DATA ANALYSIS
USING STEIN'S ESTIMATOR
AND ITS GENERALIZATIONS

BRADLEY EFRON
CARL MORRIS

The research reported herein was performed pursuant to a grant from the Office of Economic Opportunity, Washington, D.C. Reports of The Rand Corporation do not necessarily reflect the opinions or policies of the sponsors of Rand research.

Published by The Rand Corporation

DATA ANALYSIS
USING STEIN'S ESTIMATOR
AND ITS GENERALIZATIONS

PREPARED FOR THE OFFICE OF ECONOMIC OPPORTUNITY

BRADLEY EFRON
CARL MORRIS

R-1394-OEO

MARCH 1974

PREFACE

The most commonly used statistical estimators are known to be in-
efficient for problems involving estimation of several parameters simul-
taneously. In a series of earlier papers (see Efron and Morris, 1971,
1972a and c, 1973a and b, forthcoming c) we suggested and examined more
efficient rules than the usual ones and identified cases where it is
theoretically possible to make substantial improvements over the usual
methods. The purpose of this report is to analyze three sets of data
for concrete illustrations of the methods, rewards, and difficulties of
using improved estimates on real data. It is hoped that this report and
others being written will make applied researchers more able and willing
to use improved methods for multiparameter estimation, as these methods
are believed to be capable of substantially improving estimation in many
applied statistical problems.

This work has been sponsored in part by the Office of Economic
Opportunity through grants 90088 D-72-02 and 90088 D-73-01 to The Rand
Corporation. OEO also supports the Health Insurance Study (HIS) in which
both authors are involved. The methods of this report will be used to
improve the analysis and design of the HIS in several ways: (a) for
estimating interactions in fitted linear models; (b) for estimating city-
specific responses from data for all cities; (c) for stabilizing surveyed
physician responses in cities as inputs to site selection; and (d) for
stabilizing family-specific cost estimates when cost-effective experi-
mental design choices are being made. Methods similar to those of Sec.
III have also been used by Rand in New York City (G. M. Carter and

J. E. Rolph, *New York City Fire Alarm Prediction Models: I: Box Reported Serious Fires*, The Rand Corporation, R-1214-NYC, May 1973).

Dr. Efron is a consultant to The Rand Corporation and is a Professor in the Department of Statistics, Stanford University. Dr. Morris is on the staff of The Rand Corporation.

SUMMARY

In 1961, James and Stein produced an estimator for the mean of a multivariate normal distribution with uniformly lower mean squared error than the sample mean. This estimator and several of its generalizations are presented briefly in an empirical Bayes context and are then applied to three examples with real data. These estimators perform much better than the classical estimators in each example.

The first application predicts final 1970 batting averages for 14 major league players from their early season performances. The predictions resulting from Stein's estimator are better than the maximum likelihood estimator for every batter. Then toxoplasmosis prevalence rates for 36 El Salvador cities are estimated. Additional analysis indicates that the generalization of Stein's estimator used for this situation is several times better than the usual estimator. Finally, in 51 situations a computer simulation is used to estimate the exact size of Pearson's chi-square test for comparing binomial means. Comparisons of the various estimates with the true values reveal Stein's estimator and its multivariate generalizations to be approximately twice as efficient as the maximum likelihood estimator.

ACKNOWLEDGMENTS

CONTENTS

I. INTRODUCTION

Charles Stein showed that for the problem of estimating several parameters from independent normal observations it was possible to make a uniform improvement upon the maximum likelihood estimator (MLE) in terms of total squared error risk (Stein, 1955). Later James and Stein (1961) presented a particularly simple estimator for which the improvement was quite substantial near the origin, as long as there are more than two parameters. This achievement leads immediately to a uniform, nontrivial improvement over the least squares (Gauss-Markov) estimators for the parameters in the usual formulation of the linear model. One might expect a rush of applications of this powerful new statistical weapon, but such has not been the case. Resistance has formed along several lines:

1. Mistrust of the statistical interpretation of the mathematical formulation leading to Stein's result, in particular the sum of squared errors loss function;

2. Difficulties in adopting the James-Stein estimator to the many special cases that invariably arise in practice;

3. Long familiarity with the generally good performance of the MLE in applied problems;

4. A feeling that any gains possible from a "complicated" procedure like Stein's could not be worth the extra trouble. (J. W. Tukey, at the 1972 American Statistical Association meetings in Montreal stated that savings would not be more than 10 percent in practical situations.)

We have written a series of papers (Efron and Morris, 1971, 1972a and c, 1973a and b, forthcoming) that cover points 1 and 2. Our purpose here is to illustrate the methods suggested in these papers on three applied problems and in that way deal with points 3 and 4. Only one of the three problems, the toxoplasmosis data, is "real" in the sense of being generated outside the statistical world. The other two problems are contrived to illustrate in a realistic way the genuine difficulties and rewards of procedures of the Stein type. (They have the added advantage of having the true parameter values available for the comparison of different methods.) The examples chosen are the first and only ones considered for this report, and the favorable results typify our previous experience.

To review the James-Stein estimator (1961) in the simplest setting, suppose that for given θ_i

$$X_i | \theta_i \overset{ind}{\sim} N(\theta_i, 1) \quad i = 1, \ldots, k \geq 3 \tag{1.1}$$

(meaning the $\{X_i\}$ are independent and normally distributed with mean $E_{\theta_i} X_i \equiv \theta_i$ and variance $\text{Var}_{\theta_i}(X_i) = 1$). The example (1.1) typically occurs as a reduction to this canonical form from more complicated situations, as when X_i is a sample mean with known variance that is taken to be unity through an appropriate scale transformation. The unknown vector of means $\underset{\sim}{\theta} \equiv (\theta_1, \ldots, \theta_k)$ is to be estimated with loss being the sum of squared component errors

$$L(\underset{\sim}{\theta}, \underset{\sim}{\hat{\theta}}) \equiv \sum_{i=1}^{k} (\hat{\theta}_i - \theta_i)^2 \tag{1.2}$$

where $\hat{\underset{\sim}{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)$ is the estimate of $\underset{\sim}{\theta}$. The maximum likelihood estimator, which is also the sample mean, $\underset{\sim}{\delta}^0(\underset{\sim}{X}) \equiv \underset{\sim}{X} \equiv (X_1, \ldots, X_k)$ has constant risk k,

$$R(\underset{\sim}{\theta}, \underset{\sim}{\delta}^0) \equiv E_{\underset{\sim}{\theta}} \sum_{i=1}^{k} (X_i - \theta_i)^2 = k \qquad (1.3)$$

where $E_{\underset{\sim}{\theta}}$ indicates expectation over the distribution (1.1). James and Stein (1961) introduced the estimator $\underset{\sim}{\delta}^1(\underset{\sim}{X}) = (\delta_1^1(\underset{\sim}{X}), \ldots, \delta_k^1(\underset{\sim}{X}))$ for $k \geq 3$,

$$\delta_i^1(\underset{\sim}{X}) \equiv \mu_i + (1 - \frac{k-2}{S})(X_i - \mu_i) \quad i = 1, \ldots, k \qquad (1.4)$$

where $\underset{\sim}{\mu} \equiv (\mu_1, \ldots, \mu_k)'$ is any initial guess at $\underset{\sim}{\theta}$ and $S \equiv \sum(X_j - \mu_j)^2$. This estimator has risk

$$R(\underset{\sim}{\theta}, \underset{\sim}{\delta}^1) \equiv E_{\underset{\sim}{\theta}} \sum_{i=1}^{k} (\delta_i^1(\underset{\sim}{X}) - \theta_i)^2 \qquad (1.5)$$

$$\leq k - \frac{(k-2)^2}{k-2 + \sum(\theta_i - \mu_i)^2} < k, \qquad (1.6)$$

which is less than k for all $\underset{\sim}{\theta}$, and if $\theta_i = \mu_i$ for all i the risk is 2, comparing very favorably to k for the MLE. Risks like (1.5) are tabled in Efron and Morris (forthcoming).

The estimator (1.4) arises quite naturally in an empirical Bayes context. If the $\{\theta_i\}$ themselves are a sample from a prior distribution,

$$\theta_i \overset{ind}{\sim} N(\mu_i, \tau^2) \quad i = 1, \ldots, k, \qquad (1.7)$$

then the Bayes estimate of θ_i is the *a posteriori* mean of θ_i given the data

$$\delta_i^*(X_i) = E\theta_i|X_i = \mu_i + (1 - \frac{1}{1+\tau^2})(X_i - \mu_i). \tag{1.8}$$

In the empirical Bayes situation, τ^2 is unknown, but it can be estimated because marginally the $\{X_i\}$ are independently normal with means $\{\mu_i\}$ and

$$S \equiv \sum(X_i - \mu_i)^2 \sim (1 + \tau^2)\chi_k^2, \tag{1.9}$$

where χ_k^2 is the chi-square distribution with k degrees of freedom. The unbiased estimate

$$E(k-2)/S = 1/(1 + \tau^2) \tag{1.10}$$

is available, and substitution of $(k-2)/S$ for the unknown $1/(1 + \tau^2)$ in the Bayes estimate δ_i^* of (1.8) results in the James-Stein rule (1.4). The risk of δ_i^1 averaged over both $\underset{\sim}{X}$ and $\underset{\sim}{\theta}$ is, from Efron and Morris (1972a or 1973b),

$$E_\tau E_{\underset{\sim}{\theta}}(\delta_i^1(\underset{\sim}{X}) - \theta_i)^2 = 1 - \frac{k-2}{k} \frac{1}{1+\tau^2}, \tag{1.11}$$

where E_τ is expectation over the distribution (1.7). The risk (1.11) is to be compared to the corresponding risks of 1 for the MLE and $1 - 1/(1+\tau^2)$ for the Bayes estimator. Thus, if k is moderate or large, δ_i^1 is nearly as good as the Bayes estimator but avoids the possible gross errors of the Bayes estimator if τ^2 is misspecified.

It is clearly preferable to use $\min(1, (k-2)/S)$ as an estimate of $1/(1+\tau^2)$ instead of (1.10). This results in the simple improvement $\delta_i^{1+}(\underset{\sim}{X}) = \mu_i + (1-(k-2)/S)^+(X_i-\mu_i)$ where $a^+ \equiv \max(0, a)$. That $R(\underset{\sim}{\theta}, \underset{\sim}{\delta}^{1+}) < R(\underset{\sim}{\theta}, \underset{\sim}{\delta}^1)$ for all $\underset{\sim}{\theta}$ is proved in Baranchik (1964), Stein (1966), Efron and Morris (1973a). $R(\underset{\sim}{\theta}, \underset{\sim}{\delta}^{1+})$ is tabled in Efron and Morris (forthcoming).

In the applications that follow, we will adapt and generalize Stein's estimator to other situations. For the baseball problem, the familiar modification that requires estimation of the *a priori* mean $\underset{\sim}{\mu}$ will be given, and a variance stabilizing transformation for binomial data is used. Most other generalizations are new. A limited translation estimate (Efron and Morris, 1971 and 1972a) is used for the baseball problem. In the estimation of toxoplasmosis prevalence rates, Stein's rule is generalized to the case of unequal variances. For the computer simulation, Stein's estimator is used to smooth data by applying the canonical form estimator to the residuals of a regression. A multivariate generalization of Stein's estimator (Efron and Morris, 1972c) and a "two groups" estimator are among the new methods used in this application. The report ends with a discussion of the effectiveness and potential of multiparameter estimation for applied statistics. A lengthy bibliography of related work is appended.

The methods of this report can be justified from several statistical viewpoints--for example, the frequentist and decision theoretic viewpoint, and the empirical Bayes or the Bayesian viewpoint. We favor the empirical Bayes viewpoint because it has provided the most helpful hints for extending methods to the new situations. The bibliography provides theoretical and numerical justification from other

standpoints (Stein, 1955 and 1962; Efron and Morris, forthcoming).
Here, the rules are evaluated by their performance on real data and
are found to behave well independently of the assumptions for their
derivation.

II. USING STEIN'S ESTIMATOR TO PREDICT BATTING AVERAGES

Table 1 presents the batting averages of 14 major league players through their first 45 official at bats of the 1970 season. The problem is to predict each player's batting average over the remainder of the season using only the data of column (1) of Table 1. This sample was chosen because we wanted between 30 and 50 at bats to assure a satisfactory approximation of the binomial by the normal distribution and to leave the great bulk of at bats to be estimated. We also wanted to include an unusually good hitter (Clemente) to test the method with at least one extreme parameter, a situation expected to be less favorable to Stein's estimator. Batting averages are published weekly in

Table 1

1970 BATTING AVERAGES FOR 14 MAJOR LEAGUE PLAYERS

i		Y_i = Batting Average for First 45 at Bats	p_i = Batting Average for Remainder of Season	At Bats for Remainder of Season
		(1)	(2)	(3)
1	Clemente (Pitts, NL)	.400	.346	367
2	F. Robinson (Balt, AL)	.378	.298	426
3	F. Howard (Wash, AL)	.356	.276	521
4	Johnstone (Cal, AL)	.333	.221	276
5	Berry (Chi, AL)	.311	.273	418
6	Spencer (Cal, AL)	.311	.270	467
7	Kessinger (Chi, NL)	.289	.263	586
8	Santo (Chi, NL)	.244	.269	510
9	Unser (Wash, AL)	.222	.264	277
10	Williams (Chi, AL)	.222	.256	270
11	Scott (Bos, AL)	.222	.304	434
12	Petrocelli (Bos, AL)	.222	.264	538
13	Companeris (Oak, AL)	.200	.285	558
14	Munson (NY, AL)	.178	.319	405

the *New York Times*. In the April 24, 1970 publication, Clemente's

times at bat were given as 45. Stein's estimator requires equal vari-

ances, or in this situation, equal at bats, so the remaining 13 players

are all players with available batting averages at 45 at bats who also

batted at least 300 times over the entire season. (The unequal vari-

ances case is discussed in Sec. III.)

Let Y_i be the batting average of player i, i = 1, ..., 14 (k = 14)

after n = 45 at bats. Assuming base hits occur according to a binomial

distribution with independence between players, $nY_i \overset{ind}{\sim} Bin(n, p_i)$

i=1, 2, ..., 14 where p_i is the true season batting average, and

$EY_i = p_i$. Because the variance of Y_i depends on the mean, Anscombe's

(1948) modification of the arc-sin transformation for stabilizing the

variance of a binomial distribution is used. Thus, if

$$X_i \equiv f_n(Y_i) \quad i = 1, ..., 14 \tag{2.1}$$

with

$$f_n(y) \equiv (n + 0.5)^{\frac{1}{2}} \text{ arc } \sin((\frac{n}{n+0.75})(2y-1)), \tag{2.2}$$

then X_i has nearly unit variance independent of p_i. (An exact computer

computation showed that $Var(X_i)$ is within .0012 of unity for n = 45

for all p_i between 0.1 and 0.9.) Let θ_i be the mean of X_i, given

approximately by $\theta_i = f_n(p_i)$. From the central limit theorem for the

binomial distribution and the continuity of f_n we have approximately

$$X_i | \theta_i \overset{ind}{\sim} N(\theta_i, 1) \quad i = 1, 2, ..., k, \tag{2.3}$$

the situation described in the preceding section.

