Ann. Inst. Statist. Math.
Vol. 40, No. 4, 747-767 (1988)

APPROXIMATED BAYES AND EMPIRICAL BAYES CONFIDENCE INTERVALS—THE KNOWN VARIANCE CASE*

A. J. VAN DER MERWE, P. C. N. GROENEWALD AND C. A. VAN DER MERWE

Department of Mathematical Statistics, University of the Orange Free State, PO Box 339, Bloemfontein, Republic of South Africa

(Received June 11, 1986; revised September 29, 1987)

Abstract. In this paper hierarchical Bayes and empirical Bayes results are used to obtain confidence intervals of the population means in the case of real problems. This is achieved by approximating the posterior distribution with a Pearson distribution. In the first example hierarchical Bayes confidence intervals for the Efron and Morris (1975, *J. Amer. Statist. Assoc.*, **70**, 311–319) baseball data are obtained. The same methods are used in the second example to obtain confidence intervals of treatment effects as well as the difference between treatment effects in an analysis of variance experiment. In the third example hierarchical Bayes intervals of treatment effects are obtained and compared with normal approximations in the unequal variance case.

Key words and phrases: Hierarchical Bayes, empirical Bayes estimation, Stein estimator, multivariate normal mean, Pearson curves, confidence intervals, posterior distribution, unequal variance case, normal approximations.

1. Introduction

In the Bayesian approach to inference, a posterior distribution of unknown parameters is produced as the normalized product of the likelihood and a prior distribution. Inferences about the unknown parameters are then based on the entire posterior distribution resulting from the one specific data set which has actually occurred. In most hierarchical and empirical Bayes cases these posterior distributions are difficult to derive and cannot be obtained in closed form. Numerical integration, normal approximations (Morris (1977, 1983*b*) and Berger (1985)) or other asymptotic approximations of the posterior distribution are then used to obtain

*Financially supported by the CSIR and the University of the Orange Free State, Central Research Fund.

confidence intervals of the unknown parameter values.

In some special cases, however, it is possible to obtain the exact moments of the posterior distribution. By using these moments and Pearson curves good approximations of the posterior distributions and the corresponding confidence intervals can be obtained. In Section 2 the Pearson curve approximation is applied to the Efron and Morris (1975) baseball data. The same methods are used in the second example to obtain confidence intervals between treatment effects in an analysis of variance experiment. In Section 3 exact moments of the posterior distribution for the unequal variance case are obtained. By using these moments and Pearson curves confidence intervals for the effects of high school coaching programs are calculated and compared with normal approximations.

2. The use of Pearson curves to obtain confidence intervals—the equal variance case

Consider the problem of estimating and obtaining confidence intervals of the true means $\{\theta_j\}$ of k normal populations having observed the independent sample means $X_1, X_2, \ldots, X_k$. Each X_j is assumed to have the same variance V, which is known. Thus given θ_j,

$$X_j | \theta_j \overset{\text{ind}}{\sim} N(\theta_j, V) \quad j = 1, \ldots, k . \tag{2.1}$$

Suppose further that

$$\theta_j \overset{\text{ind}}{\sim} N(\mu, A) \quad j = 1, \ldots, k . \tag{2.2}$$

The hierarchical Bayes approach assumes second stage priors on the hyperparameters. Therefore suppose that A is uniformly distributed on $[0, \infty)$ and that the distribution of μ is a Lebesgue (uniform) measure on $(-\infty, \infty)$, then it is well-known (Morris (1977)) that the posterior distribution of $\theta_j | \underline{X}$ is given by

$$f(\theta_j | \underline{X}) = \int_0^1 f(\theta_j | \underline{X}, B) f(B | S) dB , \tag{2.3}$$

where

$$\theta_j | \underline{X}, B \sim N\left(\overline{X} + (1 - B)(X_j - \overline{X}),\ V(1 - B) + \frac{1}{k} BV \right) , \tag{2.4}$$

$$f(B | S) = K^{-1} B^{m-1} \exp(-BS/2) \quad 0 \le B \le 1 , \tag{2.5}$$

is the posterior density of B, where

$$K = \int_0^1 B^{m-1} \exp(-BS/2)dB, \tag{2.6}$$

$$B = \frac{V}{V+A}, \tag{2.7}$$

$$\overline{X} = \frac{1}{k}\sum_{j=1}^{k} X_j, \tag{2.8}$$

$$S = \frac{1}{V}\sum_{j=1}^{k} (X_j - \overline{X})^2, \tag{2.9}$$

and

$$\underline{X}' = (X_1, \ldots, X_k). \tag{2.10}$$

Since it is not possible to obtain the exact distribution of $\theta_j|\underline{X}$, $f(\theta_j|\underline{X})$ will be approximated by a Pearson density. For the use of Pearson curves, the mean, variance, third and fourth central moments of the posterior distribution are needed.

THEOREM 2.1. *The mean, variance, third and fourth central moments of* $\theta_j|\underline{X}$ *are given by*

$$E(\theta_j|\underline{X}) = \hat{\theta}_j = \overline{X} + (1-\hat{B})(X_j - \overline{X}), \tag{2.11}$$

$$\text{Var}(\theta_j|\underline{X}) = \sigma_j^{*2}(\underline{X}) = V\left\{\frac{1}{k} + \frac{k-1}{k}(1-\hat{B})\right\} + v(X_j - \overline{X})^2, \tag{2.12}$$

$$\begin{aligned}\mu_3 = {} & 3V\left(1-\frac{1}{k}\right)\left\{\frac{2(m+1)}{S}\hat{B} - \frac{2m}{Se_m(S)} - \hat{B}^2\right\}(X_j - \overline{X}) \\ & + 2\left\{\frac{m}{Se_m(S)} + \frac{2m(m+2)}{S^2 e_m(S)} - \frac{2(m+1)(m+2)}{S^2}\hat{B}\right. \\ & \left. - \frac{3m}{Se_m(S)}\hat{B} + \frac{3(m+1)}{S}\hat{B}^2 - \hat{B}^3\right\}(X_j - \overline{X})^3,\end{aligned} \tag{2.13}$$

and

$$\begin{aligned}\mu_4 = {} & 3V^2\left\{1 - 2\hat{B}\left(1-\frac{1}{k}\right) - \frac{2m}{Se_m(S)}\left(1-\frac{1}{k}\right)^2\right. \\ & \left. + \frac{2(m+1)}{S}\hat{B}\left(1-\frac{1}{k}\right)^2\right\}\end{aligned} \tag{2.14}$$

$$+6V\left\{\frac{4m(m+2)}{S^2e_m(S)}\left(1-\frac{1}{k}\right)-\frac{4(m+1)(m+2)}{S^2}\hat{B}\left(1-\frac{1}{k}\right)\right.$$

