The Annals of Statistics
1979, Vol. 7, No. 2, 284–290

ADMISSIBLE AND MINIMAX ESTIMATION FOR THE MULTINOMIAL DISTRIBUTION AND FOR K INDEPENDENT BINOMIAL DISTRIBUTIONS

BY INGRAM OLKIN AND MILTON SOBEL

Stanford University and University of California, Santa Barbara

Admissible and minimax estimation is discussed for estimating the parameters in the (a) multinomial distribution and in (b) k independent binomial distributions. In (a) the loss function is $\Sigma_0^n[\delta_i(x) - \theta_i]^2/\theta_i$, where $\theta_0, \ldots, \theta_k (\Sigma\theta_i = 1)$ are the parameters in the multinomial distribution, and the estimators are restricted to $\Sigma_0^k\delta_i(x) = 1$. In (b) the loss functions considered are the weighted sum of quadratic losses. The method of proof is based on a multivariate analog of the Cramér-Rao inequality, and uses the divergence theorem in a novel way.

1. Introduction.

The following results are obtained.

Model I. Let $X = (X_0, X_1, \ldots, X_k)$ have a multinomial distribution with parameters $n, \theta_0, \theta_1, \ldots, \theta_k, \theta_i \geqslant 0 (i = 0, 1, \ldots, k), \Sigma_0^k\theta_i = 1, \Sigma_0^k X_i = n$. The estimator

$$\delta_i(X) = \frac{X_i}{n}, i = 1, \ldots, k, \delta_0(X) = 1 - \frac{\Sigma_1^k X_i}{n}$$

is admissible and minimax for the loss function

$$(1.1) \qquad L(\delta(X), \theta) = (\delta(X) - \theta)\Sigma^{-1}(\delta(X) - \theta)', 0 < \theta_i < 1, \Sigma_1^k\theta_i < 1,$$
$$= \infty \qquad \text{otherwise,}$$

where $\delta(X) \equiv (\delta_1(X), \ldots, \delta_k(X)), \theta = (\theta_1, \ldots, \theta_k), \Sigma = (\sigma_{ij}), \sigma_{ii} = \theta_i(1 - \theta_i)$, $\sigma_{ij} = -\theta_i\theta_j (i \neq j), i, j = 1, \ldots, k$.

Note that $\Sigma = D_\theta - \theta'\theta$, where $D_\theta = \text{diag}(\theta_1, \ldots, \theta_k)$ so that $\Sigma^{-1} = D_\theta^{-1} + e'e/\theta_O$, where $e = (1, \ldots, 1)$. Hence an alternative expression for (1.1) is

$$L(\delta(X), \theta) = \Sigma_0^k \frac{(\delta_i(X) - \theta_i)^2}{\theta_i}, 0 < \theta_i < 1, i = 0, 1, \ldots, k,$$
$$= \infty \qquad \text{otherwise,}$$

where $\Sigma_0^k\delta_i(X) = 1$.

Model II. Let $X = (X_1, \ldots, X_k)$ be k independent binomial random variables with parameters $n, \theta_i, i = 1, \ldots, k$. The estimator

$$\delta_i(X) \equiv \frac{X_i + \frac{1}{2}n^{\frac{1}{2}}}{n + n^{\frac{1}{2}}}, \qquad i = 1, \ldots, k$$

Received February 1978; revised April 1978.
AMS 1970 *subject classifications.* 62C15, 62H15.
Key words and phrases. Admissible estimators, minimax estimators, multinomial distribution, independent binomial distributions, multivariate Cramér-Rao inequality, divergence theorem.

