}\sum_{i=0}^{m^*}\frac{v^i}{i!}\right), \end{aligned} \tag{4.4}$$

where $v = S^2/(2\sigma^2)$. (At $v = 0$, equation (4.4) is to be interpreted as $K = \sigma^{(1-p)}(m-1)/(m^*+1)$.) Formulas (3.3) and (3.4) then provide the posterior probability of H_0 and the Bayes factor.

Section 5 - Two Dimensions

5.1 General Result

For the special case $k = 2$, simplifications over previous formulas can be obtained. Indeed

$$\begin{aligned} p^* &= \Pr(\theta_2 \le c\theta_1 + b \mid \text{data, } H_0 \text{ is false}) \\ &= \int_0^\infty \Phi(U(\sigma^2_\pi))\ \pi^*_{2,2}(\sigma^2_\pi|x)\ d\sigma^2_\pi, \end{aligned} \tag{5.1}$$

where

$$U(\sigma_\pi^2) = \frac{b + \sigma_\pi^2[cx_1W_1 - x_2W_2] + c\sigma_1^2W_1(y_1u^* + d_1) - \sigma_2^2W_2(y_2u^* + d_2)}{[\sigma_\pi^2(c^2\sigma_1^2W_1 + \sigma_2^2W_2) + V^*(c\sigma_1^2W_1y_1 - \sigma_2^2W_2y_2)^2]^{1/2}} \; ;$$

here $W_i = (\sigma_i^2 + \sigma_\pi^2)^{-1}$; Φ is the standard normal c.d.f.;

$$u^* = \hat{\beta} - \frac{1}{1 + A(y^tWy)}(\hat{\beta} - \beta^0), \quad y^tWy = y_1^2W_1 + y_2^2W_2,$$

$$\hat{\beta} = \frac{1}{y^tWy}[y_1W_1(x_1 - d_1) + y_2W_2(x_2 - d_2)],$$

$$V^* = \frac{1}{A^{-1} + (y^tWy)};$$

and $\pi_{2,2}^*(\sigma_\pi^2|x) = K^{-1}L(\sigma_\pi^2)\pi_{2,2}^*(\sigma_\pi^2)$, where

$$L(\sigma_\pi^2) = \frac{\exp\left\{-\frac{1}{2}\left[\sum_{i=1}^2 (x_i - y_i\hat{\beta} - d_i)^2W_i + (\hat{\beta} - \beta^0)^2/(A + [y^tWy]^{-1})\right]\right\}}{(W_1W_2)^{-1/2}[A^{-1} + y^tWy]^{1/2}} \tag{5.2}$$

and K is the appropriate normalizing constant.

5.2 Hypothesis Testing Interpretation

Consider testing

$$H_0\colon \frac{\theta_1 - d_1}{y_1} = \frac{\theta_2 - d_2}{y_2} \text{ versus } H_1\colon \frac{\theta_1 - d_1}{y_1} > \frac{\theta_2 - d_2}{y_2} \text{ versus } H_2\colon \frac{\theta_1 - d_1}{y_1} < \frac{\theta_2 - d_2}{y_2}\,.$$

Then p_0, the posterior probability of H_0, is given by (3.3); p_1^*, the posterior probability of H_1, is given by $p_1^* = (1 - p_0)p^*$, where p^* is given by (5.1) with $c = y_2/y_1$ and $b = d_2 - y_2d_1/y_1$; and p_2^*, the posterior probability of H_2 is $p_2^* = 1 - p_0 - p_1^*$.