For most of this discussion we will regard the values of p_i of column 2, Table 1 as the quantities to be estimated, although these quantities are really only other estimates of the mean of Y_i. Therefore we actually have a prediction problem, but this consideration will be ignored at first. Thus $\theta_i = f_n(p_i)$ is determined from Table 1. The values of X_i, θ_i are given in Table 2.

Table 2

TRANSFORMED DATA

i	X_i	θ_i
1	-1.34	-2.11
2	-1.64	-2.80
3	-1.94	-3.12
4	-2.25	-3.98
5	-2.57	-3.18
6	-2.57	-3.22
7	-2.89	-3.33
8	-3.55	-3.24
9	-3.90	-3.31
10	-3.90	-3.44
11	-3.90	-2.71
12	-3.90	-3.31
13	-4.26	-2.99
14	-4.63	-2.50

We use Stein's estimator (1.4) except we assume $\mu_1 = \ldots = \mu_k = \mu$ (say) and estimate the common unknown value by $\overline{X} = \sum X_i/k$, shrinking all X_i toward $\overline{X}$, an idea suggested by Lindley (in Stein, 1962, pp. 285-287). The resulting estimate of the ith component θ_i of $\underset{\sim}{\theta}$ is therefore

$$\tilde{\delta}_i^1(\underset{\sim}{X}) = \overline{X} + (1 - \frac{k-3}{V})(X_i - \overline{X}) \qquad (2.5)$$

with $V \equiv \sum (X_i - \bar{X})^2$ and with k-3 = (k-1)-2 as the appropriate constant

since one parameter was estimated. In the empirical Bayes case, the

appropriateness of (2.5) also follows from estimating the Bayes rule

(1.8) by using the unbiased estimated $\bar{X}$ for μ and (k-3)/V for $1/(1+\tau)^2$

on the marginal distribution of $\underset{\sim}{X}$, analogous to Sec. I (also see Efron

and Morris, 1972a, Sec. 7). We may use the Bayesian model for these

data because (1.7) seems at least roughly appropriate, although (2.5)

can be justified by the non-Bayesian from the suspicion that $\sum (\theta_i - \bar{\theta})^2$

is small, since the risk of (2.5), analogous to (1.6), is bounded by

$$R(\underset{\sim}{\theta}, \underset{\sim}{\tilde{\delta}^1}) \leq k - \frac{(k-3)^2}{k-3 + \sum_i (\theta_i - \bar{\theta})^2}, \quad \bar{\theta} \equiv \sum_i \theta_i / k. \qquad (2.6)$$

For our data, the estimate of $1/(1+\tau^2)$ is (k-3)/V = .766, or

$\hat{\tau}^2 = .306$, $\hat{\tau} = 0.553$, representing considerable *a priori* information.

The value of $\bar{X}$ is -3.09 so the estimate (2.5) for these data is

$$\tilde{\delta}_i^1 (\underset{\sim}{X}) = \hat{\theta}_i = .766 \, \bar{X} + .234 \, X_i = .234 \, X_i - 2.37. \qquad (2.7)$$

The results are striking. The sample mean $\underset{\sim}{X}$ has total squared

prediction error $\sum (X_i - \theta_i)^2$ of 15.86, but $\underset{\sim}{\tilde{\delta}^1}(\underset{\sim}{X}) \equiv (\tilde{\delta}_1^1(\underset{\sim}{X}), \ldots, \tilde{\delta}_k^1(\underset{\sim}{X}))$

has total squared prediction error of only 3.18. Moreover, $\tilde{\delta}_i^1$ is closer

than X_i to θ_i for every batter. This is no fluke: with these data $\tilde{\delta}_i^1$

would be closer to θ_i than X_i for all i = 1, ..., k about 54 percent of

the time, assuming the distribution (2.6) holds and $\mu = -3.075$, $\tau^2 = .0395$,

these last two values being the "true values" estimated as the maximum

likelihood estimate of μ, τ^2 from the data in Table 1 for the season

remainder. The estimates (2.7) are retransformed in Table 3 to give estimates $\hat{p}_i^1 = f_n^{-1}(\hat{\theta}_i)$ of p_i.

Table 3

BATTING AVERAGES AND THEIR ESTIMATES

i	Batting Average for Season Remainder p_i	Maximum Likelihood Estimate Y_i	Retransform of Stein's Estimator $\hat{p}_i^1$	Retransform of $\tilde{\delta}^{0.8}$ $\hat{p}_i^{0.8}$
1	.346	.400	.303	.353
2	.298	.378	.299	.331
3	.276	.356	.294	.310
4	.221	.333	.289	.289
5	.273	.311	.284	.284
6	.270	.311	.284	.284
7	.263	.289	.278	.278
8	.269	.244	.268	.268
9	.264	.222	.263	.263
10	.256	.222	.263	.263
11	.304	.222	.263	.263
12	.264	.222	.263	.263
13	.285	.200	.257	.240
14	.319	.178	.251	.217

Stein's estimators achieve uniformly lower aggregate risk than the MLE but allow the possibility of considerably increased risk to individual components of the vector $\underset{\sim}{\theta}$. As a function of $\underset{\sim}{\theta}$, the risk for estimating θ_1 by $\tilde{\delta}_1^1$, for example, can be as large as $k/4$ times as great as the risk of the MLE X_1. This phenomenon is discussed at length in Efron and Morris (1971 and 1972a), where "limited transla-tion estimators" $\underset{\sim}{\tilde{\delta}}^s(\underset{\sim}{X})$ $0 \leq s \leq 1$ are introduced to reduce this effect. (The MLE corresponds to $s = 0$, Stein's estimator to $s = 1$.) For our situation, the estimate $\tilde{\delta}_i^s(\underset{\sim}{X})$ of θ_i is defined to be as close as pos-sible to $\tilde{\delta}_i^1(\underset{\sim}{X})$ subject to the condition that it not differ from X_i by

more than $\sqrt{\frac{k-1}{k}\frac{k-3}{V}}$ $D_{k-1}(s)$ standard deviations of X_i, $D_{k-1}(s)$ being

a constant taken from Efron and Morris (1972a, Table 1). If $s = 0.8$,

then $D_{13}(s) = 0.761$, so $\widetilde{\delta}_i^{0.8}(\underset{\sim}{X})$ may differ from X_i by no more than

0.761 $(13 \times 0.766/14)^{\frac{1}{2}} = .64$. By limiting translation, $\underset{\sim}{\widetilde{\delta}}^{0.8}$ achieves

the risk averaged over (2.6) of $rs + (1-s) = .8r + .2$, where

$r = 1 - \frac{k-3}{k}\frac{1}{1+\tau^2}$ is the risk of $\widetilde{\delta}_i^1$ averaged over (2.6). This is some-

what larger than the risk r of $\widetilde{\delta}_i^1$, .395 as compared with .244, if we

use $\tau^2 = .0395$ as before. In return, the maximum risk for estimating

θ_i by $\widetilde{\delta}_i^{0.8}$ is 1.44; the maximum risk is 3.50 for $\widetilde{\delta}_i^1$ and 1.00 for $\widetilde{\delta}_i^0$,

the MLE.

The retransformed values $\hat{p}_i^{0.8}$ of the limited translation estimates

$f_n^{-1}(\widetilde{\delta}_i^{0.8}(\underset{\sim}{X}))$ are given in the last column of Table 3, the estimates for

the top three and bottom two batters being affected. Before retrans-

formation, the aggregate prediction error was $\sum(\widetilde{\delta}_i^{0.8}(\underset{\sim}{X}) - \theta_i)^2 = 4.90$,

which is better than 15.86 for the MLE but not as good as 3.18 for $\underset{\sim}{\widetilde{\delta}^1}$.

Like Stein's rule, in this example the limited translation estimates

are better than the MLE for every batter. In other situations, limiting

translation can actually reduce aggregate risk as well as the risk to

unusually large or small θ_i values, as an example of Sec. IV will

indicate.

Clemente (i=1) was known to be an exceptionally good hitter from

his performance in other years. Limiting translation resulted in a

much better estimate for him, as we anticipated, since $\widetilde{\delta}_1^1(\underset{\sim}{X})$ differs

from X_1 by an excessive 1.44 standard deviations of X_1. The maximum

component error occurred for Munson (i=14) with all three estimators.

The Bayesian effect is so strong that this maximum error decreases from

$|X_{14} - \theta_{14}| = 2.1$ to $|\tilde{\delta}_{14}^{0.8}(\underset{\sim}{X}) - \theta_{14}| = 1.5$ to $|\tilde{\delta}_{14}^{1}(\underset{\sim}{X}) - \theta_{14}| = .95$ as s increases from 0 to 1. Therefore, limiting translation in this case increased the worst error.

In Efron and Morris (1973a, Sec. 5) we suggest that $\min(1, \frac{k-1.66}{V})$ is a better estimate of $1/(1+\tau^2)$ than $\min(1, \frac{k-3}{V})$, based on compelling numerical evidence from comparing the two risk functions. This modifies (2.5) to $\bar{X} + (1 - (k - 1.66)/V)(X_i - \bar{X})$, changing (2.7) to $.859 \bar{X} + .141 X_i$, and reducing the total squared prediction error 12 percent from 3.18 to 2.79 for the 14 batters.

The quantities that we previously denoted θ_i are actually season-remainder estimates $\tilde{\theta}_i$ (say) of the long run average θ_i. The variances of θ_i, estimated using the last column of Table 1, sum to 1.57. Therefore any predictor $\hat{\theta}_i$ of $\tilde{\theta}_i$ based on column 1 of Table 1 must satisfy

$$E \sum(\hat{\theta}_i - \tilde{\theta}_i)^2 = E \sum(\hat{\theta}_i - \theta_i)^2 + 1.57, \tag{2.8}$$

so that expected prediction error and risk differ only by a constant. The MLE therefore gives $E \sum(X_i - \tilde{\theta}_i)^2 = 15.57$ so the value of $\sum(X_i - \tilde{\theta}_i)^2 = 15.86$ shows that our data, after 45 at bats, exhibit almost exactly the expected amount of variation from the true values.

III. A GENERALIZATION OF STEIN'S ESTIMATOR TO UNEQUAL VARIANCES
FOR ESTIMATING THE PREVALENCE OF TOXOPLASMOSIS

One of the authors participated in a study of toxoplasmosis in
El Salvador (Remington et al., 1970). Sera obtained from a total sample
of 5171 individuals of varying ages from 36 El Salvador cities were
analyzed by a Sabin-Feldman dye test. From the data given in Remington
et al. (1970, Table 1), toxoplasmosis prevalence rates X_i for city i,
i = 1, ..., 36 were calculated. The prevalence rate X_i has the form
(observed minus expected)/expected, where "observed" is the number of
positives for city i and "expected" is the number of positives for the
same city based on an indirect standardization of prevalence rates to
the age distribution of city i. The variances D_i = $Var(X_i)$ are known
from binomial considerations and differ because of unequal sample sizes.
These data X_i together with the standard deviations $D_i^{\frac{1}{2}}$ are given in
columns 2 and 3 of Table 4. The prevalence rates satisfy a linear con-
straint $\sum d_i X_i = 0$ with known coefficients $d_i > 0$. The means $\theta_i = EX_i$,
which also satisfy $\sum d_i \theta_i = 0$, are to be estimated from the $\{X_i\}$. Since
the $\{X_i\}$ were constructed as sums of independent random variables, they
are approximately normal; and except for the one linear constraint on
the k = 36 values of X_i, they are independent. For simplicity, we will
ignore the slight improvement in the independence approximation that
would result from applying our methods to an appropriate 35-dimensional
subspace. We therefore assume that the $\{X_i\}$ have the distribution of
the following paragraph.

Table 4

ESTIMATES AND EMPIRICAL BAYES ESTIMATES
OF TOXOPLASMOSIS PREVALENCE RATES

i	X_i	$\sqrt{D_i}$	$\delta_i(\underset{\sim}{X})$	$\hat{A}_i$	$\hat{k}_i$	$\hat{B}_i$
1	.293	.304	.035	.0120	1334.1	.882
2	.214	.039	.192	.0108	21.9	.102
3	.185	.047	.159	.0109	24.4	.143
4	.152	.115	.075	.0115	80.2	.509
5	.139	.081	.092	.0112	43.0	.336
6	.128	.061	.100	.0110	30.4	.221
7	.113	.061	.088	.0110	30.4	.221
8	.098	.087	.062	.0113	48.0	.370
9	.093	.049	.079	.0109	25.1	.154
10	.079	.041	.070	.0109	22.5	.112
11	.063	.071	.045	.0111	36.0	.279
12	.052	.048	.044	.0109	24.8	.148
13	.035	.056	.028	.0110	28.0	.192
14	.027	.040	.024	.0108	22.2	.107
15	.024	.049	.020	.0109	25.1	.154
16	.024	.039	.022	.0108	21.9	.102
17	.014	.043	.012	.0109	23.1	.122
18	.004	.085	.003	.0112	46.2	.359
19	-.016	.128	-.007	.0116	101.5	.564
20	-.028	.091	-.017	.0113	51.6	.392
21	-.034	.073	-.024	.0111	37.3	.291
22	-.040	.049	-.034	.0109	25.1	.154
23	-.055	.058	-.044	.0110	28.9	.204
24	-.083	.070	-.060	.0111	35.4	.273
25	-.098	.068	-.072	.0111	34.2	.262
26	-.100	.049	-.085	.0109	25.1	.154
27	-.112	.059	-.089	.0110	29.4	.210
28	-.138	.063	-.106	.0110	31.4	.233
29	-.156	.077	-.107	.0112	40.0	.314
30	-.169	.073	-.120	.0111	37.3	.291
31	-.241	.106	-.128	.0114	68.0	.468
32	-.294	.179	-.083	.0118	242.4	.719
33	-.296	.064	-.225	.0111	31.9	.238
34	-.324	.152	-.114	.0117	154.8	.647
35	-.397	.158	-.133	.0117	171.5	.665
36	-.665	.216	-.140	.0119	426.8	.789

In order to obtain an appropriate empirical Bayes estimation rule
for this data we assume that

$$X_i | \theta_i \overset{ind}{\sim} N(\theta_i, D_i) \quad i = 1, \ldots, k \tag{3.1}$$

and

$$\theta_i \overset{ind}{\sim} N(0, A) \quad i = 1, \ldots, k, \tag{3.2}$$

A being an unknown constant. These assumptions are the same as (1.1),
(1.7), which lead to the James-Stein estimator, except that we have
allowed each X_i to have a different sampling variance D_i. Notice that
the choice of *a priori* mean zero for the θ_i is particularly appropriate
here because the constant $\sum d_i \theta_i = 0$ forces the parameters to be centered
near the origin.

