ESTIMATING ORDERED PROBABILITIES[1]

By Morris W. Katz

University of Wisconsin-Milwaukee

1. Summary. Let X_i, Y_i $i = 1, 2, \cdots n$ be mutually independent binomial random variables with $P\{X_i = 1\} = p_1$, $P\{Y_i = 1\} = p_2$. Let $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$ and $\bar{Y}_n = (1/n)\sum_{i=1}^n Y_i$. We wish to estimate the parameter $p = (p_1, p_2)$. If the parameter space is $\Omega = \{p \mid 0 \leqq p_1 \leqq 1, 0 \leqq p_2 \leqq 1\}$ then the usual non-sequential estimator $\delta = (\delta_1, \delta_2)$ is of the form $(f(\bar{X}_n)), f(\bar{Y}_n))$, that is, if no restriction is placed on the parameters, the estimator for the paired parameters is the pair of estimators for each parameter separately. In this paper, however, we are concerned with the parameter space $\Omega = \{p \mid 0 \leqq p_1 \leqq p_2 \leqq 1\}$. We show that estimators constructed as before are no longer admissible with respect to a class of reasonable loss functions. Square error loss is included in this class. In particular, we show that for such loss functions, estimators not ordered in the same way as the parameters are inadmissible. A class of estimators which retains the same ordering as the parameters, that is with $\delta_2 \geqq \delta_1$, is investigated and the asymptotic behavior of the minimax member is described. Finally, an asymptotic estimator based on a normal approximation is given. This estimator is minimax and admissible.

The problem of estimating ordered probabilities arises in estimating a section of an unknown distribution function.

2. Ordering requirement. Let X_i, Y_i, $i = 1, 2, \cdots n$, p and δ be as above, and let the parameter space be $\Omega = \{p \mid 0 \leqq p_1 \leqq p_2 \leqq 1\}$.

We note a well-known property of any strictly convex function ϕ:

$$\phi(s) + \phi(t) > \phi(s + u) + \phi(t - u) \tag{1}$$

where $t - s > u > 0$. We use this property in the alternate form:

$$\phi(a) + \phi(b) < \phi(a + u) + \phi(b - u) \tag{2}$$

where $a - b + 2u > u > 0$.

Theorem 1. *Let $\delta = (\delta_1, \delta_2)$ be an estimator of $p = (p_1, p_2)$ where $\Omega = \{p \mid 0 \leqq p_1 \leqq p_2 \leqq 1\}$. Let the loss function be of the form*

$$\phi(|\delta_1 - p_1|) + \phi(|\delta_2 - p_2|) \tag{3}$$

where ϕ is a convex, even, positive function. Then if $P_p\{\delta_1 > \delta_2\} > 0$, δ is inadmissible.

Proof. Define a new estimator $\delta^+ = (\delta_1^+, \delta_2^+)$ by

$$\delta_1^+ = \min\{\delta_1, \alpha\delta_1 + (1 - \alpha)\delta_2\}, \qquad \delta_2^+ = \max\{\delta_2, (1 - \alpha)\delta_1 + \alpha\delta_2\},$$

where $0 \leqq \alpha \leqq 1$.

Received August 9, 1961; revised April 2, 1963.

[1] This research was supported, in part, by the Wisconsin Alumni Research Foundation.

We show

$$\phi(\delta_1^+ - p_1) + \phi(\delta_2^+ - p_2) < \phi(\delta_1 - p_1) + \phi(\delta_2 - p_2) \tag{4}$$

whenever $\delta_1 > \delta_2$. For $\delta_1 \leqq \delta_2$ equality holds in (4). Suppose $\delta_1 > \delta_2$. Let $a = \delta_1^+ - p_1$, $b = \delta_2^+ - p_2$, $u = (\delta_1 - \delta_2)(1 - \alpha)$; then $a + u = \delta_1 - p_1$, $b - u = \delta_2 - p_2$, and (4) reduces to $\phi(a) + \phi(b) < \phi(a + u) + \phi(b - u)$. We need only show that $a - b + 2u > u > 0$ and this is identical to (2). This is easily verified and the theorem is proved.

Note that since $\delta_2^+ - \delta_1^+ = (\delta_1 - \delta_2)(1 - 2\alpha)$, we require that $\alpha \leqq \frac{1}{2}$ for $\delta_1 > \delta_2$, otherwise δ^+ is inadmissible by the theorem.

Theorem 1 states that for loss functions of the form (3), estimators not ordered in the same way as the parameters are inadmissible. The theorem is not necessarily true for other loss functions, as we now show. Let $n = 1$, then $\delta_i = \delta_i(X_1, Y_1)$ $i = 1, 2$. Consider the loss function W,

$$W(\delta, p) = [\delta_1 - p_1]^2 p_1^m + [\delta_2 - p_2]^2.$$

W is convex in $\delta = (\delta_1, \delta_2)$. We show that there is an admissible estimator δ with the property

$$\delta_1(1, 0) = (m + 2)/(m + 5), \qquad \delta_2(1, 0) = \tfrac{3}{5} \tag{5}$$

so that for $m > 2.5$, $\delta_1(1, 0) > \delta_2(1, 0)$.