Equal Variances

When $\sigma_1^2 = \sigma_2^2$, in addition to the above choices of c and b, p^* simplifies to

$$p^* = \int_0^\infty \Phi\left(\frac{\sigma_\pi\varphi}{(\sigma^2 + \sigma_\pi^2)^{1/2}}\right) \pi_{2,2}^*(\sigma_\pi^2|x)\; d\sigma_\pi^2\;, \tag{5.3}$$

where (letting $|y|^2 = y_1^2 + y_2^2$)

$$\varphi = [y_2(x_1 - d_1) - y_1(x_2 - d_2)]/[\sigma|y|].$$

Noninformative Second Stage Prior for β

If β is given the noninformative prior $\pi_{2,1}(\beta) \equiv 1$ in the equal variance case, then, since $\sum_{i=1}^{2}(x_i - y_i\hat{\beta} - d_i)^2 = \varphi^2\sigma^2$, (5.2) becomes

$$L(\sigma_\pi^2) = (\sigma^2 + \sigma_\pi^2)^{-1/2}\exp\{-\varphi^2\sigma^2/[2(\sigma^2 + \sigma_\pi^2)]\}$$

(ignoring the irrelevant multiplicative constant $|y|$).

Of particular interest for $\pi_{2,2}^*$ is the choice given in (2.16), for m being a half integer, for then p^* can be evaluated in closed form. Indeed, defining $v = \varphi^2/2$, $m^* = m - 3/2$, and assuming that m^* is a nonnegative integer, calculation gives

$$p^* = \frac{\Phi(\varphi) - e^{-v}\sum_{i=0}^{m^*}\left\{\frac{v^i}{i!}\left[\frac{1}{2} + \frac{\varphi c_i}{\sqrt{2\pi}}\right]\right\}}{1 - e^{-v}\sum_{i=0}^{m^*}(v^i/i!)}\,, \tag{5.4}$$

where $c_0 = 1$ and, for $i > 1$,

$$c_i = \frac{1}{(2i+1)}\prod_{k=0}^{(i-1)}\left(1 - \frac{1}{2(i-k)}\right)^{-1}.$$

(This is to be interpreted as $p^* = \frac{1}{2}$ if $\varphi = 0$.) Also, K is given by (4.4), and the formulas for p_0 and B are given by (3.3) and (3.4).

As a specific example, if $m = \frac{3}{2}$ (the "conventional" prior in (2.14)),

$$p^* = \frac{\Phi(\varphi) - e^{-v}\left(\frac{1}{2} + \frac{\varphi}{\sqrt{2\pi}}\right)}{1 - e^{-v}}$$

and

$$B = 2v(e^v - 1)^{-1}.$$

Section 6 - Computations and Conclusions

6.1 The Example

The entries p_i in Table 1 are the posterior probabilities that θ_i is the largest mean, conditional on the data and H_0 being false. Note that these do *not* depend on γ, the prior probability of equality of the means. The exchangeable means hierarchical prior was used, with noninformative second stage priors specified by (2.9) and (2.11) or (2.13) (the results being almost completely insensitive to the choice of (2.11) or (2.13)). Of course, one might well use informative second stage priors for β, based on knowledge of typical baseball averages.

The results bear out the intuitive discussion given in Section 1. Player 1 is *not* judged to have the greatest probability of being the best; indeed, Player 1 is now third. Player 2 has the highest probability of being the best hitter. It is also interesting to consider Player 11. Although he ranked eleventh on the basis of actual batting average, his posterior probability of being the best is now seventh largest. The reason is that his variance is large; his low batting average thus has greater probability of being the largest than those that have small variance, and hence are essentially known to be low. (This conclusion would, however, be affected by the other prior considerations discussed in Section 1.)

Although not of great interest here, the Bayes factor against H_0: $\theta_1 = \theta_2 = \ldots = \theta_k$ was 4.5 (using the second stage prior given by (2.9) and (2.13) — recall that (2.11) will not work here because it is not proper). Thus there is moderate evidence that all players are not identical; since such a hypothesis has no plausibility here, the Bayes factor is only of academic interest.