is admissible and minimax for the loss function

$$L(\delta(X), \theta) = \Sigma_1^k(\delta_i(X) - \theta_i)^2.$$

Model III. Let $X = (X_1, \ldots, X_k)$ be k independent binomial random variables with parameters n_i, $\theta_i (0 \leqslant \theta_i \leqslant 1)$, $i = 1, \ldots, k$. The estimator

$$\delta_i(X) = X_i / n_i, \qquad i = 1, \ldots, k$$

is admissible and minimax for the loss function

$$L(\delta(X)\theta) = \Sigma_1^k \frac{(\delta_i(X) - \theta_i)^2}{\theta_i(1 - \theta_i)/n_i}, \; 0 < \theta_i < 1, \qquad i = 1, \ldots, k,$$
$$= \infty \qquad \text{otherwise.}$$

Results for the univariate case ($k = 1$) in Model II were obtained by Hodges and Lehmann (1951). They developed a general procedure based on the Cramér-Rao inequality that yields admissibility results for a variety of univariate models.

Minimax estimators for the multinomial distribution with loss functions of the form $\Sigma_0^k c_i(\delta_i(X) - \theta_i)^2$, where the c_i are constants and $\Sigma_0^k \delta_i(X) = 1$, were obtained by Steinhaus (1957), by Trybula (1958), and by Rutkowska (1977). Johnson (1971) shows that $\delta_i(X) = X_i / n$, $i = 0, 1, \ldots, n$ is admissible (with squared error loss) for the multinomial distribution; an alternative proof is given by Alam (1978). Johnson (1971) also discusses Model II in a more global setting.

The present approach is through a multivariate extension of the Hodges-Lehmann (1951) procedure. An analog of the Cramér-Rao inequality is developed that is a direct extension of the development in Lehmann (1950). Multivariate versions of the Cramér-Rao inequality also appear in Stein (1955, 1973) in connection with estimating the mean of a normal distribution.

The present work was actually completed in 1961. At that time there was not the interest in these problems as there is today, and the work remained unpublished. There is another aspect of the proofs in this paper that may be of interest, namely, the use of the divergence theorem in proving admissibility. Recently Stein (1973) and Haff (1977a, b) and Hudson (1978) make a different type of use of the divergence theorem in the context of unbiased estimators of the risk.

2. Preliminaries. The extension of the Cramér-Rao inequality is based on the fact that if

$$(\delta; S) \equiv (\delta_1, \ldots, \delta_l, S_1, \ldots, S_l)$$

is a random vector with covariance matrix

$$\begin{pmatrix} \mathrm{Cov}(\delta, \delta) & \mathrm{Cov}(\delta, S) \\ \mathrm{Cov}(S, \delta) & \mathrm{Cov}(S, S) \end{pmatrix} \equiv \begin{pmatrix} M & N \\ N' & \Lambda \end{pmatrix},$$

then $M - N\Lambda^{-1}N'$ is positive semidefinite. Consequently, for any positive semidefinite matrix A we have

$$\text{(2.1)} \qquad \operatorname{tr} AM \geqslant \operatorname{tr} AN\Lambda^{-1}N'.$$

The loss function

$$(2.2) \qquad L(\delta(X), \theta) = (\delta(X) - \theta)A(\delta(X) - \theta)'$$

defines the A matrix in each of the three separate problems described above. Hence, in all three cases the risk function is

(2.3)

$$\begin{aligned} r_\delta(\theta) &= EL(\delta(X), \theta) \\ &= E(\delta(X) - E\delta(X))A(\delta(X) - E\delta(X))' + (E\delta(X) - \theta)A(E\delta(X) - \theta)' \\ &= \operatorname{tr} A \operatorname{Cov}(\delta, \delta) + bAb', \end{aligned}$$

where $b_i \equiv b_i(\theta_1, \ldots, \theta_l) = \mathrm{E}(\delta_i(X) - \theta_i)$, $i = 1, \ldots, l$. Let $S_j = \partial \log L(X, \theta)/\partial\theta_j$, $N = (\nu_{ij})$, $i, j = 1, \ldots, l$, then

$$\nu_{ij} = \delta_{ij} + \partial b_i(\theta)/\partial\theta_j, \qquad i, j = 1, \ldots, l,$$

where δ_{ij} is the Kronecker delta. Equivalently,

$$(2.4) \qquad N = I + Q,$$

where $Q = q_{ij}$, $q_{ij} = \partial b_i(\theta)/\partial\theta_j$. Then (2.3), (2.1) and (2.4) yield

$$(2.5) \qquad \begin{aligned} r_\delta(\theta) &= \operatorname{tr} AM + bAb' \\ &\geqslant \operatorname{tr} AN\Lambda^{-1}N' + bAb' \\ &= \operatorname{tr} A(I + Q)\Lambda^{-1}(I + Q)' + bAb'. \end{aligned}$$

