UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER GWU/IMSE/Serial T-486/84	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* AN APPLICATION OF EMPIRICAL BAYES TECHNIQUES TO THE SIMULTANEOUS ESTIMATION OF MANY PROBABILITIES		5. TYPE OF REPORT & PERIOD COVERED SCIENTIFIC 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s)		8. CONTRACT OR GRANT NUMBER(s) N00014-83-K-0216
9. PERFORMING ORGANIZATION NAME AND ADDRESS GEORGE WASHINGTON UNIVERSITY INSTITUTE FOR MANAGEMENT SCIENCE AND ENGINEERING WASHINGTON, DC 20052		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS OFFICE OF NAVAL RESEARCH CODE 411 S&P ARLINGTON, VIRGINIA 22217		12. REPORT DATE 3 May 1984 13. NUMBER OF PAGES 19
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*		15. SECURITY CLASS. *(of this report)* UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

APPROVED FOR PUBLIC SALE AND RELEASE; DISTRIBUTION UNLIMITED.

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

18. SUPPLEMENTARY NOTES

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

BINARY DATA
CROSS-VALIDATION
EM ALGORITHM
EMPIRICAL BAYES ESTIMATES
READINESS EVALULATION

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

Consider the following situation: each of N different combat units is presented with a number of requirements to satisfy, each requirement being classified into one of K mutually exclusive categories. For each unit and each category, an estimate of the probability of that unit satisfying any requirement in that category is desired. The problem can be generally stated as that of estimating N different K-dimensional vectors of probabilities based upon a corresponding set of K-dimensional vectors of sample proportions.

(Continued)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

An empirical Bayes model is formulated and applied to an example from the Marine Corps Combat Readiness Evaluation System (MCCRES). The EM algorithm provides a convenient method of estimating the prior parameters. The Bayes estimates are compared to the ordinary estimates, i. e. the sample proportions, by means of cross-validation and the Bayes estimates are shown to provide considerable improvement.

AN APPLICATION OF EMPIRICAL BAYES TECHNIQUES TO THE SIMULTANEOUS ESTIMATION OF MANY PROBABILITIES

by

S. S. Brier
S. Zacks
W. H. Marlow

READINESS RESEARCH
GWU/IMSE/Serial T-486/84
3 May 1984

THE GEORGE WASHINGTON UNIVERSITY
School of Engineering and Applied Science
Washington, DC 20052

Institute for Management Science and Engineering

Research Supported
by
Contract N00014-83-K-0216
Project NR 347-131
Office of Naval Research

THE GEORGE WASHINGTON UNIVERSITY
School of Engineering and Applied Science
Washington, DC 20052

Institute for Management Science and Engineering

AN APPLICATION OF EMPIRICAL BAYES TECHNIQUES TO THE SIMULTANEOUS ESTIMATION OF MANY PROBABILITIES

by

S. S. Brier
S. Zacks
W. H. Marlow

Abstract
of
READINESS RESEARCH
GWU/IMSE/Serial T-486/83
3 May 1984

Consider the following situation: each of N different combat units is presented with a number of requirements to satisfy, each requirement being classified into one of K mutually exclusive categories. For each unit and each category, an estimate of the probability of that unit satisfying any requirement in that category is desired. The problem can be generally stated as that of estimating N different K-dimensional vectors of probabilities based upon a corresponding set of K-dimensional vectors of sample proportions. An empirical Bayes model is formulated and applied to an example from the Marine Corps Combat Readiness Evaluation System (MCCRES). The EM algorithm provides a convenient method of estimating the prior parameters. The Bayes estimates are compared to the ordinary estimates, i.e. the sample proportions, by means of cross-validation and the Bayes estimates are shown to provide considerable improvement.

Research Supported
by
Contract N00014-83-K-0216
Project NR 347-131
Office of Naval Research

been defined by the Marine Corps after a careful analysis of each requirement's structure. The categories are:

1. Advance preparation and education.
2. Combat service support and logistics considerations.
3. Mission planning and preparation.
4. Command control and task organization.
5. Execution.
6. Information and communication.

Taking the category model as a starting point, there is still the problem of estimating K probabilities for each unit (in our case there are 75 units in the study). The use of empirical Bayes methodology has proven useful in the simultaneous estimation of many parameters and suggests itself as a promising tool here. We will demonstrate that an appropriate Bayes model, described below, does indeed lead to improved estimation of the category probabilities.