6.2 The Computation

It should again be emphasized that high dimensional numerical integration is not needed; the integrations over the θ_i were mostly carried out in closed form, resulting in the need for at most $(\ell + 2)$-dimensional numerical integration. A method that we found quite effective for this numerical integration was direct Monte Carlo simulation, based on the hierarchical representation for the posterior. Thus a sequence $\{({}_k\sigma_\pi^2, {}_k\beta, {}_k\theta_j); k = 1, \ldots, N\}$ of independent random vectors are generated; here ${}_k\sigma_\pi^2$ is generated according to $\pi_{2,2}^*(\sigma_\pi^2|x)$, ${}_k\beta$ according to $\pi_{2,1}^*(\beta|{}_k\sigma_\pi^2, x)$, and ${}_k\theta_j$ according to $\pi_j^*(\theta_j|{}_k\beta, {}_k\sigma_\pi^2, x)$. Then

$$I = \int \psi(\theta_j, \beta, \sigma_\pi^2)\pi_j^*(\theta_j)\pi_{2,1}^*(\beta)\pi_{2,2}^*(\sigma_\pi^2)d\theta_j \, d\beta \, d\sigma_\pi^2$$

can be approximated by

$$\hat{I} = \frac{1}{N} \sum_{k=1}^{N} \psi({}_k\theta_j, {}_k\beta, {}_k\sigma_\pi^2).$$

Since π_j^* and $\pi_{2,1}^*$ are normal distributions, generation of the ${}_k\theta_j$ and ${}_k\beta$ pose no problem. The density $\pi_{2,2}^*$ is quite complicated, but its shape is well approximated by a two-point mixture of gammas, and an accept-reject based method of generating the ${}_k\sigma_\pi^2$ is quite efficient. (Note that K, the normalizing constant for $\pi_{2,2}^*$, is determined by a one-dimensional integral which can easily be evaluated.)

Calculation of the p_i in Table 1 was done on a CDC 6000, using $N = 6000$ and requiring about 460 seconds. The standard errors of the p_i in Table 1 ranged from 0.003 for the large p_i to 0.001 for the small p_i. For most applications, standard errors in the 0.01 to 0.005 range would be quite satisfactory. Such could be achieved using $N = 600$ and at 1/10 the time and cost of the above calculation.

6.3 Conclusions and Generalizations

The hierarchical Bayesian approach is a promising method of dealing with ranking and selection problems involving unequal variances. The existence of quite easily computable "objective" versions of the analysis (those with noninformative second stage prior) should make the approach widely usable in practice.

Several generalizations are of obvious interest and are being developed. These include:

i. Unknown Variances: It turns out that the identical model with σ_i^2 unknown adds, not k additional dimensions of integration, but only one additional dimension of integration (plus an additional multiplicative complexity of k). Efficient numerical methods of evaluation are being developed.

ii. More General Models: Similar analyses are also desirable for models such as randomized block models. Preliminary work indicates that, again, this can often be done with addition of only one or two dimensions of integration.

iii. More General Rankings: Suppose one wanted probabilities of the form

$$\Pr(\theta_1 \text{ is the largest and } \theta_2 \text{ is the second largest} \mid \text{data}).$$

Similar methods can be used to calculate such, though each additional ordering adds another dimension of integration.

iv. Means and Variances: Problems such as estimating the largest mean, and providing variances and confidence regions for it can also be handled in this framework.

v. Loss Functions: It is very natural to take a decision-theoretic approach to the problem of selecting the largest mean, with a loss of the form $L(i,\theta) = W(\theta_i,\theta^*)$ for selecting θ_i where θ^* denotes the maximum of the $\{\theta_j\}$. (Typically, W would be an increasing function of $(\theta^* - \theta_j)$.) A Bayesian analysis would proceed by calculating, for each i, the posterior expected loss $E^{\pi(\theta|x)}[W(\theta_i,\theta^*)]$, and choosing that i which yields smallest posterior expected loss.

The calculation of these posterior expected losses is very similar to the calculations in the paper. Indeed, instead of (3.2) say, one obtains

$$E^{\pi(\theta|x)}[W(\theta_i,\theta^*)]$$
$$= \sum_{j=1}^{k} \int_0^\infty \int_{R^\ell} \int_{-\infty}^{\infty} W_i^*(\theta_j)\{ \prod_{\ell \neq i,j} \Phi\left(\frac{\theta_j - \mu_\ell}{\sqrt{V_\ell}}\right)\} \; \pi_j^*(\theta_j)\pi_{2,1}^*(\beta)\pi_{2,2}^*(\sigma_\pi^2|x)d\theta_j d\beta d\sigma_\pi^2,$$

where

$$W_i^*(\theta_j) = \int_{-\infty}^{\theta_j} W(\theta_i,\theta_j)\pi_i^*(\theta_i)d\theta_i.$$