If δ^* is an estimator for which equality in (2.5) is achieved, and if for every bias vector $b(\theta)$,

(2.6)

$$\operatorname{tr} A(I + Q)\Lambda^{-1}(I + Q)' + bAb' \leqslant \operatorname{tr} A(I + Q^*)\Lambda^{-1}(I + Q^*)' + b^*Ab^{*\prime}$$

for all θ implies that $b(\theta) \equiv b^*$, then δ^* is admissible. Note that $\Lambda = \operatorname{Cov}(S, S)$ is independent of δ. If, in addition to admissibility, the risk is constant, then δ^* is minimax.

Remark. We have defined $S_j = [L(X, \theta)]^{-1}\partial L(X, \theta)/\partial\theta_j$. However, higher derivatives, $[L(X, \theta)]^{-1}\partial^m L(X, \theta)/\partial\theta_j^m$, can also be used to yield a multivariate version of the Bhattacharya inequalities. Hence if θ has q components and S has l components we can take $l \geqslant q$ in some of our models. However, in the present discussion the use of first derivatives with $q = l$ suffices.

3. Multinomial distribution-Model I. Under Model I the loss function is $L(\delta(X), \theta) = (\delta(X) - \theta)\Sigma^{-1}(\delta(X) - \theta)'$, so that the matrix A of (2.2) is Σ^{-1}. The evaluation of $\Lambda = (\lambda_{ij})$ is straightforward and follows from $\lambda_{ij} = \operatorname{Var}(S_i) = n(\theta_i^{-1} + \theta_0^{-1})$, $i = 1, \ldots, k$, $\theta_0 = 1 - \Sigma_1^k\theta_i$, $\lambda_{ij} = \operatorname{Cov}(S_i, S_j) = n\theta_0^{-1}$, $i \neq j$, so that $\Lambda = n(D_\theta^{-1} + e'e/\theta_0)$, where $e = (1, \ldots, 1)$. Since $\Sigma = D_\theta - \theta'\theta$, it follows that $\Lambda = n\Sigma^{-1}$.

Remark. In this development we restrict our attention to estimators $\delta(X)$ of the form $\Sigma_0^k\delta_i(X) = 1$, so that $\delta_0(X)$ is determined from $(\delta_1(X), \ldots, \delta_k(X))$.

Consequently, in what follows only $\delta_1(X), \ldots, \delta_k(X)$ appear. This also has the advantage that the $k \times k$ covariance matrix Σ is nonsingular. An alternative approach is to use the complete vector $\delta_0(X), \delta_1(X), \ldots, \delta_k(X)$ with a singular $(k+1) \times (k+1)$ covariance matrix and then use a generalized inverse.

With $\delta^*(X) = (X_1/n, \ldots, X_k/n)$, (2.6) becomes

$$r_\delta(\theta) = \frac{1}{n} \operatorname{tr} \Sigma^{-1}(I + Q)\Sigma(I + Q)' + b\Sigma^{-1}b' \leqslant \frac{k}{n} = r_{\delta *}(\theta). \tag{3.1}$$

We need to show that (3.1) for all $0 \leqslant \theta_i$, $i = 1, \ldots, k$, $\Sigma_1^k \theta_i \leqslant 1$, implies that $b(\theta) \equiv 0$.