We require $k \geq 3$ in the following derivations. Define

$$B_i \equiv D_i / (A + D_i). \tag{3.3}$$

Then (3.1) and (3.2) are equivalent to

$$\theta_i | X_i \overset{ind}{\sim} N((1 - B_i)X_i, D_i(1 - B_i)) \quad i = 1, \ldots, k \tag{3.4}$$

and, marginally,

$$X_i \overset{ind}{\sim} N(0, A + D_i) \quad i = 1, \ldots, k. \tag{3.5}$$

From (3.4) it follows that for the loss function

$$L(\theta_i, a_i) = (\theta_i - a_i)^2 \tag{3.6}$$

(or for any other increasing function of $|\theta_i - a_i|$) the Bayes estimator is the *a posteriori* mean

$$\delta_i^*(X_i) = E\theta_i|X_i = (1 - B_i)X_i. \tag{3.7}$$

The Bayes risk is obtained from the variance in (3.4)

$$R_i(A, \delta_i^*) \equiv E_A E_\theta L(\theta_i, \delta_i^*(X_i)) = (1 - B_i)D_i \tag{3.8}$$

and is to be compared with the risk of the maximum likelihood estimator $\delta_i^0(X_i) = X_i$

$$R_i(A, \delta_i^0) = E_A E_\theta L(\theta_i, \delta_i^0(X_i)) = D_i. \tag{3.9}$$

In the empirical Bayes context A must be estimated. Denote $\underset{\sim}{X} = (X_1, \ldots, X_k)$, $S_j = X_j^2$, $j = 1, \ldots, k$, and $\underset{\sim}{S} = (S_1, \ldots, S_k)$. We have

$$S_j \overset{ind}{\sim} (A + D_j)\chi_1^2 \quad j = 1, \ldots, k. \tag{3.10}$$

The maximum likelihood estimator $\hat{A}$ of A from the distribution (3.10) is the solution to the equation

$$\hat{A} = \sum_{j=1}^{k} (S_j - D_j)I_j(\hat{A}) / \sum_{j=1}^{k} I_j(\hat{A}) \tag{3.11}$$

with

$$I_j(A) \equiv 1/Var(S_j) = 1/[2(A + D_j)^2] \tag{3.12}$$

being the Fisher information for A in S_j. We could use $\hat{A}$ from (3.11) to define the empirical Bayes estimator of θ_i as $(1 - D_i/(\hat{A} + D_i))X_i$. However, this rule does not reduce to Stein's when all D_j are equal, and we will instead use a minor variant of this estimator derived in Efron and Morris (1973a, Sec. 8), which does reduce to Stein's. The difference between the rules is minor in this case, but it might be more important were k smaller. The proposed rule is described as follows.

The preferred estimate of θ_i is

$$\delta_i(\underset{\sim}{X}) \equiv (1 - \hat{B}_i(\underset{\sim}{S}))X_i \tag{3.13}$$

where

$$\hat{B}_i(\underset{\sim}{S}) = \hat{c}_i D_i/(\hat{A}_i + D_i) \tag{3.14}$$

with

$$\hat{c}_i = \max((\hat{k}_i - 2)/(\hat{k}_i + 2), 0) \tag{3.15}$$

and

$$\hat{k}_i = \sum_{j=1}^{k} (\hat{A}_i + D_i)^2/(\hat{A}_i + D_j)^2. \tag{3.16}$$

The estimate $\hat{A}_i$ of A, which we allow to depend on i, the index of the component being estimated, is the solution to

$$\hat{A}_i = \sum_{j=1}^{k} a_j I_j(\hat{A}_i)/\sum_{j=1}^{k} I_j(\hat{A}_i) \tag{3.17}$$

with

$$a_j \equiv S_j - D_j \qquad \text{if } j \neq i$$

$$a_i = (S_i - 3D_i)/3 \quad \text{if } j = i. \qquad (3.18)$$

The functions $\{I_j\}$ are Fisher informations, defined as

$$I_j(A) = 1/[2(A + D_j)^2] \quad \text{if } j \neq i$$

$$I_i(A) = 3/[2(A + D_i)^2] \quad \text{if } j = i. \qquad (3.19)$$

The solution to (3.17) is a maximum likelihood estimate of A based on the distributions (3.10), except that when $j = i$ we count S_i three times in the estimation of A by assuming $S_i \sim (A + D_i)\chi_3^2$, a modification justified in Efron and Morris (1973a, Secs. 2, 8). The $\{a_j\}$ of (3.18) are then the independent unbiased coordinate estimates of A, and the right-hand side of (3.17) gives an estimated minimum variance linear unbiased estimate of A from the $\{a_j\}$. The only difference between (3.17) and (3.11) is that S_i is counted three times in (3.17) and once in (3.11). Formula (3.17) is solved iteratively and, in our experience, converges rapidly.

The value of $\hat{k}_i$ in (3.16) is an estimate of the equivalent number of components needed to give the actual Fisher information $\sum I_j(A)$ for A if all components had the same variance D_i as the ith. Then $\hat{c}_i$ in (3.15) is the approximate best multiple of $D_i/(\hat{A}_i + D_i)$ for estimating B_i in the sense of Efron and Morris (1973a, Sec. 8). We permit solutions $\hat{A}_i$ of (3.17) to be negative, but were $\hat{A}_i$ so negative that $\hat{B}_i(\underset{\sim}{S}) > 1$, we would use $\hat{B}_i(\underset{\sim}{S}) = 1$ instead. Therefore we require

$\hat{A}_i \geq -(1 - \hat{c}_i)D_i$, although this consideration does not arise with these data. In the case $D_i = D$ for all i, it follows that $\hat{B}_i(S) = (k - 2)D/\sum S_j$ so (3.13) reduces to the James-Stein estimator.

Our estimates $\delta_i(\underset{\sim}{X})$ of the θ_i are given in the fourth column of Table 4 and are compared with the unbiased estimate X_i in Fig. 1. Figure 1 illustrates the "pull in" effect of $\delta_i(\underset{\sim}{X})$, which is most pronounced for cities 1, 32, 34, 35, and 36. Under the empirical Bayes model, the major explanation for the large $|X_i|$ for these cities is through their large D_i rather than through large $|\theta_i|$. This figure also shows that the rankings of the cities on the basis of $\delta_i(\underset{\sim}{X})$ differs from that based on the X_i, a feature that does not arise in the case when the X_i have equal variances.
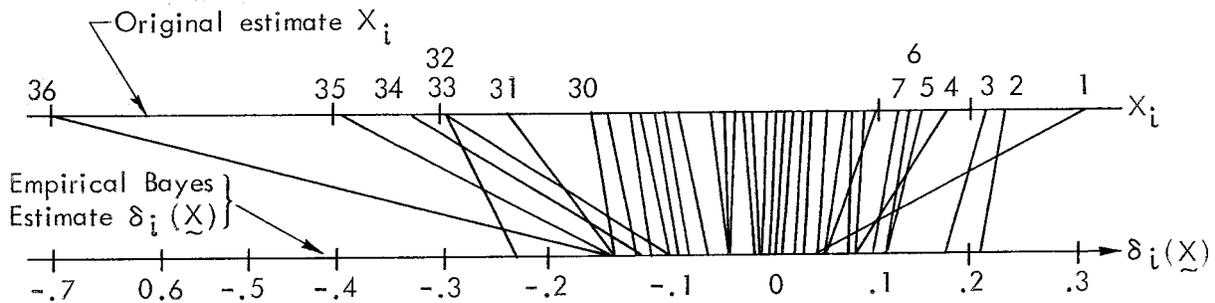

Fig. 1 — Estimates of Toxoplasmosis prevalence rates

The values $\hat{A}_i$, $\hat{k}_i$, and $\hat{B}_i(\underset{\sim}{S})$ defined in (3.17), (3.16), and (3.14) are given in the last three columns of Table 4. The value $\hat{A}$ of (3.11) is $\hat{A} = 0.0122$ with standard deviation $\sigma(\hat{A})$ estimated as 0.0041 (if A = 0.0122) by the Cramér-Rao lower bound on $\sigma(\hat{A})$. The preferred estimates $\hat{A}_i$ are all close to but slightly smaller than $\hat{A}$, and their

estimated standard deviations vary from 0.00358 for the cities with

the smallest D_i to 0.00404 for the city with the largest D_i.

The likelihood function of the data plotted as a function of A

(on a log scale) is given in Fig. 2 and Fig. 3 as LIKELIHOOD. The

curves are normalized to have unit area as a function of $\alpha = \log A$.

The maximum value of this function of α is at $\hat{\alpha} = \log(\hat{A}) = \log(.0122)$

$= -4.40 \equiv \mu_\alpha$. The curves are almost perfectly normal with mean

$\hat{\alpha} = -4.40$ and standard deviation $\sigma_\alpha \equiv .371$ (obtained from numerical in-

tegration). The likely values of A therefore correspond to α differ-

ing from μ_α by no more than three standard deviations, $|\alpha - \mu_\alpha| \leq 3\sigma_\alpha$,

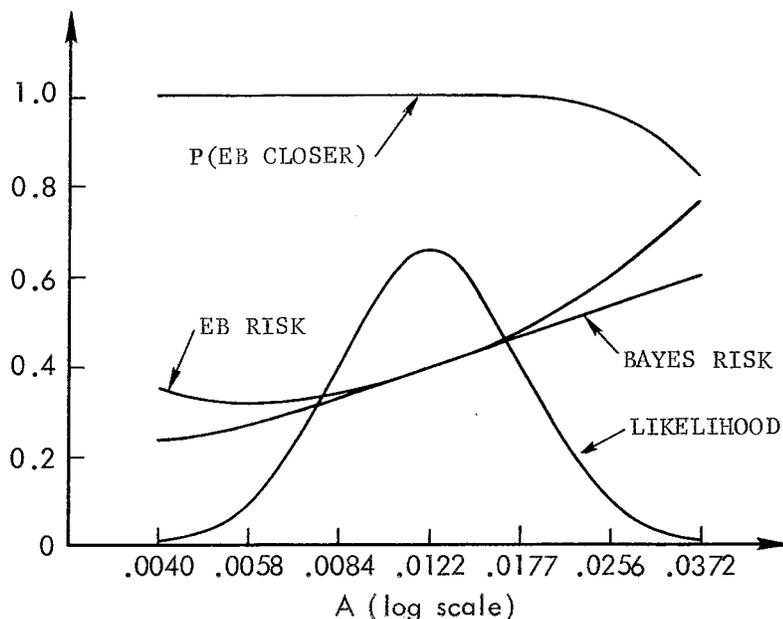or, equivalently, $.0040 \leq A \leq .0372$.

Fig. 2 - Likelihood function of A and aggregate operating
characteristics of estimates as a function of A,
conditional on the observed toxoplasmosis data

In the region of likely values of A, Fig. 2 also graphs two risks: BAYES RISK and EB RISK (for empirical Bayes risk), each conditional on the data $\underset{\sim}{X}$. EB RISK is the conditional risk of the empirical Bayes rule defined (with $D_0 \equiv \frac{1}{k} \sum_{i=1}^{k} D_i$) as

$$E_A \frac{1}{kD_0} \sum_{i=1}^{k} (\delta_i(\underset{\sim}{X}) - \theta_i)^2 |\underset{\sim}{X}, \qquad (3.20)$$

and BAYES RISK is

$$E_A \frac{1}{kD_0} \sum_{i=1}^{k} (\frac{A}{A+D_i} X_i - \theta_i)^2 |\underset{\sim}{X}. \qquad (3.21)$$

(N.B.: In (3.20) the $\delta_i(\underset{\sim}{X})$ are fixed numbers--those given in Table 4. The expectation is over the *a posteriori* distribution (3.4) of the θ_i.) Since A is not known, BAYES RISK yields only a lower envelope for empirical Bayes estimators, agreeing with EB RISK at A = .0122. Table 5 gives values to supplement Fig. 2. Not graphed because it is too large to fit in Fig. 2 is MLE RISK, the conditional risk of the maximum likelihood estimator, defined as

$$E_A \frac{1}{kD_0} \sum_{i=1}^{k} (X_i - \theta_i)^2 |\underset{\sim}{X}. \qquad (3.22)$$

MLE RISK exceeds EB RISK by factors varying from 7 to 2 in the region of likely values of A, as shown in Table 5. EB RISK tends to increase and MLE RISK to decrease as A increases, these values crossing at A = .0650, about $4\frac{1}{2}$ standard deviations above the mean of the distribution of $\hat{A}$. Our point can be stated this way: The data suggest that

almost certainly A is in the interval $.004 \leq A \leq .037$, and for all

such values of A the numbers $\delta_i(\underset{\sim}{X})$ are much better estimators of the

θ_i than are the X_i. (There is also a non-Bayesian version of this

statement based on a confidence interval for $\sum \theta_i^2/k$.)

Table 5

CONDITIONAL RISKS

A	.0040	.0122	.0372	.0650	∞
EB RISK	.35	.39	.76	1.08	2.50
MLE RISK	2.51	1.87	1.27	1.08	1.00
P(EB CLOSER)	1.00	1.00	.82	.50	.04

The remaining curve in Fig. 2 graphs the probability that the

empirical Bayes estimator is closer to $\underset{\sim}{\theta}$ than the MLE $\underset{\sim}{X}$, conditional

on the data $\underset{\sim}{X}$. It is defined as

$$P_A[\sum (\delta_i(\underset{\sim}{X}) - \theta_i)^2 < \sum (X_i - \theta_i)^2 | \underset{\sim}{X}]. \qquad (3.23)$$

This curve, denoted P(EB CLOSER), decreases as A increases, but is

always very close to unity in the region of likely values of A. It

reaches one-half at about $4\frac{1}{2}$ standard deviations from the mean of the

likelihood function and then decreases as $A \to \infty$ to its asymptotic

value .04 (see Table 5).

The "conditional relative savings loss" RSL_i for component i is

defined as

$$RSL_i = \frac{E_A\{(\delta_i(\underset{\sim}{X}) - \theta_i)^2 - (\delta_i^*(X_i) - \theta_i)^2\} | \underset{\sim}{X}}{E_A\{(X_i - \theta_i)^2 - (\delta_i^*(X_i) - \theta_i)^2\} | \underset{\sim}{X}}. \qquad (3.24)$$

This compares the risk of δ_i with that of the MLE in terms of the risk of the Bayes rule δ_i^*, conditional on the data $\underset{\sim}{X}$. Using the distribution (3.4) this is easily shown to reduce to

$$ RSL_i = \left(\frac{\hat{B}_i(\underset{\sim}{S})}{B_i} - 1 \right)^2 = \left(\frac{A - \hat{A}}{D_i + \hat{A}} \right)^2 . \tag{3.25} $$

This quantity is plotted in Fig. 3 for the case $D_i = (.071)^2$, a central value of D_i. Obviously $\delta_i(\underset{\sim}{X})$ is better than X_i for those values of A for which $RSL_i < 1$, a region that includes most of the likely A values. From (3.25), larger values of D_i would give lower RSL_i curves while smaller D_i give higher curves. For any D_i, if $A \le 2\hat{A} + D_i$, then $RSL_i \le 1$ and $\delta_i(\underset{\sim}{X})$ is better than X_i.
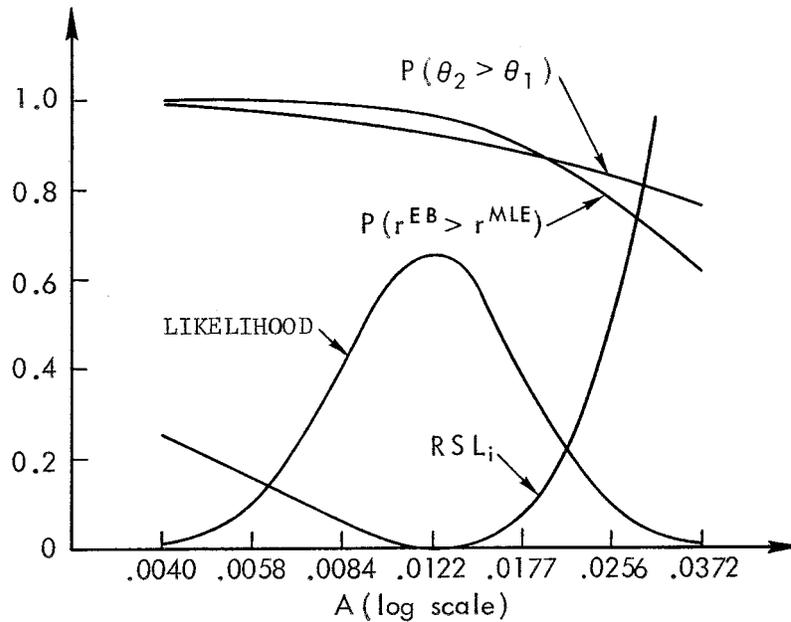

Fig. 3 - Likelihood function of A and individual and ordering characteristics of estimates as a function of A, conditional on the observed toxoplasmosis data.