2. The Category Model

For the $N = 75$ units in the present study we define X_{ij} to be the number of satisfied requirements in category j, for the i-th unit. Let M_{ij} be the number of requirements in category j that the i-th unit was tested on. Due to practical constraints units are generally not evaluated on all requirements. If θ_{ij} is defined to be the probability of unit i satisfying a requirement in the j-th category then we assume that X_{ij} has a binomial distribution, $B(M_{ij},\theta_{ij})$. We further assume that given $\underset{\sim}{\theta}_i \equiv (\theta_{i1},\ldots,\theta_{iK})'$, $X_{i1},\ldots,X_{iK}$ are independent. These two assumptions are consistent with the fact that, for a particular unit, performances on different requirements are independent. Finally, we assume that $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_N$ are independent, where $\underset{\sim}{X}_i \equiv (X_{i1},\ldots,X_{iK})$, i.e., the performances of

different units are independent. These assumptions are summarized as:

$$X_{ij} \sim B(M_{ij},\theta_{ij}) \ , \quad i=1,\ldots,N \ , \quad j=1,\ldots,K \tag{1}$$

$$X_{i1},\ldots,X_{iK} \ , \text{ given } \underset{\sim}{\theta}_i \ , \text{ are independent} \ , \tag{2}$$

$$\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_N \text{ are independent.} \tag{3}$$

The most critical assumption of the category model is that all requirements in the same category have the same probability of being satisfied. Zacks, Marlow, and Barzily (1981) and Zacks and Marlow (1982) have concluded that the assumption of equal probabilities within a category is tenable for a ten category model similar to the one that we propose. We use a cross-validation technique to investigate the appropriateness of the six categories. The 234 requirements of section A were randomly split in half producing two sets of count vectors, $\{\underset{\sim}{X}_1^{(1)},\ldots,\underset{\sim}{X}_{75}^{(1)}\}$ and $\{\underset{\sim}{X}_1^{(2)},\ldots,\underset{\sim}{X}_{75}^{(2)}\}$. If the probability of satisfying a requirement is a function only of its category then we should have equal probabilities for $X_{ij}^{(1)}$ and $X_{ij}^{(2)}$. More formally:

$$X_{ij}^{(1)} \sim B(M_{ij}^{(1)} \ , \ \theta_{ij}^{(1)}) \tag{4}$$

$$X_{ij}^{(2)} \sim B(M_{ij}^{(2)} \ , \ \theta_{ij}^{(2)}) \ , \tag{5}$$

and we want to test the hypotheses,

$$H_0: \ \theta_{ij}^{(1)} = \theta_{ij}^{(2)} \ , \quad i=1,\ldots,N \ ; \quad j=1,\ldots,K \tag{6}$$

$$\text{vs.} \quad H_A: \quad \text{some} \ \ \theta_{ij}^{(1)} \neq \theta_{ij}^{(2)} \tag{7}$$

3. Verifying the Assumptions of the Category Model

For a particular unit and category, we can test whether $\theta_{ij}^{(1)} = \theta_{ij}^{(2)}$ by using Fisher's exact test (see Fisher (1925)). This test is illustrated in Table 1 below.

Table 1

2x2 Contingency Table for Testing Whether

$$\theta_{ij}^{(1)} = \theta_{ij}^{(2)}$$

	Satisfied	Not Satisfied	
First Half	$X_{ij}^{(1)}$	$(M_{ij}^{(1)} - X_{ij}^{(1)})$	$M_{ij}^{(1)}$
Second Half	$X_{ij}^{(2)}$	$(M_{ij}^{(2)} - X_{ij}^{(2)})$	$M_{ij}^{(2)}$
	C_{ij}	$\bar{C}_{ij}$	M_{ij}