Since $b\Sigma^{-1}b' \geqslant 0$, (3.1) implies that

$$\frac{1}{n} \operatorname{tr} I_k + \frac{2}{n} \operatorname{tr} Q + \frac{1}{n} \operatorname{tr} \Sigma^{-1} Q \Sigma Q' \leqslant \frac{k}{n},$$

so that

$$2 \operatorname{tr} Q + \operatorname{tr} \Sigma^{-1} Q \Sigma Q' \leqslant 0.$$

The second term is nonnegative, which implies that

$$\operatorname{tr} Q = \Sigma_1^k \frac{\partial b_i}{\partial \theta_i} \leqslant 0. \tag{3.2}$$

We now show that $\Sigma_1^k \partial b_i / \partial \theta_i = 0$. To accomplish this we make use of the divergence theorem. Equation (5.1) implies that

$$b\Sigma^{-1}b' = \Sigma_1^k \frac{b_i^2}{\theta_i} + \frac{(\Sigma b_i)^2}{\theta_0} \leqslant \frac{k}{n} \tag{3.3}$$

in the interior of $\Omega = \{\theta : 0 \leqslant \theta_i, i = 1, \ldots, k, \Sigma_1^k \theta_i \leqslant 1\}$. Consequently, $b_i^2 \leqslant \theta_i k/n$, $i = 1, \ldots, k$, and $(\Sigma_1^k b_i)^2 \leqslant \theta_0 k/n$, so that as we approach any point on the boundary $\mathfrak{B}$ of Ω, the corresponding b-component(s) must approach zero, i.e.,

$$\begin{aligned} b_i(\theta) &= 0 \quad \text{when} \quad \theta_i = 0, \quad i = 1, \ldots, k, \\ \Sigma_1^k b_i(\theta) &= 0 \quad \text{when} \quad \theta_0 = 0 \quad \left(\text{i.e., } \Sigma_1^k \theta_i = 1\right). \end{aligned} \tag{3.4}$$

Consider the region Ω with boundary $\mathfrak{B}$. By the divergence theorem

$$\int_{\mathfrak{B}} (\Sigma b_i(\theta) \cos \alpha_i) \, d\sigma = \int_\Omega \left(\Sigma_1^k \frac{\partial b_i}{\partial \theta_i} \right) dv. \tag{3.5}$$

From (3.4) and the fact that $\cos \alpha_i (i = 0, 1, \ldots, k)$ is 0 on the boundary points where $\theta_j (j \neq i)$ is 0 or 1, the left-hand side of (3.5) is 0. Hence

$$0 = \int_\Omega \left(\Sigma_1^k \frac{\partial b_i}{\partial \theta_i} \right) dv$$

which, together with (3.2), implies that $\operatorname{tr} Q = \Sigma_1^k \partial b_i / \partial \theta_i = 0$. Consequently, for all points in the interior of Ω, (3.1) reduces to

$$b\Sigma^{-1}b' + \frac{1}{n} \operatorname{tr} \Sigma^{-1} Q \Sigma Q' \leqslant 0. \tag{3.6}$$

Since the second term in (3.6) is nonnegative, (3.6) implies that

$$b\Sigma^{-1}b' \leqslant 0, \tag{3.7}$$

and hence $b(\theta) \equiv 0$ throughout the interior (as well as on the boundary) of Ω, which completes the proof.

4. k **Independent binomial distributions.** Under Models II and III the loss functions are $L(\delta(X), \theta) = \Sigma_1^k(\delta_i(X) - \theta_i)^2$ and $\Sigma_1^k n_i(\delta_i(X) - \theta_i)^2/\theta_i(1 - \theta_i)$, respectively. When $k = 1$ the estimator δ_1 is unique Bayes with respect to a beta prior (with appropriate values for the parameters of that prior). By independence and the form of the loss function, the estimator $\delta(X)$ is therefore unique Bayes with respect to a product of such priors. Therefore $\delta(X)$ is admissible. Further, since $\delta(X)$ has constant risk, it is minimax.

From the above argument, the results on admissibility and minimax can be regarded as known. However, we note that proofs based on the multivariate Cramér-Rao inequality can also be constructed for these two models, and we give a sketch of the details.