$$-\frac{4m}{Se_m(S)}\hat{B}\left(1-\frac{1}{k}\right)+\frac{4(m+1)}{S}\hat{B}^2\left(1-\frac{1}{k}\right)$$

$$\left.-\hat{B}^3\left(1-\frac{1}{k}\right)+\frac{2(m+1)}{S}\hat{B}-\frac{2m}{kSe_m(S)}-\hat{B}^2\right\}(X_j-\overline{X})^2$$

$$+\left\{-\frac{2m}{Se_m(S)}-\frac{4m(m+3)}{S^2e_m(S)}-\frac{8m(m+2)(m+3)}{S^3e_m(S)}\right.$$

$$+\frac{8(m+1)(m+2)(m+3)}{S^3}\hat{B}+\frac{8m}{Se_m(S)}\hat{B}+\frac{16m(m+2)}{S^2e_m(S)}\hat{B}$$

$$-\frac{16(m+1)(m+2)}{S^2}\hat{B}^2-\frac{12m}{Se_m(S)}\hat{B}^2+\frac{12(m+1)}{S}\hat{B}^3$$

$$\left.-3\hat{B}^4\right\}(X_j-\overline{X})^4,$$

where

$$m=\frac{k-3}{2}, \tag{2.15}$$

$$e_m(S)=m\exp(S/2)\int_0^1 B^{m-1}\exp(-BS/2)dB, \tag{2.16}$$

$$\hat{B}=E(B\mid S)=\frac{k-3}{S}\left(1-\frac{1}{e_m(S)}\right), \tag{2.17}$$

$$v=\operatorname{Var}(B\mid S)=2\{\hat{B}-(1-\hat{B})m/e_m(S)\}/S. \tag{2.18}$$

PROOF. By making use of equation (2.5) and integration by parts, the first four moments of $B\mid S$ are obtained as

$$E(B\mid S)=\hat{B}, \tag{2.19}$$

$$E(B^2\mid S)=-\frac{2}{S}\frac{m}{e_m(S)}+\frac{2(m+1)}{S}\hat{B}, \tag{2.20}$$

$$E(B^3\mid S)=\frac{-2m}{Se_m(S)}-\frac{4m(m+2)}{S^2e_m(S)}+\frac{4(m+1)(m+2)}{S^2}\hat{B}, \tag{2.21}$$

and

$$(2.22) \qquad E(B^4|S) = \frac{-2m}{Se_m(S)} - \frac{4m(m+3)}{S^2 e_m(S)} - \frac{8m(m+2)(m+3)}{S^3 e_m(S)} + \frac{8(m+1)(m+2)(m+3)}{S^3}\hat{B}.$$

Calculating $E(\theta_j|\underline{X}, A)$, $E(\theta_j^2|\underline{X}, A)$, $E(\theta_j^3|\underline{X}, A)$, $E(\theta_j^4|\underline{X}, A)$ and using (2.19)–(2.22), $E(\theta_j|\underline{X})$, $E(\theta_j^2|\underline{X})$, $E(\theta_j^3|\underline{X})$ and $E(\theta_j^4|\underline{X})$ can be obtained.

By using the relationships between moments about the origin and central moments and after some algebraic manipulations (2.11)–(2.14) follow.

In our first example it will be shown that by using Pearson curves good approximations of the true Bayesian intervals can be obtained.

Example 2.1 *Efron and Morris* (1975) *baseball data*

The observations X_j in the second column of Table 1 are the batting averages (after transformations) of $k = 18$ batters in 1970 after 45 attempts. The variance of each X_j is known to be $V = (0.0659)^2$. The batting averages for these players during the remainder of the season, considered to be the true values θ_j, are presented in the last column of Table 2. The estimator $\hat{\theta}_j$ as well as normal approximations of the confidence intervals was calculated by Morris (1977, 1983*b*). Not only was the estimator $\underline{\hat{\theta}}' = (\hat{\theta}_1, \ldots, \hat{\theta}_{18})$ about three times as efficient as the sample mean vector $\underline{X}' = (X_1, \ldots, X_{18})$ but the intervals were 37% shorter while containing the true values with greater frequency than nominally claimed.

It will now be shown that by using Pearson curves good approximations of the exact confidence intervals can be obtained. For the use of Pearson curves;

$$\mu_3, \quad \mu_4, \quad \beta_1 = \frac{\mu_3^2}{\mu_2^3}, \quad \beta_2 = \frac{\mu_4}{\mu_2^2} \quad \text{and} \quad \kappa = \frac{\beta_1(\beta_2+3)^2}{4(4\beta_2 - 3\beta_1)(2\beta_2 - 3\beta_1 - 6)},$$

were calculated and are given in Table 1. The value of κ determines the type of Pearson curve that should be used. It is also known that $m = 7.5$, $\overline{X} = 0.265667$, $S = 18.932467$, $e_m(S) = 6.77428$ and $\hat{B} = 0.675334$.

From Table 1 it is also clear that the posterior distributions for the first eight baseball players are positively skew and those for the remaining players are negatively skew. The posterior distributions for batters 7, 8, 9 and 10 are more or less symmetric. The kurtosis β_2 for all eighteen players are greater than three. The deviations of the skewness and kurtosis values from 0 and 3, respectively, are indications that improvements in confidence intervals can be obtained by using Pearson curves instead of normal

Table 1. The maximum likelihood estimates (MLE), hierarchical Bayes estimates (HBE), $\sigma_j^{*2}(\underline{X})$, μ_3, μ_4, β_1, β_2 and κ as well as the type of Pearson curve.