For many common losses, W_i^* can be calculated in closed form. For instance, if $W(\theta_i,\theta^*) = (\theta^* - \theta_i)$, then

$$W_i^*(\theta_j) = (\theta_j - \mu_i)\Phi\left(\frac{\theta_j - \mu_i}{\sqrt{V_i}}\right) + \sqrt{\frac{V_i}{2\pi}} \exp\left\{-\frac{(\theta_j - \mu_i)^2}{2V_i}\right\}.$$

The calculation of the posterior expected losses then involves essentially the same dimensional integration as did the calculation of the p_i. Also, the dimensionality reductions in integration typically hold for the special cases we considered when W is, say, linear or quadratic. (Note that when W is an indicator function, the calculation is of the type considered in (3.2).)

Appendix

From Berger (1985, section 4.6), one has the representation, for the posterior distribution of θ,

$$\pi(\theta|x) = \int \int \pi_1^*(\theta)\pi_{2,1}^*(\beta)\pi_{2,2}(\sigma_\pi^2|x)\ d\beta\ d\sigma_\pi^2\ ,$$

where

$$\pi_1^*(\theta) = \prod_{i=1}^{k} \pi_i^*(\theta_i)$$

and

$$\pi_{2,2}(\sigma_\pi^2|x) = \gamma^* I_{\{0\}}(\sigma_\pi^2) + (1-\gamma^*)\pi_{2,2}^*(\sigma_\pi^2|x),$$

where

$$\gamma^* = \left[1 + \frac{(1-\gamma)}{\gamma}\ \cdot\ \frac{K}{L(0)}\right]^{-1}, \qquad (A1)$$

all other quantities being defined in Subsection 3.1. Since $\sigma_\pi^2 = 0$ corresponds to H_0 being true,

$$\gamma^* = \ \Pr(H_0 \text{ is true } | \text{ data}).$$

All formulas for $p_0 = \ \Pr(H_0|\text{ data})$ and the Bayes factor B thus follow from (A1).

Other probability calculations are conditional on H_0 being false, which means they are done with respect to

$$\pi^*(\theta|x, H_0 \text{ false}) = \int \int \pi_1^*(\theta)\pi_{2,1}^*(\beta)\pi_{2,2}^*(\sigma_\pi^2|x)\ d\beta d\sigma_\pi^2.$$

<u>Proof of (3.2)</u>

$$\begin{aligned} p^* &= \int_\Omega \pi^*(\theta|x, H_0 \text{ false})d\theta \\ &= \int \int \left[\int_\Omega \prod_{i=1}^{k} \pi_i^*(\theta_i)d\theta_i\right] \pi_{2,1}^*(\beta)\pi_{2,2}^*(\sigma_\pi^2|x)\ d\beta\ d\sigma_\pi^2, \end{aligned}$$

and (3.2) follows easily from the definition of $\pi_i^*(\theta_i)$ and the fact that Ω is a product set.

The changes in formulas for the noninformative prior case are also from Berger (1985).

Proof of (3.6)

Assume that $cy_i^{-1} > 0$; the result can similarly be established if $cy_i^{-1} < 0$. Then

$$a_i \le \theta_j - cy_i^{-1}\theta_i \le b_i$$

can be written

$$c^{-1}y_i(\theta_j - b_i) \le \theta_i \le c^{-1}y_i(\theta_j - a_i).$$

This defines $v_i(\theta_j)$ and $w_i(\theta_j)$ for use in (3.2). We will only analyze the w_i term, the analysis for the v_i term being identical.