Figure 1 illustrated that the MLE and the empirical Bayes esti-
mators order the $\{\theta_i\}$ differently. Define the correlation of an esti-
mator $\hat{\underset{\sim}{\theta}}$ of $\underset{\sim}{\theta}$ by

$$r(\hat{\underset{\sim}{\theta}}, \underset{\sim}{\theta}) = \frac{\sum \hat{\theta}_i \theta_i}{(\sum \hat{\theta}_i^2 \sum \theta_i^2)^{\frac{1}{2}}} \qquad (3.26)$$

as a measure of how well $\hat{\underset{\sim}{\theta}}$ orders $\underset{\sim}{\theta}$. We denote $P(r^{EB} > r^{MLE})$ as the
probability that the empirical Bayes estimate $\underset{\sim}{\delta}$ orders $\underset{\sim}{\theta}$ better than
$\underset{\sim}{X}$, i.e., as

$$P_A(r(\underset{\sim}{\delta}, \underset{\sim}{\theta}) > r(\underset{\sim}{X}, \underset{\sim}{\theta}) | \underset{\sim}{X}). \qquad (3.27)$$

The graph of (3.27) given in Fig. 3 shows that $P(r^{EB} > r^{MLE}) > .5$ for
$A \leq .0372$. The value at $A = \infty$ drops to .046.

Although $X_1 > X_2$, the empirical Bayes estimator for city 2 is
larger, $\delta_2(\underset{\sim}{X}) > \delta_1(\underset{\sim}{X})$. This is because $D_1 \gg D_2$, indicating that X_1
is large under the empirical Bayes model because of randomness while
X_2 is large because θ_2 is large. The remaining curve in Fig. 3 is

$$P_A(\theta_2 > \theta_1 | \underset{\sim}{X}) \qquad (3.28)$$

and shows that $\theta_2 > \theta_1$ is quite probable for likely values of A. This
probability declines as $A \to \infty$, being .50 at $A = .24$ (eight standard
deviations above the mean) and .40 at $A = \infty$.

A simpler solution to the unequal variances problem is to apply
Stein's rule (1.4) to $X_i/\sqrt{D_i}$ to estimate $\theta_i/\sqrt{D_i}$. Retransforming leads
to

$$\hat{\theta}_i = (1-B)X_i, \qquad B \equiv (k-2)/\sum_j X_j^2/D_j. \qquad (3.29)$$

This method is unsatisfactory because the implied *a priori* variance of θ_i is proportional to D_i, although the prior distribution should be independent of the number of observations. Furthermore, (3.29) uses an equal shrinkage factor 1-B for components with much and little sample information, violating the principle that the sample mean X_i can be trusted more as sample sizes increase. Finally, (3.29) is constrained to order the $\{\theta_i\}$ as the $\{X_i\}$. The formulation (3.2) avoids these deficiencies.

IV. IMPROVING A COMPUTER SIMULATION BY USING UNIVARIATE AND MULTIVARIATE EMPIRICAL BAYES ESTIMATORS

A Monte Carlo experiment is given below in which several forms of Stein's method all substantially improve the experimental precision. The example is realistic in that the normality and variance assumptions made are only approximations to the true situation.

We chose to investigate Pearson's chi-square statistic for its independent interest and selected the particular parameters $(m \leq 24)$ from our prior belief that empirical Bayes methods would be effective for these situations. Although our beliefs were substantiated, the outcomes in this instance did not always favor our pet methods.

In addition to demonstrating the effectiveness of empirical Bayes methods, this section illustrates several new rules: a modified rule that determines how much two groups are to be combined, Stein's rule after first fitting a linear model to the data, and a multivariate empirical Bayes rule. Some readers will also be interested in the results given for Pearson's chi-square test, summarized in Fig. 7.

The simulation was conducted to estimate the exact size of Pearson's chi-square test. Let Y_1 and Y_2 be independent binomial random variables, $Y_1 \sim \text{bin}(m, p')$, $Y_2 \sim \text{bin}(m, p'')$ so $EY_1 = mp'$, $EY_2 = mp''$. Pearson advocated the statistic and critical region

$$T = \frac{2m(Y_1 - Y_2)^2}{(Y_1 + Y_2)(2m - Y_1 - Y_2)} > 3.84 \qquad (4.1)$$

to test the composite null hypothesis H_0: $p' = p''$ against all alternatives for the nominal size $\alpha = 0.05$. The value 3.84 is the 95th

percentile of the chi-square distribution with one degree of freedom, which asymptotically approximates that of T when m is large.

The true size of the test under H_0 is defined as

$$\alpha(p, m) \equiv P(T > 3.84 | p, m), \qquad (4.2)$$

which depends on both m and the unknown value $p \equiv p' = p''$. The simulation was conducted for all combinations of the r = 3 values of p, $p_1 = 0.5$, $p_2 = 0.3$, $p_3 = 0.1$ and the k = 17 values of m with $m_j = 7 + j$, j = 1, ..., k. The rk = 51 values of $\alpha_{ij} \equiv \alpha(p_i, m_j)$ were to be estimated. For each i, j we simulated (4.1) n = 500 times on a computer and recorded Z_{ij} as the proportion of times H_0 was rejected. The data appear in Table 9 at the section's end. Since $nZ_{ij} \sim \text{bin}(n, \alpha_{ij})$ independently, Z_{ij} is the unbiased and maximum likelihood estimator usually chosen to estimate α_{ij}.

We will ignore an extensive bibliography of other methods for improving computer simulations. Empirical Bayes methods can be applied simultaneously with other methods, and if better estimates of α_{ij} than Z_{ij} were available then the empirical Bayes methods could instead be applied to them. But for simplicity we take Z_{ij} itself as the quantity to be improved.

Under H_0 the standard deviation of Z_{ij} is $\{\alpha_{ij}(1 - \alpha_{ij})/n\}^{\frac{1}{2}}$, which is approximately $\sigma = \{(.05)(.95)/500\}^{\frac{1}{2}} = .009747$. The variables $X_{ij} \equiv (Z_{ij} - .05)/\sigma$ have expectations $\theta_{ij} \equiv EX_{ij} = (\alpha_{ij} - .05)/\sigma$ and, approximately, the distribution

$$X_{ij} | \theta_{ij} \overset{\text{ind}}{\sim} N(\theta_{ij}, 1) \quad i = 1, 2, 3 = r, \quad j = 1, 2, ..., 17 = k,$$
$$(4.3)$$

described in earlier sections.

The data Z_{1j}, $j = 1, \ldots, k$ plotted in Fig. 4 pertain to the case $p = p_1 = 0.5$ only. The average value $\bar{Z}_1 = .051$ of the 17 points supports the choice of the "natural origin" $\bar{\alpha}_1 = .05$, and the slope of the plotted regression line differs insignificantly from zero. Stein's rule (1.4) applied to the transformed data (4.3) and then retransformed according to $\hat{\alpha}_{1j} = .05 + \sigma\hat{\theta}_{1j}$ yields

$$\hat{\alpha}_{1j} = (1 - \hat{B})Z_{1j} + .05 \hat{B}, \quad \hat{B} = .325, \tag{4.4}$$

where $\hat{B} \equiv (k-2)/S$ and $S \equiv \sum_{j=1}^{17}(Z_{1j} - .05)^2/\sigma^2 = 46.15$.

All 51 true values α_{ij} were obtained exactly through a separate computer program and appear in Fig. 7 and Table 9. The loss function, taken to be the normalized sum of squared errors $\sum(\hat{\alpha}_{1j} - \alpha_{1j})^2/\sigma^2$, can therefore be evaluated. The MLE has loss 18.9, Stein's estimate (4.4) has loss 10.2, and the constant estimator, which always estimates α_{1j} as .05, has loss 23.4. Hence, Stein's rule dominates both extremes between which it compromises.

Figure 5 displays the maximum likelihood estimates, Stein estimates, and true values. The true values show a surprising periodicity, which would frustrate attempts at improving the MLE by smoothing.

The MLE is closer to the true value than the empirical Bayes rule in four cases: $m = 12, 24, 10, 8$, and further in 12 cases (there was one tie). The second column of Table 6 gives the difference of errors $\{|Z_{1j} - \alpha_{1j}| - |\hat{\alpha}_{1j} - \alpha_{1j}|\}/\sigma$ in standard error units and listed in ascending order. Clearly the empirical Bayes rule is more accurate in more cases, and the improvements tend to be quite substantial.
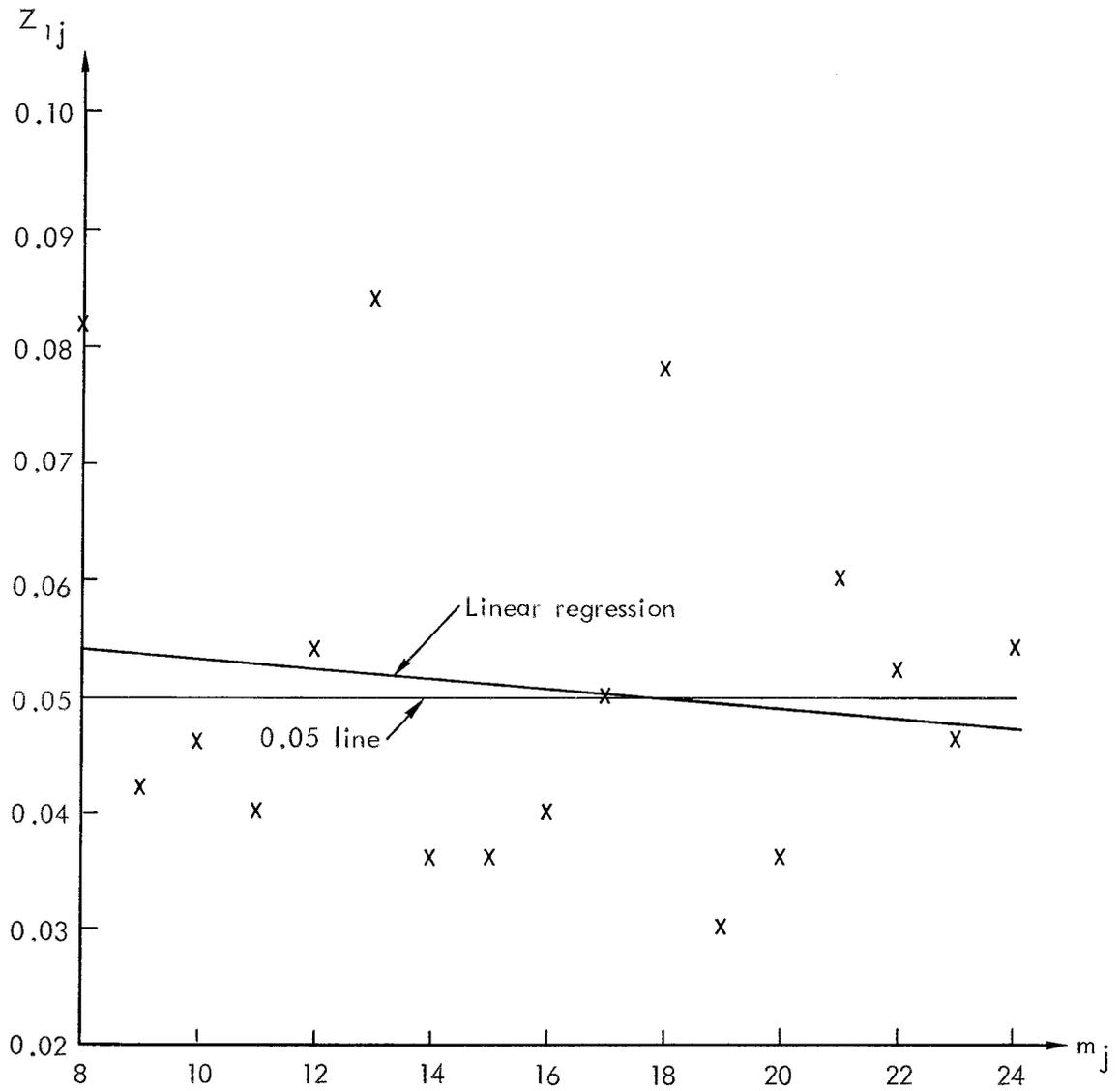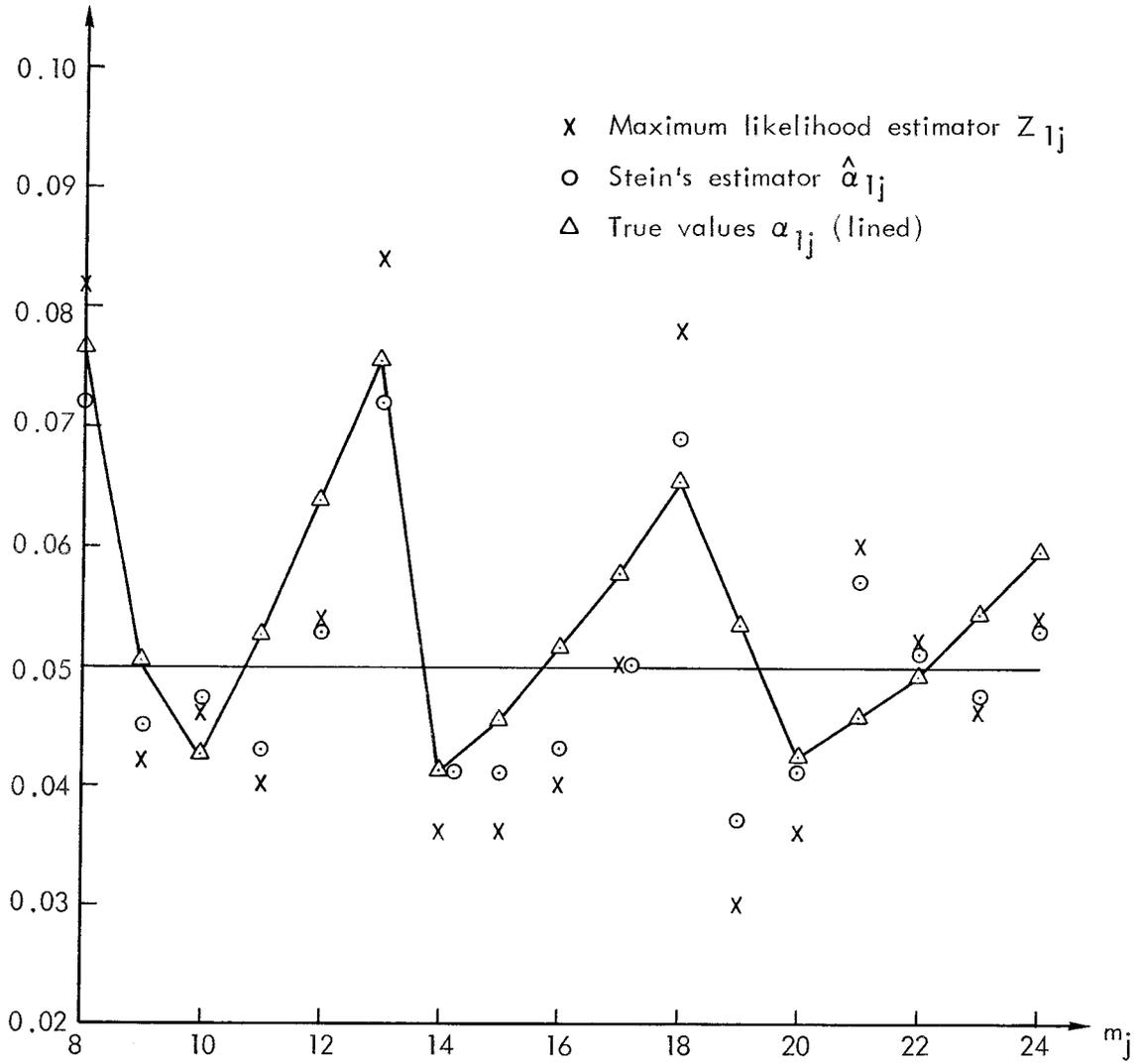
Fig. 4—Maximum likelihood estimates for p = 0.5