j	MLE X_j	HBE $\hat{\theta}_j$	$\sigma_j^{*2}(X)$	μ_3	μ_4	β_1	β_2	κ	Type of Pearson curve
1	0.395	0.307657	0.002131	5.6076×10^{-5}	1.5650×10^{-5}	0.324943	3.446301	-3.204780	I
2	0.375	0.301163	0.001972	4.6697×10^{-5}	1.3641×10^{-5}	0.284353	3.507870	1.404358	VI
3	0.355	0.294670	0.001839	3.7654×10^{-5}	1.2048×10^{-5}	0.227977	3.562670	0.409897	VI
4	0.334	0.287852	0.001729	2.8503×10^{-5}	1.0793×10^{-5}	0.157187	3.610586	0.163976	IV
5	0.313	0.281034	0.001648	1.9587×10^{-5}	9.9043×10^{-6}	0.085718	3.646796	0.063745	IV
6	0.313	0.281034	0.001648	1.9587×10^{-5}	9.9043×10^{-6}	0.085718	3.646796	0.063745	IV
7	0.291	0.273892	0.001594	1.0421×10^{-5}	9.3273×10^{-6}	0.026815	3.670970	0.016194	IV
8	0.269	0.266749	0.001573	1.3682×10^{-6}	9.1068×10^{-6}	0.000481	3.680553	0.000268	IV
9	0.247	0.259606	0.001585	-7.6769×10^{-6}	9.2334×10^{-6}	0.014801	3.675425	0.008611	IV
10	0.247	0.259606	0.001585	-7.6769×10^{-6}	9.2334×10^{-6}	0.014801	3.675425	0.008611	IV
11	0.224	0.252139	0.001631	-1.7210×10^{-5}	9.7212×10^{-6}	0.068270	3.654400	0.047498	IV
12	0.224	0.252139	0.001631	-1.7210×10^{-5}	9.7212×10^{-6}	0.068270	3.654400	0.047498	IV
13	0.224	0.252139	0.001631	-1.7210×10^{-5}	9.7212×10^{-6}	0.068270	3.654400	0.047498	IV
14	0.224	0.252139	0.001631	-1.7210×10^{-5}	9.7212×10^{-6}	0.068270	3.654400	0.047498	IV
15	0.224	0.252139	0.001631	-1.7210×10^{-5}	9.7212×10^{-6}	0.068270	3.654400	0.047498	IV
16	0.200	0.244347	0.001717	-2.7357×10^{-5}	1.0773×10^{-5}	0.147855	3.654232	0.146416	IV
17	0.175	0.236230	0.001847	-3.8246×10^{-5}	1.2142×10^{-5}	0.232155	3.559273	3.559273	IV
18	0.148	0.227464	0.002035	-5.0557×10^{-5}	1.4423×10^{-5}	0.303303	3.482905	3.482905	VI

Table 2. The 90% Pearson curve intervals (PCIS) and corresponding exact intervals (EXIS) for the eighteen baseball players.

Player	90% PCIS	90% EXIS	θ_j
1	0.240~0.390	0.235~0.382	0.346
2	0.235~0.380	0.238~0.383	0.300
3	0.230~0.370	0.227~0.365	0.279
4	0.222~0.360	0.222~0.357	0.223
5	0.218~0.350	0.215~0.348	0.276
6	0.218~0.350	0.215~0.348	0.273
7	0.210~0.340	0.209~0.340	0.266
8	0.202~0.332	0.201~0.332	0.211
9	0.194~0.323	0.194~0.324	0.271
10	0.194~0.323	0.194~0.324	0.232
11	0.184~0.315	0.185~0.317	0.266
12	0.184~0.315	0.185~0.317	0.258
13	0.184~0.315	0.185~0.317	0.306
14	0.184~0.315	0.185~0.317	0.267
15	0.184~0.315	0.185~0.317	0.228
16	0.173~0.308	0.174~0.309	0.288
17	0.161~0.301	0.160~0.300	0.318
18	0.147~0.294	0.150~0.295	0.200

approximations.

For details of how to determine the parameters of a Pearson curve, given the values of its moments, see for example Elderton (1953) or Elderton and Johnson (1969). By using a computer program which makes use of four point Lagrangean interpolation in tables of significance points of Pearson curves (see Johnson *et al.* (1963)), the 90% approximated confidence intervals for the eighteen baseball players were calculated and are given in column two of Table 2. The exact intervals were also obtained, using numerical integration and are given in column three.

A comparison of columns two and three shows that the Pearson curve intervals approximate the exact intervals quite well. In only one of the cases (player 17) the true parameter value falls outside the range of the confidence intervals.

If B is estimated by $\hat{B}_S = (k-3)/S$ instead of $\hat{B}$ as defined in equation (2.17), then equation (2.11) becomes the Stein (empirical Bayes) estimator which means that B in (2.5) goes from 0 to ∞ and the risk estimates, posterior distributions, and all other quantities computed for (2.11) can be computed for the Stein estimator. These results are obtained by replacing $e_m(S)$ by infinity (implying $1/e_m(S) = 0$) in all formulas.

Example 2.2 *Ten reaction time experiments* (*Efron* (1975))

The second example has to do with an analysis of variance experiment conducted by Dr. R. Angel of the Stanford Medical School. Each of ten subjects was asked to perform a certain task under seven different conditions. Let y_{jl} indicate the (natural) log reaction time of subject j under condition l; $j = 1, 2, \ldots, 10$; $l = 1, \ldots, 7$.

A two way analysis of variance model

$$y_{jl} = \tilde{\mu} + \theta_j + \delta_l + \varepsilon_{jl} , \tag{2.23}$$

where

$$\sum_{j=1}^{10} \theta_j = \sum_{l=1}^{7} \delta_l = 0 ,$$

was used to analyze the data. Here $\tilde{\mu}$ is the overall mean, θ_j the main effect for subject j, δ_l the main effect for condition l, and ε_{jl} the random experimental error, which is assumed to be independently normally distributed with mean zero and variance σ^2.

The idea is to use hierarchical Bayes techniques to estimate the patient main effects θ_j from the data and to obtain confidence intervals for these main effects. These estimates and confidence intervals will also be compared with the usual ones.

An unbiased estimator of θ_j is

$$X_j = \sum_{l=1}^{7} y_{jl} - \sum_{j=1}^{10} \sum_{l=1}^{7} y_{jl}/70 , \tag{2.24}$$

where $X_j \sim N(\theta_j, V), j = 1, \ldots, 10$, and

$$V = (9/70)\sigma^2 . \tag{2.25}$$

Suppose further that

$$\theta_j \sim N(0, A) \quad j = 1, \ldots, 10 . \tag{2.26}$$

The reason for taking $\mu = 0$ is that the θ_j like the X_j must sum to zero. By assuming that A is uniformly distributed on $[0, \infty)$, the posterior mean as well as the moments around the mean are given by equations (2.11)–(2.14) with $\overline{X} = 0$.

For practical and comparison purposes, V can be taken as known because its estimate is based on a chi-square random variable with 54 degrees of freedom.

This whole experiment was repeated 10 times, yielding a total of 700 observations—10 subjects, 7 conditions, 10 experiments. These repetitions can be used as a check on how a given estimator performed in any given experiment analyzed separately from the others. In Table 3 the upper

Table 3. Ten reaction-time experiments. Upper number is unbiased estimate—lower number is hierarchical Bayes estimate on that column's data.