Fisher's test is based on the conditional distribution of $X_{ij}^{(1)}$ given C_{ij}, $\bar{C}_{ij}$, $M_{ij}^{(1)}$, and $M_{ij}^{(2)}$. Under H_0, we have

$$P\{X_{ij}^{(1)} = x_{ij}^{(1)} \mid C_{ij}, \bar{C}_{ij}, M_{ij}^{(1)}, M_{ij}^{(2)}\} = \frac{\binom{M_{ij}^{(1)}}{x_{ij}^{(1)}}\binom{M_{ij}^{(2)}}{x_{ij}^{(2)}}}{\binom{M_{ij}}{C_{ij}}}, \qquad (8)$$

and the conditional significance level or "p-value" of the test is given by

$$\hat{\alpha}_{ij} = \sum_{x \in I} P\{X_{ij}^{(1)} = x \mid C_{ij}, \bar{C}_{ij}, M_{ij}^{(1)}, M_{ij}^{(2)}\}, \qquad (9)$$

where

$$I \equiv \left\{x: P\{X_{ij}^{(1)} = x | C_{ij}, \bar{C}_{ij}, M_{ij}^{(1)}, M_{ij}^{(2)}\} \leq P\{X_{ij}^{(1)} = x_{ij}^{(1)} | C_{ij}, \bar{C}_{ij}, M_{ij}^{(1)}, M_{ij}^{(2)}\}\right\} \quad (10)$$

After performing the test for each combination of (i,j) we are left with the problem of combining all of the $NK = 450$ p-values to make an inference about H_0 defined in (6). We use two different approaches to address this issue. First note that if $X_{ij}^{(1)}$ had a continuous distribution then we could use the well-known result (see Fisher (1925)) that

$$-2 \ln \hat{\alpha}_{ij} \sim \chi^2[2] , \quad (11)$$

where $\chi^2[v]$ denotes a chi-square random variable with v degrees of freedom. We could then combine the tests by noting that

$$Q \equiv \sum_{i=1}^{N} \sum_{j=1}^{K} (-2 \ln \hat{\alpha}_{ij}) \quad (12)$$

is distributed as $\chi^2[2NK]$. Of course $X_{ij}^{(1)}$ is a discrete random variable but its conditional distribution is approximately continuous when C_{ij}, $\bar{C}_{ij}$, $M_{ij}^{(1)}$, and $M_{ij}^{(2)}$ are all large.

Another approach to combining the p-values is to note that under H_0, each $\hat{\alpha}_{ij}$ is independent and has probability .95 of being greater than .05 (again assuming continuity). Thus if we define

$$W = \frac{\#\text{ of } \hat{\alpha}_{ij} \text{ greater than } .05}{NK} \quad (13)$$

then W is distributed as $B(NK, .95)$ if $\theta_{ij}^{(1)} = \theta_{ij}^{(2)}$, for all i,j.

Table 2 summarizes the results for twenty different random splittings of the 234 requirements. The splittings were obtained using GGPER, an IMSL subroutine which generates random permutations, and the $\hat{\alpha}_{ij}$ were computed using IMSL subroutine CTPR. All computations were performed on an IBM 370 computer.

If we first consider the left part of the table, which includes all cases for which Fisher's test is meaningful, we see that all values of W are greater than .95, and all values of Q are less than 788 , the number of degrees of freedom of the reference chi-square distribution. These results clearly provide no cause for rejecting the hypothesis of equal probabilities in each half. Looking at the right part of the table, we note that for some splits of the requirements, W is less than .95. Seven of the twenty values of W are below .95 but none are more than 1.6 standard deviations below, where

$$SE(W) = \sqrt{(.95)\ (.05)/(\#\text{of applicable tests})} \ . \qquad (14)$$

Using the Bonferroni method to set the overall significance level at .10, none of the twenty W's provides significant evidence against the hypothesis of equal probabilities in both halves. Note also that all twenty of the Q's are less than the corresponding degrees of freedom which is clearly consistent with the null hypothesis.

In summary, we find that the assumption of equal probabilities for all requirements in the same category to be quite tenable. Using the category model we now develop an empirical Bayes procedure to improve the estimates of the category probabilities.

4. Empirical Bayes Estimates of the Category Probabilities

Our chief interest is in estimating $\underset{\sim}{\theta}_1,\ldots,\underset{\sim}{\theta}_N$, a large number of multivariate parameters. Since these are not unrelated parameters, empirical Bayes or James-Stein type (see James and Stein (1961)) estimators are candidates for reducing the overall error of estimation of the N parameter vectors. A lucid exposition of the applicability of empirical Bayes estimators is given by Morris (1983).