Fig. 5—MLE, Stein estimates, and true values for p = 0.5

Table 6

MLE ERRORS COMPARED WITH EMPIRICAL BAYES ERRORS

m	$|z_{1j} - \alpha_{1j}|/\sigma$ minus $|\hat{\alpha}_{1j} - \alpha_{1j}|/\sigma$	m	Empirical Bayes Errors $\hat{\alpha}_{1j} - \alpha_{1j}$
12	-.133	19	-.0166
24	-.133	12	-.0113
10	-.133	11	-.0095
8	-.003	16	-.0082
17	.000	17	-.0077
22	.067	23	-.0069
23	.133	24	-.0068
9	.267	9	-.0055
11	.333	8	-.0052
16	.333	15	-.0050
21	.333	13	-.0027
15	.467	20	-.0019
20	.467	14	-.0004
14	.467	22	.0023
13	.598	18	.0036
19	.667	10	.0051
18	.934	21	.0108

In cases like this, with k fairly large and $\hat{B}$ not too close to unity, rough confidence ellipsoids and intervals for the $\hat{\alpha}_{1j}$ may be constructed by appeal to the Bayesian interpretation of Stein's estimator. Assuming the prior distribution $\alpha_{1j} \sim N(.05, \tau^2)$, the posterior distribution is $\alpha_{1j}|z_{1j} \sim N((1-B)z_{1j} + .05B, \sigma^2(1-B))$ independently, where $B \equiv \sigma^2/(\sigma^2 + \tau^2)$ depends on the unknown parameter τ^2. Were B known, probability ellipsoids for $\{\alpha_{1j}\}$ could be obtained from $\sum_{j=1}^{17}(\alpha_{1j} - (1-B)z_{1j} - .05B)^2 \sim \sigma^2(1-B)\chi_k^2$. Using the estimate (4.4) $\hat{B}$ = .325 of B in place of B, and the fact that the upper .8 significance level for χ_{17}^2 equals 21.6, suggests $\sum(\alpha_{1j} - \hat{\alpha}_{1j})^2/\sigma^2 \le 21.6 (1-\hat{B}) = 14.6$ with probability about 0.8. As stated before, the actual value of the loss $\sum(\alpha_{1j} - \hat{\alpha}_{1j})^2/\sigma^2$ is 10.2, well within the claimed bound.

For known B, Bayesian probability intervals for an individual α_{1j} would be $(1-B)Z_{1j} + .05B \pm z\sigma\sqrt{1-B}$ with z chosen from a table of the normal distribution to give the desired probability. While using $\hat{B}$ in place of B in the preceding equation gives a satisfactory result in this case, it is generally dangerous to use $\hat{B}$, and instead we prefer first to develop a confidence interval for B. An 80 percent confidence interval for B, based on $BS \sim \chi^2_{17}$, is $10.1/S \leq B \leq 24.8/S$ or $.22 \leq B \leq .54$. (This is obtained by excluding the upper and lower ten percentiles of the χ^2_{17} distribution.) The true B is unknown, but $\sigma^2/(\sigma^2 + \sum(\alpha_{1j} - .05)^2/k) = .42$ is being estimated by $\hat{B}$ and lies in the interval just given. Then $.68 \leq \sqrt{1-B} \leq .88$, and it is safer to use the upper bound $.88$ than $\sqrt{1-\hat{B}} = .82$ to establish the widths of the intervals above. A conservative probability interval for α_{1j}, which should contain the true value in at least 68 percent of all cases, is therefore $\hat{\alpha}_{1j} \pm .0086$ ($z = 1$, $\sigma = .00947$). From the fourth column of Table 6, it actually happened that $|\alpha_{1j} - \hat{\alpha}_{1j}| \leq .0086$ in $13/17 = .76$ of these cases.

The preceding method is suspect if the upper end of the confidence interval for B is not somewhat less than unity, since by definition we know that $B \leq 1$. We will not attempt to find confidence intervals for the ensuing cases $p = 0.3$ and $p = 0.1$ because $\hat{B} = 1$ both times.

The average value of Z_{1j}, being $\bar{Z}_1 = .0509$, differs from the average $\bar{\alpha}_1 = .0545$ of α_{1j} by $.0036$, or $\sqrt{17}(.0036)/\sigma = 1.52$ standard deviations of $\bar{Z}_1$. Thus $\bar{Z}_1$ differs significantly from its mean in a way that gives unwarranted support to the choice of origin $.05$. In spite of this unlikely hazard, the rule (4.4) performs well.

On theoretical grounds we know that the approximation $\alpha(p, m) = .05$ improves as m increases, which suggests dividing the data for $p = 0.5$ into two groups, say $8 \leq m \leq 16$ and $17 \leq m \leq 24$. In the Bayesian framework of Efron and Morris (1973b), this disaggregation reflects the concern that A_1, the expectation of $A_1^* \equiv \sum_{j=1}^{9} (\alpha_{1j} - .05)^2/9\sigma^2$ may be much larger than A_2, the expectation of $A_2^* = \sum_{j=10}^{17} (\alpha_{1j} - .05)^2/8\sigma^2$, or equivalently that the pull-in factor $B_1 = 1/(1 + A_1)$ for group 1 really should be smaller than $B_2 = 1/(1 + A_2)$ for group 2.

The combined estimator (4.4), having $\hat{B}_1 = \hat{B}_2$, is given in the first row of Table 7 with loss components for each group. The simplest way to utilize separate estimates of B_1 and B_2 is to apply two separate Stein rules, as shown in the second row of the table. As in the baseball section we can also use the bolder estimate $\hat{B}_i = (k_i - .66)/S_i$, $S_1 \equiv \sum_{j=1}^{9} (Z_{1j} - .05)^2/\sigma^2$, $S_2 \equiv S - S_1$, $k_1 = 9$, $k_2 = 8$. The constant $k_i - .66$ is preferred to the usual $k_i - 2$ because it gives smaller risks unless the true values of B_1 and B_2 are near zero, without costing much otherwise (Efron and Morris, 1973a). The third row of Table 7 shows the effectiveness of this choice.

In Efron and Morris (1973b) we designed rules to compromise between separate Stein estimates for two groups and a single Stein estimate for the combined group for situations where the assumption $A_1 = A_2$ is uncertain. The suggested rules put distributions on B_1/B_2 and estimate the ratio B_1/B_2 from the ratio S_2/S_1 to choose a compromise between the separate and combined Stein rules. The values $\hat{B}_1$ and $\hat{B}_2$ in the last row of Table 7 were determined from equations (3.28) of Efron and Morris (1973b) using the mass function for $\delta_{\sim 1}^{(3)}$ given in Table 4 of the same

Table 7

VALUES OF $\hat{B}$ AND LOSSES FOR p = 0.5 DATA SEPARATED INTO
TWO GROUPS, VARIOUS ESTIMATION RULES

	$8 \leq m \leq 16$ $\hat{B}_1$	Group 1 Loss	$17 \leq m \leq 24$ $\hat{B}_2$	Group 2 Loss	Total Loss	Expected Value of Total Loss (from (1.11))
Stein's rule, combined data	.325	4.2	.325	6.0	10.2	10.5
Separate Stein rules	.232	4.5	.376	5.4	9.9	10.9
Separate Stein rules, bigger constant	.276	4.3	.460	4.6	8.9	
Two groups rule	.306	4.3	.362	5.5	9.8	

paper. The estimation rule has the form (4.4) with the appropriate
$\hat{B}_1$ and $\hat{B}_2$. The two groups estimator will generally compromise quite
favorably between the separate and combined Stein estimators, although
the improvement here is only slight, partly because the latter two
estimators perform equally well on this set of data. The larger shrink-
ing constants $k_i - .66$ could have been used with the two groups rule,
and this would have improved the two groups precision to approximately
that of the third row of Table 7. Were we certain $A_1 \geq A_2$, a corre-
sponding prior distribution could be used to further improve precision,
as described for $\underset{\sim}{\delta}_1^{(5)}$ of Table 4 in Efron and Morris (1973b).

The actual values are $A_1^* = \sum_{j=1}^9 (\alpha_{1j} - .05)^2 / 9\sigma^2 = 2.036$ for group
1 and $A_2^* = \sum_{j=10}^{17} (\alpha_{1j} - .05)^2 / 8\sigma^2 = .635$, so $B_1^* \equiv 1/(1 + A_1^*) = .329$,
$B_2^* = 1/(1 + A_2^*) = .612$, and $B_2^*/B_1^* = 1.86$. These values were used in
(1.11) to obtain the last column of Table 7.

The data for $p = p_3 = 0.1$ present a different challenge because a plot of the MLE values Z_{3j}, Fig. 6, shows that .05 is clearly an unsatisfactory origin. Instead, the figure suggests that the true values may nearly follow the linear relationship $\alpha_{3j} = \beta_0 + \beta_1(m_j - 16)$. A standard regression of Z_{3j} on m_j gives $\hat{\beta}_0 = \bar{Z}_3 = .0295$, $\hat{\beta}_1 = .00218$. The estimated line is plotted in Fig. 6.

It is easy to combine linear model fitting with Stein estimation procedures. We use Stein's rule to improve estimation of the residuals $\tilde{Z}_{3j} \equiv Z_{3j} - \hat{\beta}_0 - \hat{\beta}_1(m_j - 16)$, and then estimate each α_{3j} as the sum of its regression estimate and its estimated residual:

$$\hat{\alpha}_{3j} = \hat{\beta}_0 + \hat{\beta}_1(m_j - 16) + (1 - \hat{B})\tilde{Z}_{3j},$$

$$(4.5)$$

$$\hat{B} \equiv \min(1, \ (k-2-d)\sigma^2/\textstyle\sum\tilde{Z}_{3j}^2).$$

(In (4.5) $d = 2$ is the degrees of freedom lost by fitting the linear regression.) This estimator can be written as a convex combination of the estimated prior mean and the maximum likelihood estimator, $\hat{\alpha}_{3j} = \hat{B} \cdot (\text{regression estimator}) + (1 - \hat{B}) \cdot \text{MLE}$. For these data, $(k - 4)\sigma^2/\sum\tilde{Z}_{3j}^2 = 1.93 > 1$, so $\hat{B} = 1$ and (4.5) reduces to $\hat{\alpha}_{3j} = .0295 + .00218(m_j - 16)$. The true values α_{3j}, given in Figs. 6 and 7 and Table 9, follow the estimated line closely, and the loss $\sum(\hat{\alpha}_{3j} - \alpha_{3j})^2/\sigma^2$ is only 3.1 for the rule (4.5). This would have been a better illustration had $\hat{B} < 1$ occurred, as it would in about half of all future simulations with these parameters $\{\alpha_{3j}\}$. Then (4.5), which is mainly designed to improve the MLE, would not reduce to the linear regression estimate and give the false impression that the special features of
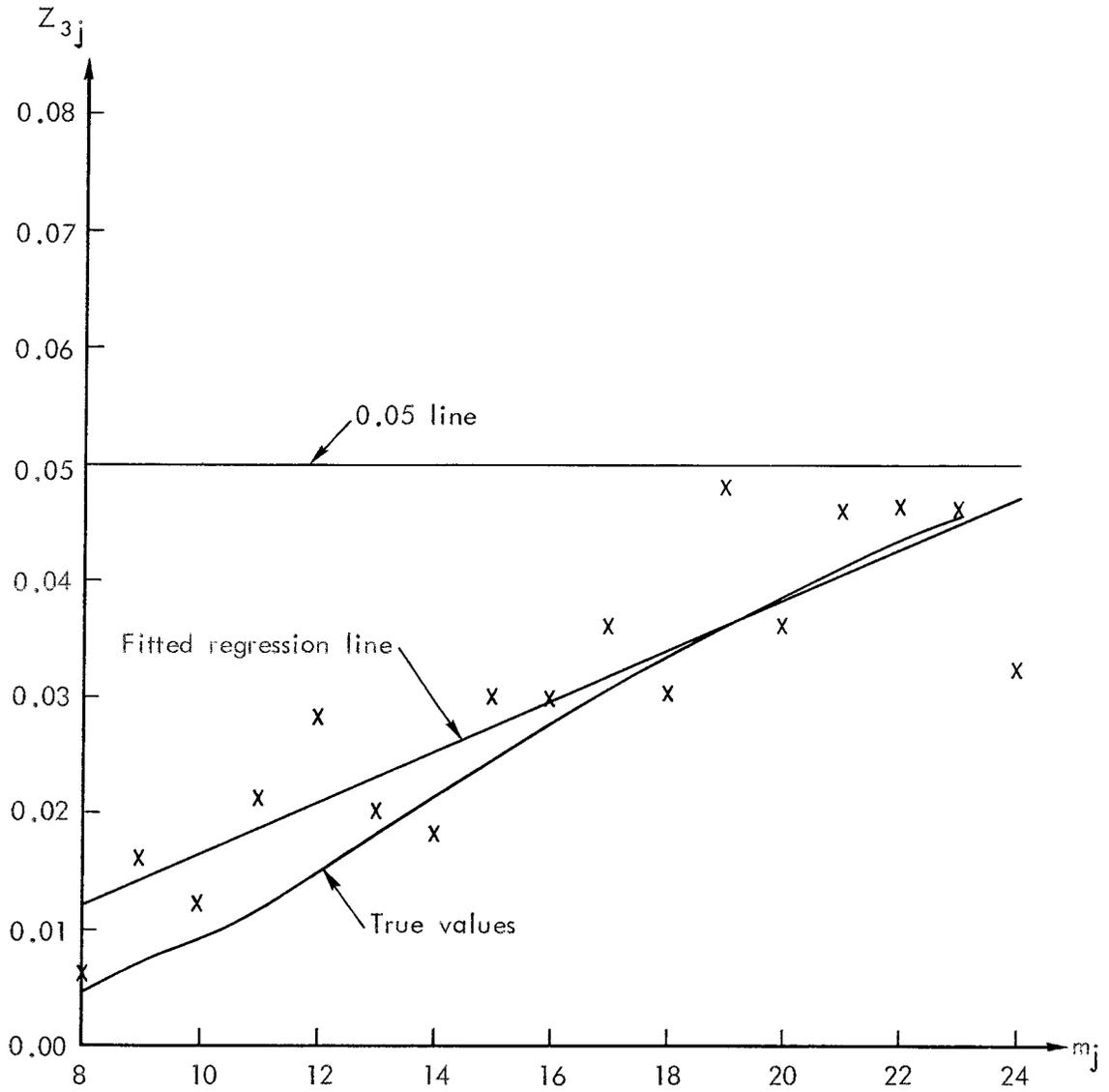
Fig. 6 — MLE and true values for $p = 0.1$

(4.5) are inoperative. The regression rule actually has slightly better risk than (4.5) for these $\{\alpha_{3j}\}$, but here (4.5) has reduced to the regression rule after first protecting against a possible non-linearity in the parameters.