Subject	Experiment 1	2	3	4	5	6	7	8	9	10	θ_j	$\sum_{i=1}^{10}(X_{ij}-\theta_j)^2 \,/\, \sum_{i=1}^{10}(\hat{\theta}_{ij}-\theta_j)^2$
1	−0.03	0.06	−0.08	−0.04	−0.05	0.15	0.05	0.03	−0.10	−0.06	−0.00	0.0545
	−0.01	0.02	−0.05	−0.02	−0.03	0.13	0.04	0.01	−0.07	−0.03		0.0287
2	0.08	0.24	0.17	0.14	−0.06	0.37	0.20	0.11	0.26	−0.07	0.14	0.1704
	0.04	0.09	0.10	0.09	−0.04	0.32	0.14	0.05	0.18	−0.03		0.1200
3	−0.16	0.07	−0.32	−0.30	−0.28	−0.18	−0.19	−0.25	−0.08	−0.09	−0.18	0.1300
	−0.07	0.03	−0.19	−0.18	−0.19	−0.15	−0.13	−0.12	−0.06	−0.04		0.0974
4	−0.04	0.02	0.02	0.04	−0.05	−0.05	−0.16	−0.02	0.13	−0.12	−0.03	0.0615
	−0.02	0.01	0.01	0.02	−0.03	−0.04	−0.11	−0.01	0.09	−0.05		0.0275
5	0.08	0.01	0.03	−0.00	0.14	0.06	−0.22	−0.00	−0.23	−0.10	−0.02	0.1307
	0.04	0.00	0.02	0.00	0.10	0.05	−0.15	−0.00	−0.16	−0.04		0.0626
6	0.25	0.15	0.17	0.35	0.21	0.24	0.29	0.02	0.28	0.22	0.22	0.0718
	0.12	0.05	0.10	0.21	0.14	0.21	0.20	0.01	0.19	0.10		0.1197
7	0.09	−0.10	−0.06	0.09	0.30	0.09	0.09	0.11	0.07	0.29	0.10	0.1431
	0.04	−0.04	−0.04	0.06	0.20	0.08	0.06	0.05	0.05	0.13		0.0623
8	−0.13	−0.14	−0.16	−0.23	−0.23	−0.34	−0.18	−0.08	−0.15	−0.10	−0.18	0.0524
	−0.06	−0.05	−0.09	−0.14	−0.16	−0.29	−0.13	−0.04	−0.10	−0.04		0.1016
9	−0.09	−0.13	0.17	−0.03	0.04	−0.20	0.03	−0.08	−0.17	−0.05	−0.04	0.1183
	−0.04	−0.05	0.10	−0.02	0.03	−0.17	0.02	−0.04	−0.12	0.02		0.0555
10	−0.04	−0.16	0.07	−0.02	−0.03	−0.14	0.09	0.17	−0.00	−0.02	−0.01	0.0898
	−0.02	−0.06	0.04	−0.01	−0.02	−0.12	0.06	0.08	−0.00	−0.01		0.0304
V	0.011	0.020	0.012	0.014	0.011	0.007	0.010	0.010	0.011	0.016	0.00122	1.025
$1-\hat{B}$	0.459	0.361	0.587	0.611	0.682	0.860	0.701	0.482	0.692	0.443		0.7057

number in each box is X_j the maximum likelihood estimate of θ_j for the experiment. For example, experiment 1 has $X_3 = -0.16$, indicating that subject 1 reacted about 16% faster than the average of the 10 subjects for that experiment. The lower number in each box is $\hat{\theta}_j$, the mean of the posterior distribution defined in (2.11). V is given at the bottom of each column, along with the shrinkage factor $(1 - \hat{B})$.

Efron (1975) averages the values of X_j over the 10 experiments to obtain a much more accurate unbiased estimate of θ_j. These are taken to be the true θ_j values even though they still have some sampling variability in them. They are listed in the second last column. The last column of Table 3 compares X_j with $\hat{\theta}_j$ over the 10 experiments. The two values given in each box $\sum_{i=1}^{10} (X_{ij} - \theta_j)^2$ and $\sum_{i=1}^{10} (\hat{\theta}_{ij} - \theta_j)^2$ are the sum of squared error risk values for the maximum likelihood and hierarchical Bayes estimators, respectively.

The first of these is greater than the second for eight out of the 10 subjects. Overall one computes

Table 4. Ten reaction-time experiments. Upper interval is the 95% maximum likelihood interval—

Subject	Experiment				
	1	2	3	4	5
1	−0.24~0.18 −0.17~0.13	−0.22~0.34 −0.16~0.21	−0.30~0.14 −0.23~0.12	−0.27~0.19 −0.22~0.16	−0.26~0.16 −0.21~0.14
2	−0.13~0.29 −0.11~0.20	−0.04~0.52 −0.12~0.30	−0.05~0.39 −0.07~0.29	−0.09~0.37 −0.10~0.29	−0.27~0.15 −0.22~0.13*
3	−0.37~0.05 −0.25~0.08	−0.21~0.35 −0.15~0.22*	−0.54~ −0.11 −0.41~0.02	−0.53~ −0.07 −0.41~0.03	−0.49~ −0.07 −0.39~0.00
4	−0.25~0.17 −0.17~0.13	−0.26~0.30 −0.17~0.19	−0.20~0.24 −0.16~0.19	−0.19~0.27 −0.16~0.22	−0.26~0.16 −0.21~0.14
5	−0.13~0.29 −0.11~0.20	−0.27~0.29 −0.18~0.19	−0.19~0.25 −0.15~0.19	−0.23~0.23 −0.19~0.19	−0.07~0.35 −0.08~0.28
6	0.04~0.46 −0.07~0.30	−0.13~0.43 −0.13~0.26	−0.05~0.39 −0.07~0.29	0.12~0.58 −0.01~0.46	0.00~0.42 −0.04~0.34
7	−0.12~0.30 −0.10~0.21	−0.38~0.18 −0.23~0.14	−0.28~0.16 −0.21~0.13	−0.14~0.32 −0.13~0.25	−0.09~0.51 −0.01~0.41
8	−0.34~0.08 −0.23~0.09	−0.42~0.14 −0.25~0.13	−0.38~0.06 −0.29~0.08	−0.46~0.00 −0.36~0.06	−0.44~ −0.02 −0.35~0.03
9	−0.30~0.12 −0.20~0.11	−0.41~0.15 −0.25~0.14	−0.05~0.39 −0.07~0.29	−0.26~0.20 −0.21~0.17	−0.17~0.25 0.15~0.20
10	−0.25~0.17 −0.17~0.13	−0.44~0.12 −0.25~0.14	−0.15~0.29 −0.13~0.22	−0.25~0.21 −0.20~0.17	−0.24~0.18 −0.20~0.15

$$\sum_{i=1}^{10}\sum_{j=1}^{10}(X_{ij}-\theta_j)^2 = 1.025\,, \tag{2.27}$$

$$\sum_{i=1}^{10}\sum_{j=1}^{10}(\hat{\theta}_{ij}-\theta_j)^2 = 0.7057\,,$$

indicating that $\hat{\theta}_j$ is about 31% more accurate than X_j over all 10 experiments. Efron obtained an improvement of 25% using Stein's positive rule estimator but did not calculate any confidence intervals of the parameter values.