Let λ be an a priori distribution uniform over $\Omega = \{p \mid 0 \leqq p_1 \leqq p_2 \leqq 1\}$. If δ is Bayes relative to λ, then we show that δ satisfies (5). For each estimator the risk is continuous on Ω, and λ puts positive probability on each open subset of Ω. Hence the Bayes estimator relative to λ is admissible. Let

$$\rho(r, s) = \int_0^1 \int_0^{p_2} [(r - p_1)^2 p_1^m + (s - p_2)^2] p_1(1 - p_2)\, dp_1\, dp_2.$$

We wish to find the pair (r, s) minimizing ρ. This minimizing pair will be the Bayes estimate at (1, 0). We have $\rho(r, s) = r^2/(m + 2)(m + 3)(m + 4) - 2r/(m + 3)(m + 4)(m + 5) + s^2/24 - s/20 +$ constant. This expression is minimized when $r = (m + 2)/(m + 5)$, $s = \frac{3}{5}$. Thus it is sometimes possible to get admissible estimators such that $P_p\{\delta_1 > \delta_2\} > 0$ if the loss is not of form (3).

3. Mixed estimators. We say that $\delta = (\delta_1, \delta_2)$ is a *mixed estimator* of $p = (p_1, p_2)$ where $\Omega = \{p \mid 0 \leqq p_1 \leqq p_2 \leqq 1\}$ if

$$\delta = (\alpha_n \bar{X}_n + (1 - \alpha_n)\bar{Y}_n, (1 - \alpha_n)\bar{X}_n + \alpha_n \bar{Y}_n)$$

where α_n is a random variable, $\alpha_n = 1$ if $\bar{X}_n - \bar{Y}_n \leqq 0$, $\alpha_n = \alpha_n^+$ otherwise. Following the remark after Theorem 1, we take $0 \leqq \alpha_n^+ \leqq \frac{1}{2}$, so that $P_p\{\delta_1 > \delta_2\} = 0$.

THEOREM 2. *Mixed estimators are invariant under a transformation that interchanges the outcomes,* 0 *and* 1, *of every binomial trial.*

PROOF. Let g be a transformation on the sample space $\mathfrak{X}$, such that $g(\bar{X}_n, \bar{Y}_n) = (1 - \bar{Y}_n, 1 - \bar{X}_n)$. The transformation g' induced on the parameter space is given by $g'(p_1, p_2) = (1 - p_2, 1 - p_1)$, and g'' transforms the estimators $g''(\delta_1, \delta_2) = (1 - \delta_2, 1 - \delta_1)$. Then for $\delta = (\delta_1, \delta_2)$ to be invariant we require $g''(\delta(\bar{X}_n, \bar{Y}_n)) = \delta[g(\bar{X}_n, \bar{Y}_n)]$ or $\delta_1(\bar{X}_n, \bar{Y}_n) + \delta_2(1 - \bar{Y}_n, 1 - \bar{X}_n) = 1$. This is obviously so for mixed estimators and the theorem is proved.

From now on we take as loss function, the square error loss,

$$W(\delta, p) = (\delta_1 - p_1)^2 + (\delta_2 - p_2)^2.$$

We wish to find the minimax mixed estimator, i.e., the minimax value of α_n^+. For simplicity we write α, α^+, $\bar{X}$, $\bar{Y}$ for α_n, α_n^+, $\bar{X}_n$, $\bar{Y}_n$ respectively. Denote the risk by $\rho_\delta(p_1, p_2; \alpha^+)$ where δ is a mixed estimator. With a little computation we have $\rho_\delta(p_1, p_2; \alpha^+) = [p_1(1 - p_1) + p_2(1 - p_2)]/n - 2E\{\alpha(1 - \alpha)(\bar{Y} - \bar{X})^2\} + 2(p_2 - p_1)E\{(1 - \alpha)(\bar{Y} - \bar{X})\}$. Now it is easy to verify that $\rho_\delta(p_1, p_2; \alpha^+) = \rho_\delta(1 - p_2, 1 - p_1; \alpha^+)$, i.e., the risk is symmetric about the line $p_1 + p_2 = 1$. Consider any line $p_2 - p_1 = \pi$ for $0 \leqq \pi \leqq 1$. Let $p_1 = p$ and $p_2 = p + \pi$. Then the risk can be written $\rho_\delta(p, p + \pi; \alpha^+) = [p(1 - p) + (p + \pi)(1 - p - \pi)]/n - [2\alpha^+(1 - \alpha^+)/n^2] \sum_{j=1}^n j^2 A_j - [2\pi(1 - \alpha^+)/n] \sum_{j=1}^n jA_j$ where

$$A_j = \sum_{k=0}^{n-j} \binom{n}{k} \binom{n}{k+j} (p + \pi)^k (1 - p - \pi)^{n-k} p^{k+j} (1 - p)^{n-k-j}.$$