Writing $\theta_j = u_j + \sqrt{V}\eta$, where η is $\mathcal{N}(0,1)$, algebra shows that

$$\frac{w_i(\theta_j) - u_i}{\sqrt{V}} = c^{-1}y_i(\eta + \frac{\sqrt{V}}{\sigma_\pi^2}(y_j - c)\beta) + \varphi,$$

where

$$\varphi = \frac{\sqrt{V}}{\sigma_\pi^2}(c^{-1}y_i d_j - d_i) + \frac{\sqrt{V}}{\sigma^2}(c^{-1}y_i x_j - x_i).$$

Since $\beta \sim \mathcal{N}(u^*, V^*)$ independent of η,

$$\eta + \frac{\sqrt{V}}{\sigma_\pi^2}(y_j - c)\beta \sim \mathcal{N}\left(\frac{\sqrt{V}}{\sigma_\pi^2}(y_j - c)u^*, 1 + \frac{V(y_j - c)^2 V^*}{\sigma_\pi^2}\right).$$

The conclusion follows from further algebra.

Proof of (3.7)

This falls within the framework of (3.6), with $c = y_j$, and either $a_i = d_j - d_i y_j / y_i$ and $b_i = \infty$ (if $y_j/y_i > 0$) or $b_i = d_j - d_i y_j / y_i$ and $a_i = -\infty$ (if $y_j/y_i < 0$).

Proof of (4.4)

Upon transformation, it is easy to recognize K as an incomplete gamma function, for which (4.4) is one representation.

Proof of (5.1)

In a fashion similar to the proof of (3.6), one can show that

$$p^* = \int_0^\infty \left[E^Z\ \Phi(B + \sqrt{c}Z)\right] \pi_{2,2}^*(\sigma_\pi^2|x)\,d\sigma_\pi^2,$$

where

$$B = V_2^{-1/2}\Big\{(c\sigma_1^2 W_1 y_1 - \sigma_2^2 W_2 y_2)u^* + b + \sigma_\pi^2(cW_1x_1 - W_2x_2) + (c\sigma_1^2 W_1 d_1 - \sigma_2^2 W_2 d_2)\Big\},$$

and

$$C = c^2 \frac{V_1}{V_2} + \frac{(cV_1 y_1 - V_2 y_2)^2 V^*}{\sigma_\pi^4 V_2}.$$

The conclusion follows from the fact that

$$E^Z \Phi(B + \sqrt{c} Z) = \Phi(B/\sqrt{1+c}).$$

<u>**Proof of (5.4)**</u>

For the given situation,

$$p^* = \int_0^\infty \Phi\left(\frac{\sigma_\pi}{\sqrt{\sigma^2 + \sigma_\pi^2}} \cdot \varphi\right) \cdot K^{-1} \frac{(m-1) \exp\left\{-\varphi^2 \sigma^2 / [2(\sigma^2 + \sigma_\pi^2)]\right\}}{(\sigma^2 + \sigma_\pi^2)^{1/2} \sigma^2 (1 + \sigma_\pi^2/\sigma^2)^m} \, d\sigma_\pi^2,$$

where K is given by (4.4). Defining $w = \sigma_\pi/\sigma$ and changing variables from σ_π^2 to w yields

$$p^* = \frac{2(m-1)}{K\sigma} \int_0^\infty \Phi\left(\left[1 + w^{-2}\right]^{-1/2} \varphi\right) \frac{w \exp\left\{-\varphi^2 / \left[2(1+w^2)\right]\right\}}{(1+w^2)^{m^*+2}} \, dw. \qquad (A2)$$

Note that

$$\int \frac{w \exp\left\{-\varphi^2/[2(1+w^2)]\right\}}{(1+w^2)^{m^*+2}} dw = \sum_{i=0}^{m^*} \frac{K_i}{(1+w^2)^i} \exp\left\{-\varphi^2/[2(1+w^2)]\right\},$$

where

$$K_i = (m^*!) 2^{(m^*-i)} / \left[i! (\varphi^2)^{(m^*+1-i)}\right].$$

Thus, integration by parts in (A2) yields

$$p^* = \frac{2(m-1)}{K\sigma} \left\{ K_0 \Phi(\varphi) - \frac{1}{2} \sum_{i=0}^{m^*} K_i \exp\left\{-\varphi^2/2\right\} \right.$$
$$\left. - \frac{\varphi}{\sqrt{2\pi}} \sum_{i=0}^{m^*} K_i \int_0^\infty (1+w^2)^{-(i+3/2)} \exp\left\{ -\frac{1}{2} \left[\frac{\varphi^2}{(1+w^{-2})} + \frac{\varphi^2}{(1+w^2)}\right] \right\} dw \right\}.$$