By calculating the second, third and fourth central moments of the posterior distribution and by using Pearson curves, the 95% approximated hierarchical Bayes confidence intervals can be obtained. These intervals are the lower values in each box in Table 4. The upper values are the usual maximum likelihood intervals, i.e. $X_{ij} \pm 1.96\sqrt{V_i}$, $i = 1,\ldots, 10, j = 1,\ldots, 10$. From the table it can also be seen that the hierarchical Bayes intervals are substantially shorter than the usual intervals. They are in fact 19% shorter while containing the true values with the same frequency as the ordinary intervals.

lower interval is the 95% hierarchical Bayes interval.

Experiment					
6	7	8	9	10	θ_j
− 0.01~0.31 − 0.03~0.29	− 0.15~0.25 − 0.13~0.21	− 0.17~0.23 − 0.13~0.16	− 0.31~0.11 − 0.25~0.10	− 0.31~0.19 − 0.21~0.15	0.00
0.21~0.53* 0.16~0.48*	0.00~0.40 − 0.03~0.32	− 0.09~0.31 − 0.09~0.22	− 0.05~0.47 − 0.01~0.38	− 0.32~0.18 − 0.22~0.141	0.14
− 0.34~ − 0.02 − 0.31~0.00	− 0.39~0.01 − 0.32~0.04	− 0.45~ − 0.05 − 0.30~0.06	− 0.29~0.13 − 0.24~0.12	− 0.34~0.16 − 0.23~0.13	− 0.18
− 0.21~0.11 − 0.20~0.11	− 0.36~0.04 − 0.29~0.06	− 0.22~0.18 − 0.16~0.13	− 0.08~0.34 − 0.08~0.28	− 0.37~0.13 − 0.25~0.12	− 0.03
− 0.10~0.22 − 0.10~0.21	− 0.42~ − 0.024* − 0.34~0.02	− 0.20~0.20 − 0.14~0.15	− 0.44~ − 0.024* − 0.35~0.02	− 0.35~0.15 − 0.24~0.13	− 0.02
0.08~0.40 0.05~0.37	0.09~0.49 0.02~0.40	− 0.18~0.216* − 0.13~0.16*	0.07~0.49 0.00~0.40	− 0.03~0.57 − 0.10~0.31	0.22
− 0.07~0.25 − 0.08~0.23	− 0.11~0.29 − 0.10~0.24	− 0.09~0.31 − 0.09~0.22	− 0.14~0.28 − 0.12~0.23	0.04~0.54 − 0.09~0.35	0.10
− 0.50~ − 0.176 − 0.45~ − 0.13	− 0.38~0.02 − 0.31~0.05	− 0.28~0.12 − 0.19~0.10	− 0.36~0.06 − 0.29~0.07	− 0.35~0.15 − 0.24~0.13	− 0.18
− 0.36~ − 0.036 − 0.33~ − 0.02	− 0.17~0.23 − 0.15~0.19	− 0.28~0.12 − 0.19~0.10	− 0.38~0.04 − 0.31~0.06	− 0.20~0.30 − 0.15~0.21	− 0.04
− 0.30~0.02 − 0.28~0.03	− 0.11~0.29 − 0.10~0.24	− 0.03~0.37 − 0.07~0.26	− 0.21~0.21 − 0.18~0.18	− 0.27~0.23 − 0.19~0.17	− 0.01

Both procedures were successful 96 out of 100 times. The four unsuccessful cases for each procedure i.e., the intervals that do not contain the parameter values are denoted by asterisks in Table 4.

For some of the posterior distributions the kurtosis coefficients (β_2-values) were as large as 11.997, 13.960, 15.680 and 22.815, while the largest skewness coefficients (β_1-values) were 0.1994, 0.2110, 0.2042 and 0.3097. This shows that the Pearson curve approach should be more appropriate to obtain confidence intervals than the normal approximations $\hat{\theta}_{ij} \pm 1.96\sqrt{\sigma_{ij}^{*2}(\underline{X})}$ where $\hat{\theta}_{ij}$ and $\sigma_{ij}^{*2}(\underline{X})$ are the mean and variance of the posterior distribution.

If we are also interested in obtaining confidence intervals of $(\theta_j - \theta_{j'})$, the difference between two treatment effects, then the posterior distribution of $(\theta_j - \theta_{j'})|\underline{X}$ is needed. We can again approximate this distribution by a Pearson distribution. By using the same arguments as in Theorem 2.1, it can be shown that the mean, variance, third and fourth central moments of $(\theta_j - \theta_{j'})|\underline{X}$ are given by

$$m_1' = (1 - \hat{B})(X_j - X_{j'}), \tag{2.28}$$

$$m_2 = 2V(1 - \hat{B}) - \left\{ \hat{B}^2 - \frac{2(m+1)}{S}\hat{B} + \frac{2m}{Se_m(S)} \right\}(X_j - X_{j'})^2, \tag{2.29}$$

$$\begin{aligned} m_3 = {} & 6V\left\{ \frac{2(m+1)}{S}\hat{B} - \frac{2m}{Se_m(S)} - \hat{B}^2 \right\}(X_j - X_{j'}) \\ & + 2\left\{ \frac{m}{Se_m(S)} + \frac{2m(m+2)}{S^2e_m(S)} - \frac{2(m+1)(m+2)}{S^2}\hat{B} - \hat{B}^3 \right. \\ & \left. + \frac{3(m+1)}{S}\hat{B}^2 - \frac{3m}{Se_m(S)}\hat{B} \right\}(X_j - X_{j'})^3, \end{aligned} \tag{2.30}$$

and

$$\begin{aligned} m_4 = {} & 12V^2\left\{ 1 - 2\hat{B} - \frac{2m}{Se_m(S)} + \frac{2(m+1)}{S}\hat{B} \right\} \\ & + 12V\left\{ \frac{2(m+1)}{S}\hat{B} + \frac{4m(m+2)}{S^2e_m(S)} - \frac{4(m+1)(m+2)}{S^2}\hat{B} \right. \\ & \left. - \hat{B}^2 - \frac{4m}{Se_m(S)}\hat{B} + \frac{4(m+1)}{S}\hat{B}^2 - \hat{B}^3 \right\}(X_j - X_{j'})^2 \\ & + \left\{ -\frac{2m}{Se_m(S)} - \frac{4m(m+3)}{S^2e_m(S)} - \frac{8m(m+2)(m+3)}{S^3e_m(S)} \right. \end{aligned} \tag{2.31}$$

$$+\frac{8(m+1)(m+2)(m+3)}{S^3}\hat{B}+\frac{8m}{Se_m(S)}\hat{B}$$

$$+\frac{16m(m+2)}{S^2e_m(S)}\hat{B}-\frac{16(m+1)(m+2)}{S^2}\hat{B}^2+\frac{12(m+1)}{S}\hat{B}^3$$

$$-\left.\frac{12m}{Se_m(S)}\hat{B}^2-3\hat{B}^4\right\}(X_j-X_{j'})^4\,,$$

where $j=1,\ldots,k$, $j'=1,\ldots,k$, $j\neq j'$.