Table 2

Combining Fisher's Tests Across All Evaluations and Categories

Split	Number of Applicable Tests*	Q*	W*	Number of Applicable Tests**	Q**	W**
1	394	426.9	.992	55	77.8	.982
2	394	463.0	1.000	55	74.6	1.000
3	393	505.3	.987	51	113.5	.922
4	394	428.7	.990	52	65.9	1.000
5	394	506.3	.980	53	85.0	.981
6	394	456.1	.977	47	65.7	.979
7	394	570.3	.975	50	73.2	.980
8	394	463.2	.987	45	68.6	.978
9	394	499.5	.977	55	84.1	.964
10	394	407.3	.982	55	75.3	.964
11	394	431.9	.987	50	69.4	.940
12	394	421.9	.992	55	91.0	1.000
13	394	483.8	.975	44	68.4	.932
14	394	469.5	.997	58	113.8	.983
15	394	520.9	.967	50	95.7	.920
16	394	424.7	.982	54	87.0	.926
17	394	410.5	.980	58	58.6	.983
18	394	445.8	.982	45	68.0	.933
19	394	503.4	.975	51	106.4	.902
20	394	462.1	.982	55	78.4	.964

*Cases in which any of C_{ij}, $\bar{C}_{ij}$, $M_{ij}^{(1)}$ or $M_{ij}^{(2)}$ are 0 lead to a degenerate distribution for $X_{ij}^{(1)}$. These cases were omitted.

**Only cases in which $C_{ij}\, M_{ij}^{(\ell)}/M_{ij} > 5$, $\ell = 1,2$ and $\bar{C}_{ij}\, M_{ij}^{(\ell)}/M_{ij} > 5$, $\ell = 1,2$, are included.

The basic assumptions we have made are that different evaluations are independent, i.e. $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_N$ are independent random variables, and that each X_{ij}, given θ_i, is distributed as $B(M_{ij},\theta_{ij})$. If the M_{ij} are all moderately large, then the X_{ij} are approximately normal and many well-known results can be applied. To further simplify the analysis we make use of the variance stabilizing transformation

$$Y_{ij} = 2 \arcsin \left(\sqrt{(X_{ij} + .5)/(M_{ij} + 1)} \right) . \tag{15}$$

Y_{ij} is distributed approximately as a normal variable with mean

$$\eta_{ij} = 2 \arcsin \left(\sqrt{\theta_{ij}}\right) \tag{16}$$

and variance M_{ij}^{-1} (see Johnson and Kotz (1970)). The usefulness of this transformation is that the variance of Y_{ij} does not depend on any unknown parameters. In this transformed scale assumption (2) implies that $Y_{i1},\ldots,Y_{iK}$ are independent, given $\eta_{i1},\ldots,\eta_{iK}$. Let $\underset{\sim}{Y}_i = (Y_{i1},\ldots,Y_{iK})$ and $\underset{\sim}{\eta}_i = (\eta_{i1},\ldots,\eta_{iK})$. We summarize the distributional assumptions by stating that

$$\underset{\sim}{Y}_i \,|\, \underset{\sim}{\eta}_i \sim N(\underset{\sim}{\eta}_i, D_i) , \quad i=1,\ldots,N , \tag{17}$$

where

$$\underset{KxK}{D_i} = \text{diag} \{M_{i1}^{-1},\ldots,M_{iK}^{-1}\} . \tag{18}$$

A Bayesian model postulates that the vectors $\underset{\sim}{\eta}_i$ are independently drawn from a (prior) distribution which, in our context, can be thought of as representing the "superpopulation" of combat units. It is convenient to assume that the $\underset{\sim}{\eta}_i$ are drawn from a normal distribution,

i.e., $\underset{\sim}{\eta}_1,\ldots,\underset{\sim}{\eta}_N$ are i.i.d. vectors

$$\underset{\sim}{\eta}_i \sim N(\underset{\sim}{\mu},\Sigma) , \quad i=1,\ldots,N , \tag{19}$$

where $\underset{\sim}{\mu}$ is an arbitrary mean vector and Σ is an arbitrary positive definite covariance matrix. It should be noted here that other prior distributions are possible. Leonard (1972) considers models where $\log(\theta_{ij}/(1-\theta_{ij}))$ is normally distributed; Good (1965) allows θ_{ij} to have a beta distribution. While these alternative models are somewhat tractable in a single category situation, it is difficult to incorporate a dependence structure into the distribution of $\underset{\sim}{\theta}_i$ in either case. Since the M_{ij} are large enough for the assumption of normality to be reasonable, there are no compelling reasons, in this application, for using either of these alternative models.