4.1. *Model II.* With $\delta_i^*(X) = (X_i + \frac{1}{2}n^{\frac{1}{2}})/(n + n^{\frac{1}{2}})$, $i = 1, \ldots, k$, (2.6) becomes

$$r_\delta(\theta) = \text{tr}(I + Q)D_a(I + Q)' + bb' \leqslant \frac{k}{4(1 + n^{\frac{1}{2}})^2} = r_{\delta *}(\theta), \tag{4.1}$$

where $D_a = \text{diag}(a_1, \ldots, a_k)$, $a_i = [\theta_i(1 - \theta_i)]^{-1}$, $i = 1, \ldots, k$. We need to show that (3.1), holding for all $0 \leqslant \theta_i \leqslant 1$, $i = 1, \ldots, k$, implies that

$$b_i(\theta) = \left(\tfrac{1}{2} - \theta_i\right)/\left(1 + n^{\frac{1}{2}}\right), \qquad i = 1, \ldots, k. \tag{4.2}$$

It is easily checked that δ^* has the bias given in (4.2).

To prove (4.2) note that

$$\text{tr}(I + Q)D_a(I + Q)' = \Sigma_1^k a_i + 2\Sigma_1^k q_{ii}a_i + \Sigma_1^k q_{ii}^2 a_i + \Sigma_{i \neq j} q_{ij}^2 a_j,$$

which implies that

$$\frac{1}{k}\Sigma_1^k\left\{b_i^2 + \frac{\theta_i(1 - \theta_i)}{n}\left[1 + \frac{\partial b_i}{\partial \theta_j}\right]^2\right\} \leqslant \frac{1}{4(1 + n^{\frac{1}{2}})^2}. \tag{4.3}$$

The left-hand side is the mean of k positive terms and, hence, at least one of the terms (say the first) must be less than or equal to the right-hand side. That is,

$$b_1^2 + \frac{\theta_1(1 - \theta_1)}{n}\left[1 + \frac{\partial b_1}{\partial \theta_1}\right]^2 \leqslant \frac{1}{4(1 + n^{\frac{1}{2}})^2}, \qquad 0 \leqslant \theta_1 \leqslant 1. \tag{4.4}$$

By the univariate binomial result of Hodges and Lehmann (1951), (4.4) implies that $b_1(\theta) \equiv (\frac{1}{2} - \theta_1)/(1 + n^{\frac{1}{2}})$ for which equality in (4.4) is achieved. Consequently, (4.3) becomes

$$\frac{1}{k-1}\Sigma_2^k\left\{b_i^2 + \frac{\theta_i(1-\theta_i)}{n}\left[1 + \frac{\partial b_i}{\partial \theta_i}\right]^2\right\} \leqslant \frac{1}{4(1+n^{\frac{1}{2}})^2}.$$

A repetitive argument then yields (4.2), which completes the proof.

4.2 *Model III.* With $\delta_i^* = X_i/n_i$, (2.6) becomes

$$(4.5)\qquad r_\delta(\theta) = \operatorname{tr} D_a^{-1}(I+Q)D_a(I+Q)' + bD_a^{-1}b' \leqslant k = r_{\delta*}(\theta).$$

We need to show that (4.5) for all $0 \leqslant \theta_i \leqslant 1$, $i = 1, \ldots, k$ implies that $b(\theta) \equiv 0$.