Suppose for example we want to obtain a hierarchical Bayes confidence interval of $(\theta_4-\theta_{10})$. Using only the data in experiment 1, we find that $S=10.7642$, $e_m(S)=5.9502939$, $m_1'=0$, $m_2=0.0100976$, $m_3=0$, $m_4=0.00037901$, $\beta_1=0$ and $\beta_2=m_4/m_2^2=3.7170883$. The 95% empirical Bayes confidence interval is given by $-0.196954\leq(\theta_4-\theta_{10})\leq0.196954$ which includes the real difference -0.02 and is substantially shorter than the usual interval $-0.2907\sim0.2907$.

It is perhaps not surprising that the hierarchical Bayes intervals given in Examples 2.1 and 2.2 have such good frequentist properties because as mentioned by Morris (1983) and Berger ((1985), p. 172, the last paragraph), hierarchical Bayes procedures (or related procedures) have frequentist risks (for sum of squares-error loss) and coverage probability comparable to or better than those of the corresponding classical procedures (unless k is very small).

For the unknown variance case corresponding results can be derived as given in (2.1)–(2.18) and (2.28)–(2.31). Instead of the normal distribution obtained in equation (2.4), a t-distribution will be derived and the distribution defined in (2.5) will become a Beta-Type II distribution on the interval 0 to 1.

3. Hierarchical Bayes and empirical Bayes confidence intervals—the unequal variance case

In this section the variances are considered known and unequal in contrast with the previous section where they were considered known and equal. Suppose that

(3.1) $$X_j|\theta_j\sim N(\theta_j,V_j)\qquad j=1,\ldots,k\,,$$

and

(3.2) $$\theta_j\sim N(\mu,A)\qquad j=1,\ldots,k\,,$$

then it is well-known that

(3.3) $$\theta_j|\underline{X}, \mu, A \sim N(\mu + (1 - B_j)(X_j - \mu); V_j(1 - B_j)),$$

where

(3.4) $$B_j = \frac{V_j}{V_j + A}.$$

By assuming that the distribution of μ is a Lebesgue uniform measure on $(-\infty, \infty)$ and independent of A, we have

(3.5) $$\mu|\underline{X}, A \sim N\left(\hat{\mu}, \left(\sum_{j=1}^{k} \frac{B_j}{V_j}\right)^{-1}\right),$$

where

(3.6) $$\hat{\mu} = \left(\sum_{j=1}^{k} \frac{B_j}{V_j}\right)^{-1} \sum_{j=1}^{k} \frac{B_j X_j}{V_j}.$$

From equations (3.3) and (3.5) it follows that given A and $\underline{X}$, θ_j will be normally distributed with mean.

(3.7) $$E(\theta_j|\underline{X}, A) = \hat{\mu} + (1 - B_j)(X_j - \hat{\mu}),$$

and variance

(3.8) $$\text{Var}(\theta_j|\underline{X}, A) = V_j(1 - B_j) + B_j^2\left(\sum_{j=1}^{k} \frac{B_j}{V_j}\right)^{-1}.$$

The posterior distribution of A can be obtained by assuming that A is uniformly distributed on the interval $[0, \infty)$ (Rubin (1981)). Using the transformation (3.4) the posterior density of B_j is then given by

(3.9) $$f(B_j|\underline{X}) = K_j \left(\frac{\prod_{i=1}^{k} \dfrac{1}{V_i B_j + V_j(1 - B_j)}}{\sum_{i=1}^{k} \dfrac{1}{V_i B_j + V_j(1 - B_j)}} \right)^{1/2} V_j B_j^{(k-5)/2}$$

$$\cdot \exp\left\{ -\frac{B_j}{2}\left[\sum_{i=1}^{k} \frac{X_i^2}{V_i B_j + V_j(1 - B_j)} \right.\right.$$

$$
\left. \left. - \frac{\left(\sum_{i=1}^{k} \frac{X_i}{V_i B_j + V_j(1 - B_j)} \right)^2}{\sum_{i=1}^{k} \frac{1}{V_i B_j + V_j(1 - B_j)}} \right] \right\}
$$

$$
= K_j \, g(B_j), \qquad 0 \leq B_j \leq 1 \,,
$$

where

$$
K_j = \left\{ \int_0^1 g(B_j) dB_j \right\}^{-1} . \tag{3.10}
$$

In the unequal variance case it is not possible to obtain exact closed form statistics like those in the preceding sections. It is, however, possible to calculate the exact moments of $\theta_j | \underline{X}$ numerically. Let

$$
P_j = B_j \frac{\sum_{i=1}^{k} \frac{X_i}{V_i B_j + V_j(1 - B_j)}}{\sum_{i=1}^{k} \frac{1}{V_i B_j + V_j(1 - B_j)}} + (1 - B_j) X_j \,,
$$

$$
R_j = V_j(1 - B_j) + \frac{B_j}{\sum_{i=1}^{k} \frac{1}{V_i B_j + V_j(1 - B_j)}} \,,
$$

and

$$
M_{1j} = \int_0^1 P_j f(B_j | \underline{X}) dB_j = E(P_j) \,.
$$

Also,

$$
M_{2j} = E(R_j), \qquad M_{3j} = E(P_j^2), \qquad M_{4j} = E(R_j P_j), \qquad M_{5j} = E(P_j^3),
$$

$$
M_{6j} = E(R_j^2), \qquad M_{7j} = E(R_j P_j^2), \qquad M_{8j} = E(P_j^4) \,.
$$

Then the moments about zero are given by

$$
E(\theta_j | \underline{X}) = M_{1j} = \hat{\theta}_j \,, \tag{3.11}
$$

$$
E(\theta_j^2 | \underline{X}) = M_{2j} + M_{3j} \,, \tag{3.12}
$$

$$
E(\theta_j^3 | \underline{X}) = 3 M_{4j} + M_{5j} \,, \tag{3.13}
$$

and

$$E(\theta_j^4|\underline{X}) = 3M_{6j} + 6M_{7j} + M_{8j} \quad j = 1,\ldots,k\,. \tag{3.14}$$

By making use of the relationship between moments about zero and central moments, $\mathrm{Var}\,(\theta_j|\underline{X})$, the coefficients of skewness $\sqrt{\beta_1} = E\{(\theta_j - \hat{\theta}_j)^3|\underline{X}\}/\{\mathrm{Var}\,(\theta_j|\underline{X})\}^{3/2}$ and kurtosis $\beta_2 = E\{(\theta_j - \hat{\theta}_j)^4|\underline{X}\}/\{\mathrm{Var}\,(\theta_j|\underline{X})\}^2$ can be calculated. By using Pearson curves good approximations of the hierarchical Bayes confidence intervals can be obtained.