Using (17) and (18), the predictive or unconditional distribution of $\underset{\sim}{Y}_i$ is $N(\underset{\sim}{\mu}, \Sigma + D_i)$ and the posterior distribution of $\underset{\sim}{\eta}_i$ is

$$\underset{\sim}{\eta}_i \,|\, \underset{\sim}{Y}_i \sim N(\underset{\sim}{Y}_i - B_i(\underset{\sim}{Y}_i - \underset{\sim}{\mu}_i) , (I-B_i)D_i) , \tag{20}$$

where

$$B_i \equiv D_i(D_i + \Sigma)^{-1} . \tag{21}$$

If $\underset{\sim}{\mu}$ and Σ were known, the Bayes estimate of $\underset{\sim}{\eta}_i$, assuming squared error loss, would be the posterior mean,

$$\hat{\underset{\sim}{\eta}}_i = \underset{\sim}{Y}_i - B_i(\underset{\sim}{Y}_i - \underset{\sim}{\mu}) . \tag{22}$$

This estimator has been proposed by Efron and Morris (1972) for the case in which $D_1,\ldots,D_N$ are equal.

In order to use the estimator in (22), we need estimates of $\underset{\sim}{\mu}$ and Σ. The unconditional distribution of $\underset{\sim}{Y}$ leads to the following simple estimates:

$$\underset{\sim}{\mu}^* = \bar{\underset{\sim}{Y}} \quad . \tag{23}$$

$$\Sigma^* = N^{-1} \sum_{i=1}^{N} (\underset{\sim}{Y}_i - \bar{\underset{\sim}{Y}})(\underset{\sim}{Y}_i - \bar{\underset{\sim}{Y}})' - D \ , \tag{24}$$

where $\bar{\underset{\sim}{Y}} = N^{-1} \sum_{i=1}^{N} \underset{\sim}{Y}_i$, $D = N^{-1} \sum_{i=1}^{N} D_i$. Although these estimates are not unreasonable, they are not the most efficient ones available and Σ^* has the added drawback of allowing negative estimates of variances. We instead suggest estimates of $\underset{\sim}{\mu}$, Σ derived by maximizing the predictive likelihood of $\underset{\sim}{Y}_1,\ldots,\underset{\sim}{Y}_N$. This likelihood function, $L(\underset{\sim}{\mu}, \Sigma; \underset{\sim}{Y}_1,\ldots,\underset{\sim}{Y}_N)$ is proportional to:

$$\prod_{i=1}^{N} |\Sigma + D_i|^{-1/2} \exp\{-\frac{1}{2}(\underset{\sim}{Y}_i - \underset{\sim}{\mu})'(\Sigma + D_i)^{-1}(\underset{\sim}{Y}_i - \underset{\sim}{\mu})\} \ . \tag{25}$$

Instead of directly maximizing (25) we will make use of the EM algorithm of Dempster, Laird, and Rubin (1977). The EM algorithm is designed for missing data situations but it can be applied here if we think of $\underset{\sim}{\eta}_1,\ldots,\underset{\sim}{\eta}_N$ as the missing "data." The algorithm consists of an E-step, computing the expectations of the sufficient statistics $(\underset{\sim}{\eta}_1,\ldots,\underset{\sim}{\eta}_N)$ given the observed data together with the current estimates of the parameters $(\underset{\sim}{\mu},\Sigma)$, and an M-step in which the likelihood of the complete data based on the estimated sufficient statistics is maximized. To define the algorithm more specifically, let $\hat{\underset{\sim}{\mu}}^{(p)}$, $\hat{\Sigma}^{(p)}$ be the current estimates of $(\underset{\sim}{\mu}, \Sigma)$ after p iterations. From (22) and the independence of $\underset{\sim}{\eta}_1,\ldots,\underset{\sim}{\eta}_N$, the E-step is given by

$$\hat{\underset{\sim}{\eta}}_i^{(p+1)} = E\left(\underset{\sim}{\eta}_i \middle| \hat{\underset{\sim}{\mu}}^{(p)}, \hat{\Sigma}^{(p)}, \underset{\sim}{Y}_1,\ldots,\underset{\sim}{Y}_N\right) = \underset{\sim}{Y}_i - \hat{B}_i^{(p)}\left(\underset{\sim}{Y}_i - \hat{\underset{\sim}{\mu}}^{(p)}\right) . \qquad (26)$$