Forcing $\hat{B} \leq 1$ is important here, for otherwise $\hat{B} = 1.93$ produces a loss of 9.7. The MLE has loss 9.0. None of these values is as favorable as it appears, for the value $\sigma^2 = (.01)^2$ for the binomial variance is based on the incorrect assumption that the average rejection probability is .05. For the case $p = 0.1$, the mean rejection probability is about .03 implying that σ^2 is actually about $(3/5)(.01)^2$. The expected MLE loss is therefore about $(3/5)17 \doteq 10$, not $k = 17$. The variances vary by a factor of about 5, contrary to the assumptions for Stein's rule. We could have used the arc sin rule of Sec. II to correct this. That Stein's rule still works well without this variance stabilizing transformation is further evidence of its ruggedness.

Had we ignored the linear trend and used Lindley's modification of Stein's estimator ((2.5) with $X_j = Z_{1j}/\sigma$, $\sigma^2 = (.0295)(.9705)/500 = (.00757)^2$), the loss would have been 1.244 times the MLE loss. The reason is that for this set of data the $\{Z_{3j}\}$ are unusually highly concentrated about their mean, an improbable event causing too much pull-in. We have calculated for this set $\{\alpha_{3j}\}$ that Lindley's rule will have squared error as much as 1.244 times that of the MLE in less than 5 percent of all samples, and will have squared error as much as the MLE in less than 17 percent of all samples. This is an unfortunate set of data! The data clearly contradict the use of Lindley's rule however, a t-test for $\beta_1 = 0$ rejecting at the .00001 level, and in this case the linear fit saves the situation.

With k as large as 17 it costs very little to fit one or two extra parameters by maximum likelihood as in (4.5), which can improve Stein-type estimators. See Efron and Morris (1973a, Sec. 7). Higher order polynomials could be used here in place of the linear regression. For $p = .1$ the quadratic regression coefficient differs significantly from zero at the .20 level, but not at the .10 level. Were the quadratic fit retained and Stein's rule again applied to the residuals as in (4.5), but with $d = 3$, then $\hat{B} = 1$ and the loss would be 3.7, slightly more than in the linear case. The cubic coefficient is not close to being significant.

The case $p = p_2 = 0.3$ requires no new ideas. A plot of these data, taken from Table 9, supports .05 as an appropriate origin. Stein's rule in the form (4.4) gives $\hat{B} = 1$ so $\hat{\alpha}_{2j} = .05$ is the estimate. This estimate has loss 4.8 (11.5 for the MLE). The Lindley modification does about as well. The true values α_{2j} appear in Table 9.

The three problems $p = 0.5$, $p = 0.3$, and $p = 0.1$ have been kept separate so far but can be combined. Most simply, one rule of the form (4.4) or (4.5) could be used simultaneously on all 51 estimation problems. Taking .05 as the origin leads to $\hat{B} = .31$ for all components, but this strong pull is disastrous in the $p = 0.1$ case, producing a loss of 24.6 there and 42.2 overall. For the MLE, the loss for $p = 0.1$ is 9.0 and the overall loss is 39.4. If, instead, linear regressions are fitted separately for $p = 0.5, 0.3, 0.1$, and a rule similar to (4.5) is used on the residuals, then $\hat{B} = .69$ and the losses for the three groups are quite similar to those derived by using separate rules. There is little to gain and much to lose from combining so many problems in the simple form (4.4) (see Efron and Morris, 1973b, and Stein,

1966), so separating the three groups would be preferable. An appropriate generalization of the two groups rule to three groups would probably be better still, but the theory has not yet been developed.

Suspected correlation between α_{1j}, α_{2j} and α_{3j} would provide a good reason for combining the estimation problems. No method presented yet accounts for the possibility, for example, that large α_{1j} suggests large α_{2j}. We have suggested a multivariate empirical Bayes rule in Efron and Morris (1972c), with risk smaller than the MLE, to cover this situation. A brief description of the method and estimates follows.

In the simplified notation of (4.3), suppose $(\theta_{1j}, \theta_{2j}, \theta_{3j})' \sim N_3((0, 0, 0)', \underset{\sim}{A})$ is an independent sample from a trivariate normal distribution, $j = 1, \ldots, k = 17$. Then the Bayes estimate of the 3×17 matrix $\underset{\sim}{\theta} = (\theta_{ij})$ from the 3×17 matrix $\underset{\sim}{X} = (X_{ij})$ is

$$E\underset{\sim}{\theta}|\underset{\sim}{X} = [\underset{\sim}{I} - (\underset{\sim}{I} + \underset{\sim}{A})^{-1}]\underset{\sim}{X}, \tag{4.6}$$

requiring knowledge of $\underset{\sim}{A}$, of course. Stein rules applied separately to p = 0.5, 0.3, and 0.1 correspond to assuming that $\underset{\sim}{A}$ is diagonal in (4.6) and then estimating the diagonal elements. The Stein rule (1.4) applied to all 51 problems is (4.6) with $\underset{\sim}{A}$ known to be diagonal with equal diagonal elements, and the common diagonal element estimated from the pooled data. If A is completely unknown, the uniformly best unbiased estimate of $(\underset{\sim}{I} + \underset{\sim}{A})^{-1}$ based on the marginal distribution of $\underset{\sim}{X}$ is $(k-r-1)(\underset{\sim}{X}\,\underset{\sim}{X}')^{-1}$ for k = 17, r = 3. Substitution into (4.4) gives the multivariate empirical Bayes estimator, which is shown in Efron and Morris (1972c) to dominate the MLE.

To estimate the $\{\alpha_{ij}\}$, we rescaled Z_{ij} and α_{ij} to give (4.3), used the linear regressions already described for each p = 0.5, 0.3, 0.1, and applied the multivariate estimator to the regression residuals as described in Efron and Morris (1972c, Sec. 7), losing d = 2 degrees of freedom in the process (instead of k-r-1 = 13, we must use k-r-d-1 = 11 as the constant). Before retransforming the scale, the estimate of $(\underset{\sim}{I} + \underset{\sim}{A})^{-1}$ was forced to be less than $\underset{\sim}{I}$ by making certain that no eigenvalue of the estimator of $(\underset{\sim}{I} + \underset{\sim}{A})^{-1}$ exceed unity (see Efron and Morris, 1972c, Sec. 6). This is more important in the multivariate than in the univariate case. Denote $\theta_{1j}^{*} = .0509 - .00048(m_j - 16)$, $\theta_{2j}^{*} = .0502 - .00030(m_j - 16)$ and $\theta_{3j}^{*} = .0295 + .00218(m_j - 16)$ as the estimates from regressing Z_{ij} linearly on $m_j - 16$ and let $\widetilde{Z}_{ij} \equiv Z_{ij} - \theta_{ij}^{*}$ be the residuals. The multivariate estimates are

$$\hat{\theta}_{ij} = \theta_{ij}^{*} + \{\widetilde{Z}_{ij} - \hat{B}_{i1}\widetilde{Z}_{1j} - \hat{B}_{i2}\widetilde{Z}_{2j} - \hat{B}_{i3}\widetilde{Z}_{3j}\} \qquad (4.7)$$

with estimates

$$\hat{B}_{11} = .269, \quad \hat{B}_{22} = .961, \quad \hat{B}_{33} = .991, \quad \hat{B}_{12} = \hat{B}_{21} = .169,$$

$$\hat{B}_{13} = \hat{B}_{31} = .082, \quad \hat{B}_{23} = \hat{B}_{32} = -.019.$$

Thus, Z_{2j} and Z_{3j}, as well as Z_{1j}, are used linearly to estimate θ_{1j}, and similarly for the other cases.

The losses of the multivariate rule (4.7) for the three components p = 0.5, 0.3, 0.1 are 11.8, 7.6, 3.4, respectively, totaling 22.7. This is slightly worse in each case than the separate Stein estimators. It happens that the true partial correlations $\rho_{hi} \equiv \sum \widetilde{\alpha}_{hj} \widetilde{\alpha}_{ij} / (\sum \widetilde{\alpha}_{hj}^{2} \sum \widetilde{\alpha}_{ij}^{2})^{\frac{1}{2}}$

(where $\tilde{\alpha}_{ij} \equiv \alpha_{ij} - \alpha_{ij}^*$ and $\alpha_{ij}^* \equiv \beta_{0i} - \beta_{1i}(m_j - 16)$, the best linear fit) are small, $\rho_{12} = .222$, $\rho_{13} = -.081$, $\rho_{23} = .410$, so the multi-variate rule is not at its best here. See Efron and Morris (1972c, Sec. 5).

The losses $\sum(\hat{\alpha}_{ij} - \alpha_{ij})^2/\sigma^2$ of this section are tabulated in Table 8. The numbers in parentheses are efficiencies of the given rules relative to the MLE, eff = (MLE loss/given rule loss) · 500. For reasons given earlier, only rules 3, 4, and 7 of Table 8 would be used, and these rules are roughly twice as efficient as the MLE.

Table 8

LOSSES AND EFFICIENCIES FOR RULES OF THIS SECTION[a]

		p = 0.5	p = 0.3	p = 0.1	Total
1.	MLE	18.9 (500)	11.5 (500)	9.0 (500)	39.4 (500)
2.	Separate Stein rules (4.4) origin minus .05	10.2 (926)	4.8 (1198)	14.1 (319)	29.1 (677)
3.	Separate Stein rules (4.5) after a linear regression	10.3 (917)	5.6 (1027)	3.1 (1452)	19.0 (1037)
4.	Separate Stein rules after a linear regression for p = 0.1 only	10.2 (926)	4.8 (1198)	3.1 (1452)	18.1 (1088)
5.	Two groups rule (Table 7)	9.8 (969)	--	--	--
6.	Combined Stein rule after linear regressions	10.8 (875)	5.2 (1106)	3.5 (1286)	19.6 (1010)
7.	Multivariate rule (4.7) after linear regressions	11.8 (801)	7.6 (757)	3.4 (1324)	22.7 (864)

[a]Efficiency = $\dfrac{\text{MLE loss}}{\text{given rule loss}} \times 500$, given in parentheses, the "equivalent MLE sample size."

The true rejection probabilities $\alpha(p, m)$ are plotted in Fig. 7. Surprisingly, $p = 0.3$ is closest of the three to the null value $\alpha = 0.5$, as Fig. 7 shows.

The data Z_{ij} and true values α_{ij} are given in Table 9. The Z_{ij}, being the ratio of successes to $n = 500$ trials, are reported exactly, but the α_{ij} are rounded to four significant figures.

We also reported on the application of empirical Bayes rules to this computer simulation in Efron and Morris (1972c, Sec. 9), only there $m_1 = 10$, $m_2 = 15$, $m_3 = 20$, and $p_i = .525 - .025i$ for $i = 1$, ..., 17 were studied. That experience was similarly favorable. The situation in this report is more interesting because $\alpha(p, m)$ is a smooth function of p for each m, but it can be quite discontinuous in m for fixed p, as Fig. 7 illustrates.