Since $(1+w^{-2})^{-1} + (1+w^2)^{-1} = 1$, the integral in the above expression reduces to

$$\int_0^\infty (1+w^2)^{-(i+3/2)} \exp\left\{-\frac{1}{2}\varphi^2\right\} dw = \exp\left\{-\frac{1}{2}\varphi^2\right\} c_i.$$

Collecting terms and simplifying yields the desired conclusions.

REFERENCES

Bechhofer, R. E. (1974), "A Single-Sample Multiple Decision Procedure for Ranking Means of Normal Populations with Known Variances", *Annals of Mathematical Statistics*, 25, 16–39.

Berger, J. (1985), *Statistical Decision Theory and Bayesian Analysis*, New York: Springer-Verlag.

Berger, J., and Sellke, T. (1987), "Testing a Point Null Hypothesis: the Irreconcilability of Significance Levels and Evidence (with Discussion)", *Journal of the American Statistical Association*, 82, 112–122.

Box, G., and Tiao, G. (1968), "Bayesian Estimation of Means for the Random Effects Model", *Journal of the American Statistical Association*, 63, 174–181.

Dudewicz, E. J., and Koo, J. O. (1982), *The Complete Categorized Guide to Statistical Selection and Ranking Procedures*", Columbus: American Science Press.

Ghosh, M., and Meeden, G. (1984), "A New Bayesian Analysis of a Random Effects Model", *Journal of the Royal Statistical Society*, Ser. B, 46, 474–482.

Gibbons, J. D., Olkin, I., and Sobel, M. (1977), *Selecting and Ordering Populations*, New York: Wiley.

Goel, P., and Rubin, H. (1977), "On Selecting a Subset Containing the Best Population - a Bayesian Approach", *Annals of Statistics*, 5, 969–983.

Gupta, S. S. (1956), "On a Decision Rule for a Problem in Ranking Means", Ph.D. Thesis, Institute of Statistics, University of North Carolina, Chapel Hill.

Gupta, S. S., and Panchapakesan, S. (1979), *Multiple Decision Procedures*, New York: Wiley.

Gupta, S. S., and Yang, H. M. (1985), "Bayes-P^* Subset Selection Prodedures for the Best Population", *Journal of Statistical Planning and Inference*, 12, 213–233.

Hill, B. (1965), "Inference about Variance Components in the One-Way Model", *Journal of the American Statistical Association*, 60, 806–825.

Hsu, J. (1982), "Simultaneous Inference with Respect to the Best Treatment in Block Designs", *Journal of the American Statistical Association*, 77, 461–467.

Jeffreys, H. (1961), *Theory of Probability*, London: Oxford University Press.

Keselman, H. J., and Rogan, J. C. (1978), "A Comparison of the Modified Tukey and Scheffé Methods for Multiple Comparisons for Pairwise Contrasts," *Journal of the American Statistical Association* 73, 47–52.

Lindley, D., and Smith, A. F. M. (1972), "Bayes Estimates for the Linear Model", *Journal of the Royal Statistical Society*, Ser. B, 34, 1–41.

Miller, R. G., Jr. (1977), "Developments in Multiple Comparisons", *Journal of the Amer-*

ican Statistical Association, 72, 779–788.

Morley, C. L. (1982), "A Simulation Study of the Powers of Three Multiple Comparison Statistics", *Australian Journal of Statistics*, 24, 201–210.

Morris, C. (1983), "Parametric Empirical Bayes Inference: Theory and Applications", *Journal of the American Statistical Association*, **78**, 47–65.

Zellner, A. and Siow, A. (1980), "Posterior Odds Ratios for Selected Regression Hypotheses", in *Bayesian Statistics*, eds. J. Bernardo, M. DeGroot, D. Lindley, and A. Smith, Valencia, Spain: University Press, pp. 585–604.