From standard results, if $\hat{\underset{\sim}{\eta}}_1^{(p+1)},\ldots,\hat{\underset{\sim}{\eta}}_N^{(p+1)}$ were in fact the observed values of $\underset{\sim}{\eta}_1,\ldots,\underset{\sim}{\eta}_N$, the maximum likelihood estimates of $\underset{\sim}{\mu}$, Σ would be

$$\hat{\underset{\sim}{\mu}}^{(p+1)} = N^{-1} \sum_{i=1}^{N} \hat{\underset{\sim}{\eta}}_i^{(p+1)} \qquad (27)$$

$$\hat{\Sigma}^{(p+1)} = N^{-1} \sum_{i=1}^{N} \left(\hat{\underset{\sim}{\eta}}_i^{(p+1)} - \hat{\underset{\sim}{\mu}}^{(p+1)}\right)\left(\hat{\underset{\sim}{\eta}}_i^{(p+1)} - \hat{\underset{\sim}{\mu}}^{(p+1)}\right)' . \qquad (28)$$

Equations (26) - (28) define the algorithm for our problem. Using $\underset{\sim}{\mu}^*$, Σ^*, defined by (23), (24) as initial estimates led to convergence of the algorithm (to four decimal places in all components of $\underset{\sim}{\mu}$ and Σ) in nine iterations. Table 3 gives the initial and final estimates. Note that $\hat{\underset{\sim}{\mu}}$ and $\underset{\sim}{\mu}^*$ are very close but there are substantial differences between $\hat{\Sigma}$ and Σ^*.

Using (22) we obtain the empirical Bayes estimates

$$\hat{\underset{\sim}{\eta}}_i = \underset{\sim}{Y}_i - \hat{B}_i\left(\underset{\sim}{Y}_i - \hat{\underset{\sim}{\mu}}\right) , \quad i=1,\ldots,N , \qquad (29)$$

where $\hat{B}_i = D_i\left(D_i + \hat{\Sigma}\right)^{-1}$. Our interest is in estimating $\underset{\sim}{\theta}_i$ and the natural estimates to take are

$$\hat{\theta}_{ij} = \left[\sin\left(\eta_{ij}/2\right)\right]^2 , \quad i=1,\ldots,N \; ; \; j=1,\ldots,K , \qquad (30)$$

although these are not exactly the Bayes estimates of $\{\theta_{ij}\}$ for the assumed model.

Table 3

Initial and Final Estimates of $\underset{\sim}{\mu}$, Σ

$\underset{\sim}{\mu}^*$: 2.3916 2.3837 2.2332 2.4693 2.2211 2.4081

$\hat{\underset{\sim}{\mu}}$: 2.3943 2.3843 2.2338 2.4691 2.2233 2.4097

Σ^*:

$$\begin{bmatrix} .0693 & .0426 & .0648 & .0558 & .0597 & .0428 \\ & .0087 & .0474 & .0475 & .0616 & .0376 \\ & & .0856 & .0809 & .0574 & .0449 \\ & & & .0866 & .0589 & .0509 \\ & & & & .1182 & .0553 \\ & & & & & .0471 \end{bmatrix}$$

$\hat{\Sigma}$:

$$\begin{bmatrix} .0461 & .0370 & .0557 & .0573 & .0541 & .0399 \\ & .0297 & .0447 & .0460 & .0434 & .0321 \\ & & .0673 & .0692 & .0653 & .0482 \\ & & & .0712 & .0672 & .0496 \\ & & & & .0634 & .0468 \\ & & & & & .0346 \end{bmatrix}$$

5. Cross-Validation of the Bayesian Model

The motivation for using empirical Bayes estimates instead of estimating the probabilities with the relative frequencies is the hope that the overall error across all units will be smaller. Other empirical studies (see Efron and Morris (1975) and Fay and Herriot (1979)) have demonstrated that empirical Bayes techniques do lead to improved estimation. In both of these studies the true parameter values were available so the actual error of the estimates could be computed. We do not have this luxury so we will again use cross-validation to compare the Bayes estimates with the sample proportions.