Table 9

MAXIMUM LIKELIHOOD ESTIMATES AND TRUE VALUES

p:		0.5	0.3	0.1	0.5	0.3	0.1
			MLE			True Values	
j	m_j	Z_{1j}	Z_{2j}	Z_{3j}	α_{1j}	α_{2j}	α_{3j}
1	8	.082	.060	.006	.07681	.05266	.00466
2	9	.042	.060	.016	.05011	.05525	.00715
3	10	.046	.054	.012	.04219	.03711	.00904
4	11	.040	.044	.022	.05279	.04271	.01187
5	12	.054	.048	.028	.06403	.04918	.01491
6	13	.084	.040	.020	.07556	.05648	.01809
7	14	.036	.048	.018	.04102	.05185	.02134
8	15	.036	.052	.030	.04559	.05477	.02459
9	16	.040	.062	.030	.05151	.05739	.02780
10	17	.050	.046	.036	.05766	.04592	.03057
11	18	.078	.036	.030	.06527	.04988	.03343
12	19	.030	.054	.048	.05306	.05340	.03614
13	20	.036	.052	.036	.04253	.05326	.03867
14	21	.060	.054	.046	.04588	.05589	.04102
15	22	.052	.046	.046	.04896	.04807	.04319
16	23	.046	.038	.046	.05417	.04971	.04518
17	24	.054	.060	.032	.05950	.05157	.04701

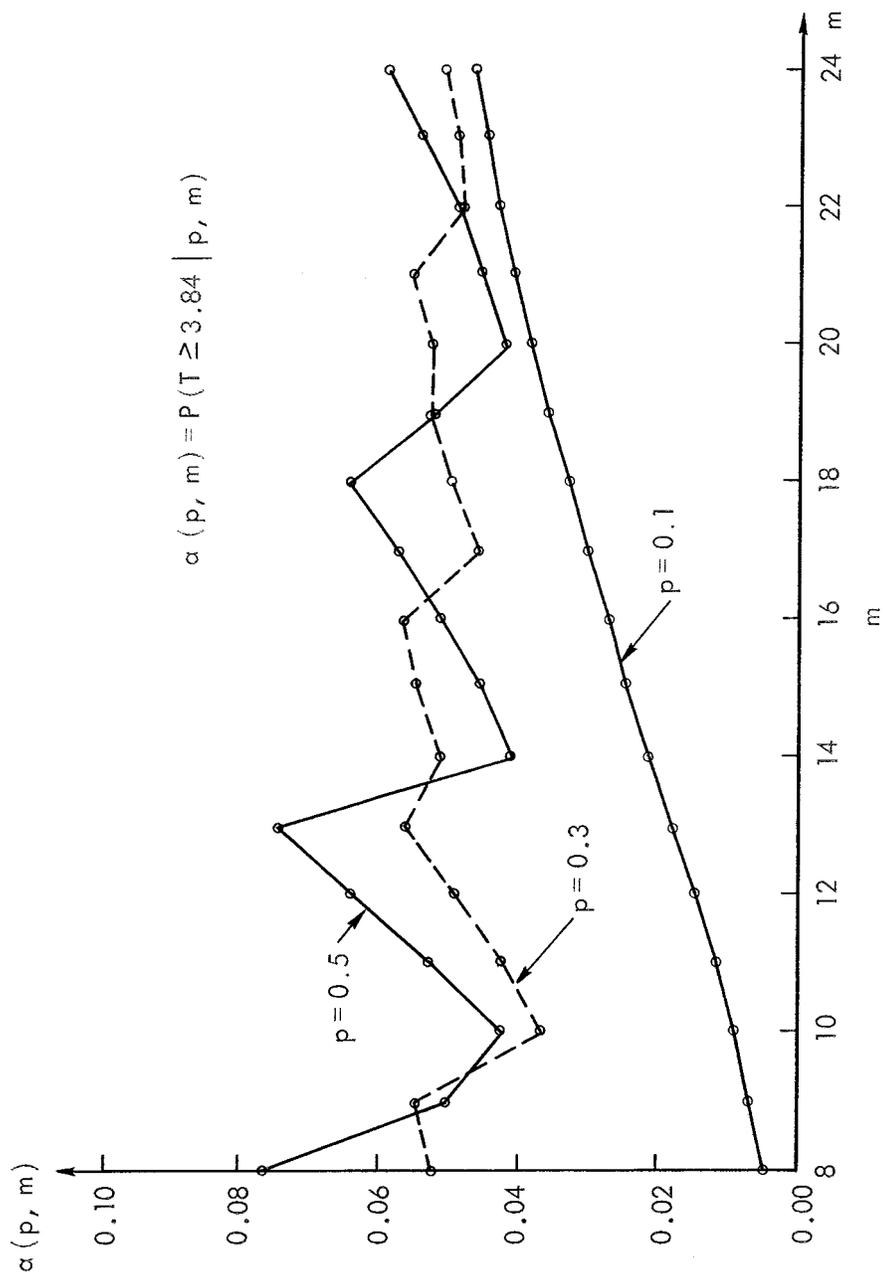$$\alpha(p, m) = P(T \geq 3.84 \mid p, m)$$

Fig. 7— Graph of the rejection probabilities for Pearson's chi-square test (4.1)

V. DISCUSSION

In the baseball, toxoplasmosis, and computer simulation examples of this study, Stein's estimator and its generalizations increased efficiencies relative to the maximum likelihood estimator by about 400 percent, 200 percent, and 100 percent. These examples were chosen, of course, because we expected empirical Bayes methods to work well for them and because their efficiencies could be determined. But we are aware of other successful applications to real data and have suppressed no negative results. Although blind application of these methods would gain little in most instances, the statistician who uses them sensibly and selectively can expect major improvements.

Even when they do not significantly increase efficiency, there is little penalty for using the rules of this report because they cannot give larger total mean squared error than the maximum likelihood estimator and because the limited translation modification (see Sec. II) protects individual components. As several authors have indicated, these rules are also robust to the assumption of the normal distribution, because their operating characteristics depend primarily on the means and variances of the sampling distributions and of the unknown parameters. Nor is the sum of squared error criterion especially important. This robustness is borne out by the experience in this report, for the sampling distributions were actually binomial rather than normal. Here, the rules not only worked well in the aggregate, but for most components the empirical Bayes estimators ranged from slightly to substantially better than the maximum likelihood estimator, with no substantial errors in the other direction.

Professor Tukey's comment, that empirical Bayes benefits are un-
appreciable (Sec. I), was actually directed at a method of D. V.
Lindley. Lindley's rules, though more formally Bayesian, are similar
to ours in that they are designed to pick up the same intercomponent
information in possibly related estimation problems. We have not done
justice here to the many other contributors to multiparameter estima-
tion, but refer the reader to the bibliography. We have concentrated
only on Stein's rule and its generalizations to illustrate the power
of the empirical Bayes theory, because the main gains are derived by
recognizing the applicability of the theory, with lesser benefit
attributable to which particular method is used. Nevertheless, we
hope other authors will compare their methods with ours on these or
other data.

The rules of this report are neither Bayes nor admissible, so they
can be uniformly beaten (but not by much; see Efron and Morris, 1973a).
There are several published, admissible, minimax rules that would also
do well on the baseball data, although probably not much better than
the rule used there, for none yet given dominates the versions of
Stein's rule with the positive part modification. The authors in the
bibliography approach these multiparameter estimation problems from
many perspectives, but all produce rules that compete with Stein's in
that they seek and use intercomponent information. For applications,
we happily recommend the combination of simplicity, generalizability,
efficiency, and robustness found in the estimators presented here.

The most favorable situation for these estimators occurs when the
statistician wants to estimate the parameters of a linear model that
are known to lie in a high dimensional parameter space H_1, but he

suspects that they may lie close to a specified lower dimensional

parameter space $H_0 \subset H_1$. Then estimates unbiased for every parameter

vector in H_1 may have large variance, while estimates restricted to

H_0 have smaller variance but possibly large bias. The statistician

need not choose between these extremes but can instead view them as

endpoints on a continuum and use the data to determine the compromise[*]

between bias and variance through an appropriate empirical Bayes rule,

perhaps Stein's or one of the generalizations of this report.

We believe many applications embody these features and that most

data analysts will have good experiences with the sensible use of the

rules we recommend. In view of their potential, we regard empirical

Bayes methods to be among the most underutilized tools in applied data

analysis.

---

[*] The amount of compromise is usually a smooth function of the
likelihood ratio statistic for testing H_0 versus H_1.

BIBLIOGRAPHY

Alam, Khursheed, "A Family of Admissible Minimax Estimators of the
  Mean of a Multivariate Normal Distribution," *Annals of Statist.*,
  Vol. 1, No. 3, May 1973, pp. 517-525.

------, *Minimax Estimators of the Mean of a Multivariate Normal Dis-
  tribution*, Clemson University, Department of Mathematical Sciences,
  Technical Report No. 142, August 1973.

------ and J. R. Thompson, *Estimation of the Mean of a Multivariate
  Normal Distribution*, Indiana University, Technical Report, 1964.

Anscombe, F., "The Transformation of Poisson, Binomial and Negative-
  Binomial Data," *Biometrika*, Vol. 35, 1948, pp. 246-254.

Atchison, T. A., and H. F. Martz, Jr. (eds.), *Proceedings of the
  Symposium on Empirical Bayes Estimation and Computing in Statistics*,
  Texas Tech University, Department of Mathematics, Mathematics Series
  No. 6, August 8-9, 1969.

Baranchik, A. J., *Multiple Regression and Estimation of the Mean of a
  Multivariate Normal Distribution*, Stanford University, Department
  of Statistics, Technical Report No. 51, 1964.

------, *Subminimax Estimation of the Mean of a Normal Random Variable*,
  Columbia University, Department of Mathematical Statistics, February
  1966.

------, "A Family of Minimax Estimators of the Mean of a Multivariate
  Normal Distribution," *Ann. Math. Statist.*, Vol. 41, No. 2, April
  1970, pp. 642-645.

------, "Inadmissibility of Maximum Likelihood Estimators in Some
  Multiple Regression Problems with Three or More Independent Vari-
  ables," *Annals of Statist.*, Vol. 1, No. 2, 1973, pp. 312-321.

Bhattacharya, P. K., "Estimating the Mean of a Multivariate Normal
  Population with General Quadratic Loss Function," *Ann. Math. Statist.*,
  Vol. 37, 1966, pp. 1819-1824.

Bock, M. E., G. G. Judge, and T. A. Yancey, "Estimation in Regression
  After Preliminary Tests of Significance," University of Illinois,
  November 1971 (unpublished).

Box, G.E.P., and G. C. Tiao, "Multiparameter Problems from a Bayesian
  Point of View," *Ann. Math. Statist.*, Vol. 36, 1965, pp. 1468-1482.

------, "Bayesian Estimation of Means for the Random Effects Model,"
  J. Amer. Statist. Assoc., Vol. 63, 1968, pp. 174-181.

Box, G.E.P., and G. C. Tiao, *Bayesian Inference in Statistical Analysis*, Addison-Wesley Publishing Company, Reading, Mass., 1973.

Brown, L. D., "On the Admissibility of Invariant Estimators of One or More Location Parameters," *Ann. Math. Statist.*, Vol. 37, October 1966, pp. 1087-1136.

—————, "Inadmissibility of the Usual Estimators of Scale Parameters in Problems with Unknown Location and Scale Parameters," *Ann. Math. Statist.*, Vol. 39, 1968, pp. 29-48.

—————, "Admissible Estimators, Recurrent Diffusions, and Insoluble Boundary Value Problems," *Ann. Math. Statist.*, Vol. 42, 1971, pp. 855-903.

—————, *Estimation with Incompletely Specified Loss Functions (the Case of Several Location Parameters)*, Cornell University, 1973.

Carter, G. M., and J. E. Rolph, *Empirical Bayes Methods Applied to Spatial Analysis Problems*, The Rand Corporation, P-5076, August 1973.

—————, *New York City Fire Alarm Prediction Models I: Box-Reported Serious Fires*, The Rand Corporation, R-1214-NYC, 1973.

Christiansen, H. Dalgas, "A Better Estimator for the Mean of a Multivariate Normal Distribution," Technical University of Denmark, 1973 (unpublished).

Clemmer, B. A., and R. G. Krutchkoff, "The Use of Empirical Bayes Estimators in a Linear Regression Model," *Biometrika*, Vol. 55, 1968, pp. 525-534.

Cogburn, R., "On the Estimation of a Multivariate Location Parameter with Squared Error Loss," in J. Newman and L. LeCam (eds.), *Bernoulli (1723), Bayes (1763), and Laplace (1813) Anniversary Volume*, Springer-Verlag, Berlin, 1965, pp. 24-29.

—————, "Stringent Solutions to Statistical Decision Problems," *Ann. Math. Statist.*, Vol. 38, 1967, pp. 447-463.

Cohen, A., and W. E. Strawderman, "Admissibility of Estimators of the Mean Vector of a Multivariate Normal Distribution with Quadratic Loss," *Ann. Math. Statist.*, Vol. 42, 1971, pp. 270-296.

Copas, J. B., "Compound Decisions and Empirical Bayes (with Discussion), *J. Royal Statist. Soc.*, B, Vol. 31, No. 3, 1969, pp. 387-425.

—————, "Empirical Bayes Methods and the Repeated Use of a Standard," *Biometrika*, Vol. 59, No. 2, 1972, pp. 349-360.

Dempster, A. P., "Model Searching and Estimation in the Logic of Inference," in V. P. Godambe and D. A. Sprott (eds.), *Foundations of Statistical Inference*, Holt, Rinehart and Winston, Toronto, Canada, 1971.

Draper, N. R., "'Ridge Analysis' of Response Surfaces," *Technometrics*, Vol. 5, 1963, pp. 469-479.

Efron, B., and C. Morris, "Limiting the Risk of Bayes and Empirical Bayes Estimators--Part I: The Bayes Case," *J. Amer. Stat. Assoc.*, Vol. 66, No. 336, December 1971, pp. 807-815.

------, "Limiting the Risk of Bayes and Empirical Bayes Estimators--Part II: The Empirical Bayes Case," *J. Amer. Stat. Assoc.*, Vol. 67, No. 337, March 1972a, pp. 130-139.

------, *Simultaneous Estimation of Parameters*, The Rand Corporation, P-4835, May 1972b.

------, "Empirical Bayes on Vector Observations--An Extension of Stein's Method," *Biometrika*, Vol. 59, No. 2, August 1972c, pp. 335-347.

------, "Stein's Estimation Rule and Its Competitors--An Empirical Bayes Approach," *J. Amer. Stat. Assoc.*, Vol. 68, No. 341, March 1973a, pp. 117-130.

------, "Combining Possibly Related Estimation Problems," *JRSS*, B, November 1973b (with discussion).

------, *Families of Minimax Estimators of the Mean of a Multivariate Normal Distribution*, submitted to *Ann. Math. Statist.*, 1974.

------, "Estimating Several Parameters Simultaneously," to be published in *Statistica Neerlandica* (also available The Rand Corporation, R-1395-OEO, forthcoming).

Farrar, D. E., and R. R. Glauber, "Multicollinearity in Regression Analysis: The Problem Revisited," *Review of Economics and Statistics*, Vol. 49, 1967, pp. 96-108.

Fienberg, S. E., and P. W. Holland, "Methods for Eliminating Zero Counts in Contingency Tables," in G. P. Patil (ed.), *Random Counts in Models and Structures*, Pennsylvania State University Press, University Park, 1970, pp. 233-260.

------, "On the Choice of Flattening Constants for Estimating Multinomial Probabilities," *J. Multivariate Analysis*, Vol. 2, 1972, pp. 127-134.

------, "Simultaneous Estimation of Multinomial Cell Probabilities," *J. Amer. Statist. Assoc.*, Vol. 68, No. 343, September 1973, pp. 683-691.

Geisser, S., "Bayes Estimation in Multivariate Analysis," *Ann. Math Statist.*, Vol. 36, 1965, pp. 150-159.

George, S. L., "Evaluation of Empirical Bayes Estimators for Small Numbers of Past Samples," *Biometrika*, Vol. 58, 1971, p. 244.

Goodman, A. S., P.A.W. Lewis, and H. E. Robbins, *Simultaneous Estimation of Large Numbers of Extreme Quantiles in Simulation*, Naval Postgraduate School, Monterey, California, Technical Report, December 1971.

Griffin, B. S., and R. G. Krutchkoff, "Optimal Linear Estimators: An Empirical Bayes Version with Application to the Binomial Distribution," *Biometrika*, Vol. 58, No. 1, 1971, pp. 195-201.

Haff, Leonard R., "Bayesian Regression with First Order Autoregressive Priors," Department of Mathematics, University of California, San Diego, 1973 (unpublished).

Hartigan, J., "Linear Bayes Methods," *J. Royal Statist. Soc.*, B, Vol. 31, December 1969, pp. 446-454.

Hoerl, A. E., "Application of Ridge Analysis to Regression Problems," *Chemical Engineering Progress*, Vol. 58, 1970, pp. 55-67.

-----, and R. W. Kennard, "Ridge Regression. Applications to Nonorthogonal Problems," *Technometrics*, Vol. 12, 1970a, pp. 69-82.