Fix π, and write $\rho_\delta(p, p + \pi; \alpha^+) = \rho_\delta(p; \alpha^+)$. By the Mean Value Theorem, $\rho_\delta(p; \alpha^+) = \rho_\delta((1 - \pi)/2; \alpha^+) + \rho_\delta'(q; \alpha^+)(p - (1 - \pi)/2)$ where q is between p and $(1 - \pi)/2$. It follows by an induction on n that $\rho_\delta'(q; \alpha^+)(p - (1 - \pi)/2) \leqq 0$ for all values of π. Hence $\rho_\delta(p; \alpha^+) \leqq \rho_\delta((1 - \pi)/2; \alpha^+) = \rho_\delta((1 - \pi)/2, (1 + \pi)/2; \alpha^+)$ that is, the point (p_1, p_2) maximizing $\rho_\delta(p_1, p_2; \alpha^+)$ lies on the line $p_1 + p_2 = 1$. Now let $R = (1 - \pi)/2$ and $S = 1 - R$. Then

$$\begin{aligned}
\rho_\delta(R, S; \alpha^+) &= \frac{1 - \pi^2}{2n} - \frac{2\alpha^+(1 - \alpha^+)}{n^2} \sum_{j=1}^n j^2 R^{n+j} S^{n-j} \sum_{k=0}^{n-j} \binom{n}{k} \binom{n}{k+j} \\
&\quad - \frac{2\pi(1 - \alpha^+)}{n} \sum_{j=1}^n j R^{n+j} S^{n-j} \sum_{k=0}^{n-j} \binom{n}{k} \binom{n}{k+j} \\
&= \frac{1 - \pi^2}{2n} - \frac{2\alpha^+(1 - \alpha^+)}{n^2} \sum_{j=n+1}^{2n} (j - n)^2 \binom{2n}{j} R^j S^{2n-j} \\
&\quad - \frac{2\pi(1 - \alpha^+)}{n} \sum_{j=n+1}^{2n} (j - n) \binom{2n}{j} R^j S^{2n-j}.
\end{aligned} \tag{6}$$

The last step follows from a change in summation and the fact that

$$\sum_{k=0}^{n-j} \binom{n}{k} \binom{n}{k+j} = \binom{2n}{n+j}.$$

The minimax value of α^+ is that value of α^+ minimizing $\sup_{0 \leqq \pi \leqq 1} \rho_\delta(R, S; \alpha^+)$.

For $n = 1$ it is simple to show that $\pi = 0$ maximizes the risk and the minimax value of α^+ is $\frac{1}{2}$. For $n > 1$ the computations become formidable. We investigate the behavior of α^+ when n is large.

The incomplete moments, from $j = n + 1$ to $j = 2n$, of the binomial distribution, μ_r, are given by

$$\mu_r = \sum_{j=n+1}^{2n} (j - 2nR)^r \binom{2n}{j} R^j S^{2n-j} \qquad r = 0, 1, 2, \cdots.$$

Expressing (6) in terms of μ_0, μ_1 and μ_2, we have $\rho_\delta(R, S; \alpha^+) = (1 - \pi^2)/2n - [2\alpha^+(1 - \alpha^+)/n^2][\mu_2 + n^2\pi^2\mu_0 - 2n\pi\mu_1] - [2\pi(1 - \alpha^+)/n][\mu_1 - n\pi\mu_0]$. Then, using the recursion formulae cited in ([5], p. 135) we have $\rho_\delta(R, S; \alpha^+) = (1 - \pi^2)/2n + (1 - \alpha^+)RT_n[-2\alpha^+S/n + 2\pi(\alpha^+ - 1)] + (1 - \alpha^+)\mu_0 \cdot [-4\alpha^+RS/n + 2\pi^2(1 - \alpha^+)]$ where

$$T_j = \binom{2n}{j} R^j S^{2n-j} \quad \text{and} \quad \mu_0 = \sum_{j=n+1}^{2n} T_j.$$

For n sufficiently large, we ignore the terms $-4\alpha^+RS/n$ and $-2\alpha^+S/n$, and $\rho_\delta(R, S; \alpha^+)$ is given approximately by $(1 - \pi^2)/2n + (1 - \alpha^+)^2 2\pi[\pi\mu_0 - RT_n]$. The term $[\pi\mu_0 - RT_n]$ is non-positive for all n and for all $0 \leqq \pi \leqq 1$. Thus for large n, the minimax value of α^+ must be close to zero.