As described in section 2, we split the 234 requirements into two halves. The first half was used to estimate the probabilities and the relative frequencies in the second half were compared with these estimates. Define $\hat{p}^B_{ij}$ to be the Bayes estimate computed from the first half and let $\hat{p}_{ij}$ denote the (modified) sample proportion from the first half, i.e.

$$\hat{p}_{ij} = (X^{(1)}_{ij} + .5)/(M^{(1)}_{ij} + 1) \ . \tag{31}$$

Let p_{ij} denote the (modified) sample proportion from the second half,

$$p_{ij} = (X^{(2)}_{ij} + .5)/(M^{(2)}_{ij} + 1) \ . \tag{32}$$

For each unit we would like measures of agreement between $\hat{\underset{\sim}{p}}^B_i$ and $\underset{\sim}{p}_i$ and between $\hat{\underset{\sim}{p}}_i$ and $\underset{\sim}{p}_i$. We propose two measures:

$$C^B_i = \sum_{j=1}^{K} \frac{M^{(2)}_{ij} \ (P_{ij} - \hat{p}^{(B)}_{ij})^2}{\hat{p}^B_{ij} \ (1 - \hat{p}^B_{ij})} \ , \tag{33}$$

$$L^B_i = 2 \sum_{j=1}^{K} \{M^{(2)}_{ij} \ p_{ij} \log (p_{ij}/\hat{p}^B_{ij}) + M^{(2)}_{ij} \ (1-p_{ij}) \log ((1-p_{ij})/(1-\hat{p}^B_{ij}))\}. \tag{34}$$

In a similar manner define C_i and L_i by replacing $\hat{p}^B_{ij}$ with $\hat{p}_{ij}$ in (33) and (34). Note that C_i is the Pearson chi-square statistic and L_i the likelihood-ratio statistic for comparing $\hat{\underset{\sim}{p}}_i$ and $\underset{\sim}{p}_i$. By using these measures we are attempting to incorporate the variability of $\underset{\sim}{p}$ into the comparisons. Bishop, Fienberg, and Holland (1975) give a complete discussion of the merits of using these statistics to compare observed and estimated probabilities.

Table 4 summarizes our findings. We have summed C_i, C^B_i L_i, and L^B_i over the N=75 evaluations. The results are strikingly in favor of the Bayes estimates. For every one of the 20 splits of the requirements the Bayes estimates produced a smaller total "error" as measured by either C or L. We have, for a number of splits, looked at the 75 values of C and L for each unit and it is clear that the Bayes estimators are providing protection against gross errors in estimating p_{ij}, as the large values of C_i or L_i are almost exclusively observed for $\hat{p}_{ij}$ as opposed to $\hat{p}^B_{ij}$.

In summary, we have found that the empirical Bayes methodology leads to a very tractable means of estimating a collection of vector parameters and offers a considerable improvement over the typical method of estimating each vector separately. To our knowledge this is the first empirical study that assesses the efficacy of empirical Bayes estimators for vector parameters. We hope that it leads to a greater usage of these techniques in the future.

Table 4

Cross-Validation of Bayes and Ordinary Estimates of Probabilities

Split	$\sum_{i=1}^{75} C_i$	$\sum_{i=1}^{75} C_i^B$	$\sum_{i=1}^{75} L_i$	$\sum_{i=1}^{75} L_i^B$
1	824.3	539.2	614.6	506.7
2	898.5	547.2	702.7	512.8
3	788.1	563.7	678.0	555.1
4	784.9	572.6	596.5	513.8
5	809.2	533.9	709.9	552.8
6	780.4	439.6	667.2	470.4
7	906.3	614.9	760.4	587.6
8	818.3	620.3	686.1	594.9
9	815.0	582.8	704.4	565.8
10	895.7	665.4	655.0	558.3
11	863.0	587.3	657.9	533.4
12	970.2	588.8	637.2	511.0
13	850.5	563.5	673.8	508.6
14	704.8	541.1	656.4	565.1
15	1213.8	735.9	814.7	636.8
16	989.4	695.7	702.6	620.0
17	880.6	554.8	606.9	495.5
18	956.5	586.6	753.2	564.2
19	1125.1	611.8	793.9	570.8
20	914.7	574.1	652.2	529.1

REFERENCES

BARZILY, Z. (1980). Analyzing MCCRES data. Technical Paper Serial T-427.