-----, "Ridge Regression. Biased Estimation for Nonorthogonal Problems," *Technometrics*, Vol. 12, 1970b, pp. 55-68.

Holland, B. S., *On the Estimation of Regression Coefficients with a Coordinatewise Mean Square Error Criterion of Goodness*, Institute of Statistics Mimeograph Series No. 693, North Carolina State University, Raleigh, July 1970.

Holland, Paul W., *Weighted Ridge Regression: Combining Ridge and Robust Regression Methods*, National Bureau of Economic Research, Inc., Working Paper No. 11, September 1973.

Huntsberger, D. V., "A Generalization of a Preliminary Testing Procedure for Pooling Data," *Annals of Mathematical Statistics*, Vol. 26, 1955, pp. 734-743.

Jackson, P. H., M. R. Novick, and D. T. Thayer, "Estimating Regression in m-Groups," *Brit. J. Math. Statist. Psychol.*, Vol. 24, 1971, pp. 129-153.

James, W., and C. Stein, "Estimation with Quadratic Loss," *Proc. Fourth Berkeley Symp. Math. Statist. Prob.*, Vol. 1, University of California Press, 1961, pp. 361-379.

Johns, M. V., Jr., "Two-Action Compound Decision Problems," *Proc. Fifth Berkeley Symp. Math. Statist. Prob.*, University of California Press, 1967, pp. 463-478.

Kantor, M., "Estimating the Mean of a Multivariate Normal Distribution with Applications to Time Series and Empirical Bayes Estimation," Ph.D. dissertation, Columbia University, 1967.

Kitagawa, Tosio, "Estimation After Preliminary Tests of Significance," *University of California Publications in Statistics,* Vol. 3, No. 4, 1963, pp. 147-186.

Lemon, G. H., and R. G. Krutchkoff, "An Empirical Bayes Smoothing Technique," *Biometrika,* Vol. 56, No. 2, 1969, pp. 361-365.

Leonard, T., "Bayesian Methods for Binomial Data," *Biometrika,* Vol. 59, No. 3, 1972, pp. 581-589.

-----, "A Bayesian Method for Histograms," *Biometrika,* Vol. 60, No. 2, 1973, pp. 297-308.

Lewis, Charles, Ming-Mei Wang, and Melvin R. Novick, *Marginal Distributions for the Estimation of Proportions in m Groups,* American College Testing Program, Iowa City, Iowa, ACT Technical Bulletin No. 13, May 1973.

Lin, Pi-Erh, and Hui-Liang Tsai, "Generalized Bayes Minimax Estimators of the Multivariate Normal Mean with Unknown Covariance Matrix," *Annals of Statist.,* Vol. 1, No. 1, 1973, pp. 142-145.

Lindley, D. V., "The Estimation of Many Parameters," in V. P. Godambe and D. A. Sprott (eds.), *Foundations of Statistical Inference,* Holt, Rinehart and Winston, Toronto, 1971, pp. 435-455.

-----, *A Bayesian Solution for Two-Way Analysis of Variance,* American College Testing Program, Iowa City, Iowa, ACT Technical Bulletin No. 8, September 1972a.

-----, *Multiple Regression in a Two-Way Layout,* American College Testing Program, Iowa City, Iowa, ACT Technical Bulletin No. 9, September 1972b.

-----, *Bayesian Statistics, A Review,* Regional Conference Series in Applied Mathematics, 2, SIAM, Philadelphia, 1972c.

-----, and A.F.M. Smith, "Bayes Estimates for the Linear Model (with Discussion)," *J. Royal Statist. Soc.,* B, Vol. 34, 1972, pp. 1-42.

Macky, D., *Empirical Bayes Estimation in an Exponential Family,* Michigan State University, Department of Statistics and Probability, RM-176, 1966.

Maritz, J. S., "Smooth Empirical Bayes Estimation for One-Parameter Discrete Distributions," *Biometrika,* Vol. 53, 1966, pp. 417-429.

-----, "Smooth Empirical Bayes Estimation for Continuous Distribution," *Biometrika,* Vol. 54, 1967a, pp. 435-450.

-----, "Empirical Bayes Estimation for the Poisson Distribution," *Biometrika,* Vol. 54, 1967b, pp. 367-374.

Maritz, J. S., "On the Smooth Empirical Bayes Approach to Testing of Hypotheses and the Compound Decision Problem," *Biometrika*, Vol. 55, 1968, pp. 83-100.

------, *Empirical Bayes Methods*, Methuen & Co., Ltd., London, 1970.

Marquardt, D. W., "Generalized Inverses, Ridge Regression, Biased Linear Estimation and Nonlinear Estimation," *Technometrics*, Vol. 12, 1970, pp. 591-612.

Martz, H., and R. Krutchkoff, "Empirical Bayes Estimators in a Multiple Linear Regression Model," *Biometrika*, Vol. 56, 1969, pp. 367-374.

McDonald, G. C., and D. I. Galarneau, *An Evaluation of Two Ridge-Type Estimators of Linear Regression Coefficients*, General Motors Research Laboratories, Mathematics Department, GMR-1322, December 1972.

Miyasawa, K., "An Empirical Bayes Estimator of the Mean of a Normal Population," *Bull. Inst. Internat. Statist.*, Vol. 38, 1961, pp. 181-188.

Newhouse, J. P., and S. D. Oman, *An Evaluation of Ridge Estimators*, The Rand Corporation, R-716-PR, 1971.

Neyman, J., "Two Breakthroughs in the Theory of Statistical Decision Making," *Rev. Int. Statist. Inst.*, Vol. 30, 1962, pp. 11-27.

Novick, M. R., "Multiparameter Bayesian Indifference Procedures," *J. Royal Statist. Soc.*, B, Vol. 31, 1969, pp. 29-64.

------, and D. T. Thayer, *A Comparison of Bayesian Estimates of True Score*, Educational Testing Service, Princeton, New Jersey, RB-69-74, September 1969.

Novick, M. R., P. H. Jackson, D. T. Thayer, and N. S. Cole, "Applications of Bayesian Methods to the Prediction of Educational Performance," *Brit. J. Math. Statist. Psychol.*, Vol. 25, 1972, pp. 33-50.

Novick, M. R., C. Lewis, and P. H. Jackson, "The Estimation of Proportions in m Groups," *Psychometrika*, Vol. 28, No. 1, March 1973, pp. 19-46.

Portnoy, S., *Formal Bayes Estimation with Application to a Random Effects Analysis of Variance Model*, Stanford University, Department of Statistics, Technical Report No. 142, 1968.

Press, S. James., *Statistical Estimation by the Empirical Bayes Method: Some Extensions and Logistical Applications*, The Rand Corporation, RM-4442-PR, 1965.

Radhakrishnan, R., "Estimation of the Mean of Y Using Observations on (Y, X) and Extra Observations on X^1," Carnegie-Mellon University, Department of Statistics, 1972 (unpublished).

Remington, J. S., et al., "Studies on Toxoplasmosis in El Salvador Prevalence and Incidence of Toxoplasmosis as Measured by the Sabin-Feldman Dye Test," *Transactions of the Royal Society of Tropical Medicine and Hygiene*, Vol. 64, No. 2, 1970, pp. 252-267.

Robbins, H., "Asymptotically Subminimax Solutions of Compound Statistical Decision Problems," in *Proc. Second Berkeley Symp. Math. Statist. Prob.*, University of California Press, 1950, pp. 131-148.

──────, "An Empirical Bayes Approach to Statistics," in *Proc. Third Berkeley Symp. Math. Statist. Prob.*, Vol. 1, University of California Press, 1955, pp. 157-164.

Robbins, H., "The Empirical Bayes Approach to Statistical Decision Problems," *Ann. Math. Statist.*, Vol. 35, 1964, pp. 1-20.

Rolph, J. E., "Bayesian Estimation of Mixing Distributions," *Ann. Math. Statist.*, Vol. 39, No. 4, 1968, pp. 1289-1302.

Rutherford, J. R., and R. G. Krutchkoff, "The Empirical Bayes Approach: Estimating the Prior Distribution," *Biometrika*, Vol. 54, 1967, pp. 326-328.

──────, "Some Empirical Bayes Techniques in Point Estimation," *Biometrika*, Vol. 56, No. 1, 1969a, pp. 133-137.

──────, "ε Asymptotic Optimality of Empirical Bayes Estimators," *Biometrika*, Vol. 56, No. 1, 1969b, pp. 220-223.

Samuel, E., "The Compound Statistical Decision Problem," *Sankhyā*, A, Vol. 29, 1967, pp. 123-140.

Sclove, S. L., "Improved Estimators for Coefficients in Linear Regression," *J. Amer. Statist. Assoc.*, Vol. 63, No. 322, 1968, pp. 596-606.

──────, "Improved Estimation of Parameters in Multivariate Regression," *Sankhyā*, A, Vol. 33, Part 1, 1971, pp. 61-66.

──────, C. Morris, and R. Radhakrishnan, "Non-Optimality of Preliminary-Test Estimators for the Mean of a Multivariate Normal Distribution," *Ann. Math. Statist.*, Vol. 43, No. 5, October 1972, pp. 1481-1490.

Sprott, D. A., "The Estimation of Prior Distributions in Problems of Decision," *Sankhyā*, A, Vol. 29, 1967, pp. 227-238.

Stein, C., "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Normal Distribution," *Proc. Third Berkeley Symp. Math. Statist. Prob.*, Vol. 1, University of California Press, 1955, pp. 197-206.

──────, "Multiple Regression," in I. Olkin (ed.), *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, Stanford University Press, Stanford, California, 1960, pp. 424-443.

Stein, C., "Estimation of Many Parameters," *Inst. Math. Statist.*, Wald Lectures, 1961.

-----, "Confidence Sets for the Mean of a Multivariate Normal Distribution," *J. Royal Statist. Soc.*, B, Vol. 24, 1962, pp. 265-296.

-----, "Inadmissibility of the Usual Estimator for the Variance of a Normal Distribution with Unknown Mean," *Ann. Inst. Statist. Math.*, Vol. 16, 1965, pp. 155-156.

-----, "An Approach to the Recovery of Inter-Block Information in Balanced Incomplete Block Designs," in F. N. David (ed.), *Festschrift for J. Neyman*, John Wiley & Sons, Inc., New York, 1966, pp. 351-366.

-----, *Estimation of the Mean of a Multivariate Normal Distribution*, Stanford University, Department of Statistics, Technical Report No. 48, June 1973 (submitted to *Annals of Statistics*).

-----, B. Efron, and C. Morris, *Improving the Usual Estimator of a Normal Covariance Matrix*, Stanford University, Department of Statistics, Technical Report No. 37, March 1972.

Stone, M., "Cross-validatory Choice and Assessment of Statistical Predictions," *J. Royal Statist. Soc.*, 1973 (forthcoming).

Strawderman, W. E., "On the Existence of Proper Bayes Minimax Estimators of the Mean of a Multivariate Normal Distribution," *Proc. Sixth Berkeley Symp. Math. Statist. Prob.*, University of California Press, 1970.

-----, "Proper Bayes Minimax Estimators of the Multivariate Normal Mean," *Ann. Math. Statist.*, Vol. 42, No. 1, 1971, pp. 385-388.

-----, "Proper Bayes Minimax Estimators of the Multivariate Normal Mean Vector for the Case of Common Unknown Variances," *Ann. Math. Statist.*, Vol. 1, No. 6, 1973, pp. 1189-1194.

-----, and Arthur Cohen, "Admissibility of Estimators of the Mean Vector of a Multivariate Normal Distribution with Quadratic Loss," *Ann. Math. Statist.*, Vol. 42, No. 1, 1971, pp. 270-296.

Susarla, V., *Rates of Convergence in Sequence-Compound Squared-Distance Loss Estimation and Two-Action Problems*, Michigan State University, SLP-24, September 1970.

Sutherland, M., "Estimation in Large Sparse Multinomials," Ph.D. dissertation, Harvard University, 1973.

-----, P. W. Holland, and S. E. Fienberg, *Combining Bayes and Frequency Approaches to Estimating a Multinomial Parameter*, University of Minnesota, School of Statistics, Technical Report No. 198, 1973.

Thatcher, A. R., "Relationships Between Bayesian and Confidence Limits for Predictions," *J. Royal Statist. Soc.*, B, Vol. 26, 1964, pp. 176-210.

Thompson, J. R., "On the Inadmissibility of X as the Estimate of the Mean of a p-Dimensional Normal Distribution for $p \geq 3$," Indiana University, 1968 (unpublished).

Tiao, G. C., and G.E.P. Box, "Bayesian Analysis of a Three Component Hierarchical Design Model," *Biometrika*, Vol. 54, 1967, pp. 109-125.

-----, and W. Y. Tan, "Bayesian Analysis of Random-Effect Models in the Analysis of Variance. I. Posterior Distribution of Variance-Components," *Biometrika*, Vol. 52, 1965, pp. 37-53.

-----, "Bayesian Analysis of Random-Effect Models in the Analysis of Variance. II. Effect of Autocorrelated Errors," *Biometrika*, Vol. 53, 1966, pp. 477-495.

-----, and A. Zellner, "Bayes's Theorem and the Use of Prior Knowledge in Regression Analysis," *Biometrika*, Vol. 51, 1964a, pp. 219-230.

-----, "On the Bayesian Estimation of Multivariate Regression," *J. Royal Statist. Soc.*, B, Vol. 26, 1964, pp. 277-285.

Toro-Vizcarrondo, C., and T. D. Wallace, "A Test of the Mean Square Criterion for Restrictions in Linear Regression," *J. Amer. Statist. Assoc.*, Vol. 63, 1968, pp. 558-572.

Trybula, S., "Some Problems of Simultaneous Minimax Estimation," *Ann. Math. Statist.*, Vol. 29, 1958, pp. 245-253.

Wang, Ming-mei, and Charles Lewis, "Estimation of Proportions in Two-Way Tables," The American College Testing Program, The Research and Development Division, *ACT Technical Bulletin No. 16*, October 1973.

Wermuth, N. E., "An Empirical Comparison of Regression Methods," Ph.D. dissertation, Harvard University, September 1972.

Wind, S. L., "Stein-James Estimators of a Multivariate Location Parameter," *Ann. Math. Statist.*, Vol. 43, 1972, pp. 340-343.

-----, "An Empirical Bayes Approach to the Multiple Linear Regression Problem," *Annals of Statist.*, Vol. 1, No. 1, 1973, pp. 93-103.

Yancey, T. A., G. G. Judge, and M. E. Bock, "A Mean Square Error Test When Stochastic Restrictions are Used in Regression," University of Illinois, July 1971 (unpublished).

Zellner, A., and W. Vandaele, "Bayes-Stein Estimators for K-Means, Regression and Simultaneous Equation Models," presented at the third *NBER-NSF Symposium on Bayesian Inference in Econometrics*, Harvard University, 15 October 1971.

Zidek, J. V., "Sufficient Conditions for the Admissibility Under Squared Error Loss of Formal Bayes Estimators," *Ann. Math. Statist.*, Vol. 41, No. 2, 1970, pp. 446-456.