To prove that $b(\theta) \equiv 0$, note that (4.5) implies that

$$\operatorname{tr} D_a(I+Q)D_a^{-1}(I+Q)' = \operatorname{tr} I_k + 2\operatorname{tr} Q + \operatorname{tr} D_aQD_a^{-1}Q' \leqslant k,$$

so that

$$2\operatorname{tr} Q + \operatorname{tr} D_aQD_a^{-1}Q' \leqslant 0.$$

Since the second term is nonnegative, we must have that

$$(4.6)\qquad \operatorname{tr} Q = \Sigma_1^k\frac{\partial b_i}{\partial \theta_i} \leqslant 0.$$

We wish to show that $\Sigma_1^k \partial b_i/\partial\theta_i = 0$. To accomplish this we use the divergence theorem (3.5). Equation (4.5) implies that

$$(4.7)\qquad bD_a^{-1}b' = \Sigma_1^k\frac{b_i^2}{\theta_i(1-\theta_i)/n_i} \leqslant k,$$

$$0 < \theta_i < 1,\ i = 1, \ldots, k.$$

As any θ_i approaches the boundary, the denominator goes to zero, so that b_i must go to zero. Consequently,

$$b_i(\theta_1, \ldots, \theta_{i-1}, 0, \theta_{i+1}, \ldots, \theta_k) = b_i(\theta_1, \ldots, \theta_{i-1}, 1, \theta_{i+1}, \ldots, \theta_k) = 0,$$

$$i = 1, \ldots, k.$$

Using (4.7) and the fact that $\cos\alpha_i$ is 0 on the boundary for which $\theta_j (j \neq i)$ is 0 or 1, the left-hand side of (3.5) is 0. Hence by the divergence theorem

$$(4.8)\qquad 0 = \int_\Omega\left(\Sigma_1^k\frac{\partial b_i}{\partial\theta_i}\right)dv,$$

which together with (4.6) implies that $\operatorname{tr} Q = \Sigma_1^k \partial b_i/\partial\theta_i = 0$.

Consequently, (4.5) reduces to

$$(4.9)\qquad \operatorname{tr} D_aQD_a^{-1}Q' + \Sigma_1^k\frac{[b_i(\theta)]^2}{\theta_i(1-\theta_i)}n_i \leqslant 0, \qquad 0 < \theta_i < 1.$$

Each term on the left-hand side is nonnegative, so that each term must be zero. Thus, $b_i(\theta) = 0$, $i = 1, \ldots, k$ in the interior of Ω, which, together with (4.7), completes the proof.

Acknowledgment. The authors are indebted to L. Brown, J. Kiefer, J. Zidek and the Associate Editor for their comments and suggestions.

REFERENCES

ALAM, K. (1979). Estimation of multinomial probabilities. *Ann. Math. Statist.* **7** 301–302.

HAFF, L. R. (1977a). Minimax estimators for a multinormal precision matrix. *J. Multivariate Analysis* **7** 374–385.

HAFF, L. R. (1977b). Estimation of the inverse covariance matrix: random matrices of the inverse Wishart matrix and the identity. Unpublished manuscript.

HODGES, H. L. JR. and LEHMANN, E. L. (1951). Some applications of the Cramér-Rao inequality. *Proc. Second Berkeley Symp. Math. Statist. Probability.* 13–22. Univ. California Press, Berkeley and Los Angeles.

HUDSON, M. E. (1978). A natural identity for exponential families with applications in multiparameter estimation. *Ann. Statist.* **6** 473–484.

JOHNSON, B. MCK. (1971). On admissible estimators for certain fixed sample binomial problems. *Ann. Math. Statist.* **42** 1579–1587.

LEHMANN, E. L. (1950). *Notes on the Theory of Estimation.* Univ. California.

RUTKOWSKA, M. (1977). Minimax estimation of the parameters of the multivariate hypergeometric and multinomial distributions. *Zastos. Mat.* **16** 9–21.

STEIN, C. (1955). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proc. Third Berkeley Symp. Statist. Probability* **I** 197–206. Univ. California Press, Berkeley and Los Angeles.

STEIN, C. (1973). Estimation of the mean of a multivariate normal distribution. *Proc. Prague Symp. Asymptotic Statist.* **II** 345–381.

STEINHAUS, H. (1957). The problem of estimation. *Ann. Math. Statist.* **28** 633–648.

TRYBULA, S. (1958). Some problems of simultaneous minimax estimation. *Ann. Math. Statist.* **29** 245–253.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA
SANTA BARBARA, CALIFORNIA 93106