BISHOP, Y. M. M., S. E. FIENBERG, and P. W. HOLLAND (1975). Discrete Multivariate Analysis. MIT Press.

DEMPSTER, A. P., N. M. LAIRD, and D. B. RUBIN (1977). Maximum likelihood from incomplete data via the EM algorithm. J. Royal Statis. Soc. (B), 39, pp. 1-38.

EFRON, B. and C. MORRIS (1972). Empirical Bayes on vector observations: an extension of Stein's method. Biometrika 59, pp. 335-347.

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

FAY, R. E. and R. A. HERRIOT (1979). Estimates of income for small places: an application of James-Stein procedures to census data. J. Amer. Statist. Assoc. 74, pp. 269-277.

FISHER, R. A. (1925). Statistical Methods for Research Workers. Hafner Press.

GOOD, I. J. (1965). The Estimation of Probabilities. MIT Press.

JAMES, W. and C. STEIN (1961). Estimation with quadratic loss. Proceedings Fourth Berkeley Symposium 1, pp. 361-79.

JOHNSON, N. L. and S. I. KOTZ (1970). Discrete Distributions. Wiley.

LEONARD, T. (1972). Bayesian methods for binomial data. Biometrika 59, pp. 581-589.

MORRIS, C. (1983). Parametric empirical Bayes inference: theory and applications. J. Amer. Statist. Assoc. 78, pp. 47-54.

ZACKS, S., W. H. MARLOW, and Z. BARZILY (1981). Category analysis of the Marine Corps Combat Readiness Evaluation System. Technical Paper Serial T-450.

ZACKS, S. and W. H. MARLOW (1982). Estimating the structural parameters of the Marine Corps Combat Readiness Evaluation System on the basis of the primary categories model. Technical Memorandum Serial TM-69200.

Copies

Director
AMSAA
Attn: DRXSY-MP, H. Cohen
Aberdeen Proving Ground, MD 21005 1

Dr. Gerhard Heiche
Naval Air Systems Command (NAIR 03)
Jefferson Plaza No. 1
Arlington, VA 20360 1

Dr. Barbara Bailar
Associate Director, Statistical Standards
Bureau of Census
Washington, DC 20233 1

Leon Slavin
Naval Sea Systems Command (NSEA 05H)
Crystal Mall #4, Rm. 129
Washington, DC 20036 1

B. E. Clark
RR #2, Box 647-B
Graham, NC 27253 1

Naval Underwater Systems Center
Attn: Dr. Derrill J. Bordelon
Code 601
Newport, Rhode Island 02840 1

Naval Coastal Systems Center
Code 741
Attn: Mr. C. M. Bennett
Panama City, FL 32401 1

Naval Electronic Systems Command (NELEX 612)
Attn: John Schuster
National Center No. 1
Arlington, VA 20360 1

Defense Logistics Studies Information Exchange
Army Logistics Management Center
Attn: Mr. J. Dowling
Fort Lee, VA 23801 1

Copies

Reliability Analysis Center (RAC)
RADC/RBRAC
Attn: I. L. Krulac
Data Coordinator/
Government Programs
Griffiss AFB, New York 13441 1

Technical Library
Naval Ordnance Station
Indian Head, MD 20640 1

Library
Naval Ocean Systems Center
San Diego, CA 92152 1

Technical Library
Bureau of Naval Personnel
Department of the Navy
Washington, DC 20370 1

Mr. Dan Leonard
Code 8105
Naval Ocean Systems Center
San Diego, CA 92152 1

Dr. Alan F. Petty
Code 7930
Naval Research Laboratory
Washington, DC 20375 1

Dr. M. J. Fischer
Defense Communications Agency
Defense Communications Engineering Center
1860 Wiehle Avenue
Reston, VA 22090 1

Mr. Jim Gates
Code 9211
Fleet Material Support Office
U. S. Navy Supply Center
Mechanicsburg, PA 17055 1

Mr. Ted Tupper
Code M-311C
Military Sealift Command
Department of the Navy
Washington, DC 20390 1

	Copies	Copies
Mr. F. R. Del Priori Code 224 Operational Test and Evaluation Force (OPTEVFOR) Norfolk, VA 23511	1	