Since numerical integration and simulation studies can be time consuming, it will be helpful if closed form approximations of certain expected values can be derived. Approximations of $E(\theta_j|\underline{X})$ and $\mathrm{Var}\,(\theta_j|\underline{X})$ can be obtained by approximating $E(B_j|\underline{X})$. This can be achieved by deriving an approximately unbiased estimate of A and substituting it into (3.4). An unbiased estimate of A, relative to the marginal distributions of X_j and the grand mean $\overline{\overline{X}} = (1/k)\sum_{j=1}^{k} X_j$, is given by

$$\hat{A}_j = \frac{k}{k-1}(X_j - \overline{\overline{X}})^2 - \frac{k-2}{k-1}V_j - \frac{1}{k-1}\overline{\overline{V}} \quad j = 1,\ldots,k\,, \tag{3.15}$$

with variance

$$\mathrm{Var}\,(\hat{A}_j) = 2\left(\frac{k-2}{k-1}V_j + \frac{1}{k-1}\overline{\overline{V}} + A\right)^2 \quad j = 1,\ldots,k\,, \tag{3.16}$$

where $\overline{\overline{V}} = (1/k)\sum_{j=1}^{k} V_j$.

The final estimator $\hat{A}_p$, for A, based on all the data, is then the weighted mean of all the $\hat{A}_j$'s with weights inversely proportional to their standard deviations (from (3.16)). Notice that the weights themselves must be estimated by finding an initial estimate for A. To avoid iteration without losing too much accuracy, A is estimated by the unweighted mean of the $\hat{A}_j$'s. So

$$\hat{A}_p = \frac{\Sigma W_j \hat{A}_j^+}{\Sigma W_j}\,,$$

where

$$W_j = \frac{k-1}{(k-1)\,V_j + \overline{\overline{V}} + (k-1)\hat{\overline{A}}^+}\,,$$

and

$$\hat{\hat{A}} = \frac{1}{k-1} \sum_{j=1}^{k} (X_j - \overline{\overline{X}})^2 - \overline{\overline{V}} ,$$

with

$$y^+ = \max(0, y) .$$

The estimator $\hat{A}_p$ is not the maximum likelihood estimator for A, but is similar to the one given by Morris (1983*b*). Other related estimates appear in Carter and Rolph (1974), Dempster *et al.* (1977) and Fay and Herriot (1979).

The estimator for $E(B_j|\underline{X})$ is then given by

$$\hat{B}_j = \frac{k-3}{k-1} \frac{V_j}{V_j + \hat{A}_p} \quad j = 1, \ldots, k , \tag{3.17}$$

where the constant $(k-3)/(k-1)$ helps to correct for the non-linear dependence of B_j on A.

Approximations of $E(\theta_j|\underline{X})$ and Var $(\theta_j|\underline{X})$ are given by

$$\hat{E}(\theta_j|\underline{X}) = \hat{\hat{\mu}} + (1 - \hat{B}_j)(X_j - \hat{\hat{\mu}}) , \tag{3.18}$$

(where $\hat{\hat{\mu}}$ is $\hat{\mu}$ in (3.6) with B_j replaced by $\hat{B}_j$) and

$$\hat{\mathrm{V}}\mathrm{ar}\,(\theta_j|\underline{X}) = V_j(1 - \hat{B}_j) + \hat{B}_j^2\,\hat{\gamma} + v_j[\hat{\gamma} + (X_j - \hat{\hat{\mu}})^2] , \tag{3.19}$$

where

$$\hat{\gamma} = \left(\sum_{j=1}^{k} \frac{\hat{B}_j}{V_j} \right)^{-1} , \tag{3.20}$$

and

$$v_j = \frac{2(k-2)}{(k-1)} \frac{\hat{B}_j^4}{V_j} \left(\frac{1}{k} \sum_{j=1}^{k} (V_j - \overline{\overline{V}})^2 + \frac{k-1}{k-2} (\overline{\overline{V}} + \hat{A}_p)^2 \right) , \tag{3.21}$$

approximates the variance of $\hat{B}_j$. The derivations of the approximations ((3.15)–(3.21)) are quite lengthy and are given in Groenewald and van der Merwe (1985). An advantage of these approximations is that $\hat{A}_p$ will only be zero if all the component estimates of A, $\hat{A}_j$, $j = 1, \ldots, k$, are zero. It is therefore less likely that an (unrealistic) zero estimate of A will be obtained.

By assuming that $\theta_j|\underline{X}$ is approximately normal the $(1-\alpha)100\%$ empirical Bayes confidence intervals can be obtained. As mentioned by Morris (1977) it is not precisely true that $\theta_j|\underline{X}$, $j=1,\ldots,k$, has the normal distribution. But θ_j does have a normal distribution for fixed A and if A is estimated by $\hat{A}_p$ without large variance, then the normal distribution should hold approximately. On the other hand if the posterior variance of A is large in comparison with V_j, $j=1,\ldots,k$, then large skewness and kurtosis values will occur.

Example 3.1 *Hierarchical Bayes and empirical Bayes confidence intervals in the unequal variance case*

Alderman and Powers (1979) studied the effects of high school coaching programs on SAT-V (Scholastic Aptitute Test-Verbal) scores in eight schools, each school conducting a separate randomized experiment. Rubin (1981) analyzed these data with the purpose of illustrating hierarchical Bayesian techniques that can be used to help summarize the evidence in such data about differences among treatments, thereby obtaining improved estimates of the treatment effect in each experiment. Here X_j, $j=1,\ldots,8$, is the unbiased estimate of the j-th treatment. By using equations (3.9)–(3.21), $E(\theta_j|\underline{X})$, $\text{Var}(\theta_j|\underline{X})$, $\sqrt{\beta_1}$, β_2, $\hat{E}(\theta_j|\underline{X})$ and $\hat{\text{V}}\text{ar}(\theta_j|\underline{X})$ were calculated. These values as well as X_j and V_j obtained from Rubin (1981) are given in Table 5.

From Table 5 it can be seen that $\hat{E}(\theta_j|\underline{X})$ and $\hat{\text{V}}\text{ar}(\theta_j|\underline{X})$ are good approximations of the true parameter values. The kurtosis β_2 for all eight schools are greater than three. The deviations of the skewness and kurtosis values from 0 and 3, respectively, are indications that improvements in confidence intervals can be obtained by using Pearson curves. In Table 6 the following intervals are given:

PHBCIS: The Pearson curve approximations to the exact hierarchical Bayes confidence intervals.

EBCIS: The empirical Bayes confidence intervals with normal approximation from (3.18) and (3.19); $\hat{E}(\theta_j|\underline{X}) \pm 1.96\,(\hat{\text{V}}\text{ar}(\theta_j|\underline{X}))^{1/2}$.

Table 5. X_j, V_j, exact and estimated moments of $\theta_j|\underline{X}$.

School	A	B	C	D	E	F	G	H
X_j	28.4	8.0	− 2.8	6.8	− 0.6	0.6	18.0	12.2
V_j	222.01	104.04	265.69	121.00	88.36	129.96	108.16	309.76
$E(\theta_j\|\underline{X})$	14.94	8.00	4.86	7.52	3.71	4.90	12.66	9.32
$\text{Var}(\theta_j\|\underline{X})$	110.17	57.73	101.17	63.24	56.41	68.81	66.28	104.00
$\sqrt{\beta_1}$	0.610	0.018	− 0.371	− 0.025	− 0.261	− 0.250	0.325	0.166
β_2	3.741	3.408	3.990	3.484	3.284	3.504	3.305	4.090
$\hat{E}(\theta_j\|\underline{X})$	15.10	7.98	4.36	7.54	4.34	5.15	12.04	9.38
$\hat{\text{V}}\text{ar}(\theta_j\|\underline{X})$	108.69	56.24	107.09	61.32	64.13	69.19	70.30	116.62

Table 6. The 95% hierarchical Bayes, empirical Bayes and usual confidence intervals of the treatment effects for the eight schools.

School	A	B	C	D	E	F	G	H
PHBCIS	(− 2.7 ; 38.3)	(− 6.9 ; 23.1)	(− 16.5 ; 23.4)	(− 8.3 ; 23.1)	(− 11.9 ; 17.6)	(− 12.4 ; 20.14)	(− 2.1 ; 29.8)	(− 10.4 ; 30.2)
EBCIS	(− 5.3 ; 35.6)	(− 6.7 ; 22.7)	(− 15.9 ; 24.6)	(− 7.8 ; 22.9)	(− 11.4 ; 20.0)	(− 11.2 ; 21.5)	(− 4.4 ; 28.5)	(− 11.8 ; 30.6)
SHBCIS	(− 2 ; 36)	(− 6 ; 19)	(− 10 ; 22)	(− 7 ; 21)	(− 9 ; 16)	(− 8 ; 20)	(− 1 ; 24)	(− 3 ; 24)
MLCIS	(− 0.8 ; 57.6)	(− 12.1 ; 27.9)	(− 34.7 ; 29.2)	(− 14.7 ; 28.4)	(− 19.1 ; 17.8)	(− 21.7 ; 23.0)	(− 2.4 ; 38.4)	(− 22.3 ; 46.7)

SHBCIS: Simulated Bayes confidence intervals conducted by Rubin (1981).

MLCIS: Usual maximum likelihood confidence intervals, $X_j \pm 1.96$ $(\text{Var}(X_j))^{1/2}$.

A comparison of the PHBCIS and the SHBCIS shows that the latter is not very accurate. This is not surprising because these intervals were obtained from only 200 simulations. It is also clear that the empirical Bayes intervals (EBCIS) are good approximations of the true intervals (PHBCIS). In other experiments with larger skewness and kurtosis values these approximations will, however, be less good. Both the PHBCIS and EBCIS are much shorter than the usual intervals (MLCIS) which have an average length of 49.6. In fact the average hierarchical Bayes confidence interval (PHBCI) is 30.34% shorter while the average empirical Bayes confidence interval is 29.3% shorter than the maximum likelihood intervals.

Acknowledgement

The authors are grateful to the referee for valuable comments and suggestions.

REFERENCES

Alderman, D. and Powers, D. (1979). The effects of special preparation on SAT-Verbal scores, *Research Report*, 79-1, Princetown, New Jersey: Educational Testing Service.

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

Carter, G. and Rolph, J. (1974). Empirical Bayes methods applied to estimating fire alarm probabilities, *J. Amer. Statist. Assoc.*, **69**, 880-885.

Dempster, A. P., Schatzoff, M. and Wermuth, H. (1977). A simulation study of alternatives to ordinary least squares, *J. Amer. Statist. Assoc.*, **72**, 77-106.

Efron, B. (1975). Biased versus unbiased estimation, reprinted from *Adv. in Math.*, **16**, 259-277.

Efron, B. and Morris, C. (1975). Data analysis using Stein's estimator and its generalizations, *J. Amer. Statist. Assoc.*, **70**, 311-319.

Elderton, W. P. (1953). *Frequency Curves and Correlation*, Cambridge University Press.

Elderton, W. P. and Johnson, N. L. (1969). *Systems of Frequency Curves*, Cambridge University Press.

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

Groenewald, P. C. N. and van der Merwe, A. J. (1985). Model selection—using normal priors and predictive sample re-use, Tech. Report No. 106, Dept. of Math. Statistics, Univ. of the Orange Free State.

Johnson, N. L., Nixon, E., Amos, D. and Pearson, E. S. (1963). Table of percentage points of Pearson curves for given $\sqrt{\beta_1}$ and β_2 expressed in standard measure, *Biometrika*, **50**, 459-498.

Morris, C. N. (1977). *Interval Estimation for Empirical Bayes Generalizations of Stein's Estimator*, The Rand Paper Series, Rand Corp., California.

Morris, C. N. (1983*a*). *Parametric Empirical Bayes Confidence Intervals, Scientific Inference, Data Analysis, and Robustness*, 25–50, Academic Press, New York.

Morris, C. N. (1983*b*). Parametric empirical Bayes inference. Theory and applications, *J. Amer. Statist. Assoc.*, **78**, 47–65.

Rubin, D. B. (1981). Estimation in parallel randomized experiments, *Journal of Educational Statistics*, **6**, 377–400.