The author is indebted to Mr. Richard Black for computing Table 1 of minimax values. The time-consuming procedure allowed computations to two decimals only.

Ayer et al. [1], have found the maximum likelihood value of α_n^+ to be $\frac{1}{2}$, for all n.

4. Asymptotic estimator. Whether the minimax mixed estimator is admissible is an open question. In this section we present for a normal distributional problem asymptotically equivalent to the binomial one, an estimator that is both minimax and admissible. The estimator is readily computed.

Let X_i $i = 1, 2, \cdots n$ be mutually independent normally distributed random variables each with mean θ and variance V^2. In [4] it was shown that if the parameter space is $\Omega = \{\theta \mid \theta \geqq 0\}$, then an admissible and minimax estimator of θ is $\bar{X} + Vn^{-\frac{1}{2}}\nu(\bar{X}n^{\frac{1}{2}}/V)$ where $\nu(t) = \exp[-t^2/2]/\int_{-\infty}^{t} \exp[-s^2/2]ds$.

Consider the case of two ordered means. X_i, Y_i $i = 1 \cdots n$ are mutually independent normal random variables, with $E\{X_i\} = \theta_1$, $E\{Y_i\} = \theta_2$ and

TABLE 1

n	α_n^+	n	α_n^+	n	α_n^+	n	α_n^+	n	α_n^+	n	α_n^+
1	.50	6	.31	11	.22	16	.17	21	.14	26	.12
2	.49	7	.29	12	.20	17	.16	22	.14	27	.12
3	.44	8	.26	13	.19	18	.16	23	.13	28	.12
4	.39	9	.25	14	.19	19	.15	24	.13	29	.11
5	.35	10	.23	15	.18	20	.15	25	.12	30	.11

var. $\{X_i\}$ = var. $\{Y_i\} = 1$, $i = 1, 2, \cdots n$. The parameter space is

$$\Omega = \{(\theta_1, \theta_2) \mid \theta_2 \geqq \theta_1\}.$$

We require an estimator $\delta = (\delta_1, \delta_2)$ of $\theta = (\theta_1, \theta_2)$. The square error loss function is

$$(\delta_1 - \theta_1)^2 + (\delta_2 - \theta_2)^2 = \tfrac{1}{2}[\delta_1 + \delta_2 - (\theta_1 + \theta_2)]^2 + \tfrac{1}{2}[\delta_2 - \delta_1 - (\theta_2 - \theta_1)]^2. \tag{7}$$

On the right-hand side of (7) we are estimating, in the first term, a parameter $\theta_1 + \theta_2$ where $-\infty < \theta_1 + \theta_2 < \infty$, and in the second term a parameter $\theta_2 - \theta_1$ where $\theta_2 - \theta_1 \geqq 0$. This suggests estimators satisfying $\delta_2 + \delta_1 = \bar{Y} + \bar{X}$, $\delta_2 - \delta_1 = \bar{Y} - \bar{X} + (2/n)^{\frac{1}{2}}\nu((n/2)^{\frac{1}{2}}(\bar{Y} - \bar{X}))$, or

$$\begin{aligned} \delta = (\delta_1, \delta_2) = (\bar{X} - (2n)^{-\frac{1}{2}}\nu((n/2)^{\frac{1}{2}}(\bar{Y} - \bar{X})), \\ \bar{Y} + (2n)^{-\frac{1}{2}}\nu((n/2)^{\frac{1}{2}}(\bar{Y} - \bar{X}))). \end{aligned} \tag{8}$$

We have in fact,

THEOREM 3. *The estimator* (8) *is admissible and minimax for*

$$\Omega = \{(\theta_1, \theta_2) \mid \theta_2 \geqq \theta_1\}.$$

PROOF. The proof parallels that given in [4], and is only sketched here. With no loss of generality we take $n = 1$.

We take as a priori distribution

$$\begin{aligned} \lambda_\sigma(\theta) &= (1/\pi\sigma^2) \exp(-\tfrac{1}{2})(\theta_1^2 + \theta_2^2)/\sigma^2 \qquad && \theta \varepsilon \Omega \\ &= 0 && \theta \not\varepsilon \Omega. \end{aligned}$$

Then the Bayes estimator $\delta_\sigma = (\delta_1(\sigma), \delta_2(\sigma))$ of θ with respect to $\lambda_\sigma(\theta)$ is computed to be

$$(k^2(X_1 - 2^{-\frac{1}{2}}\nu(k(Y_1 - X_1)/2^{\frac{1}{2}})), \qquad k^2(Y_1 + 2^{-\frac{1}{2}}\nu(k(Y_1 - X_1)/2^{\frac{1}{2}})))$$

where $k^2 = \sigma^2/(1 + \sigma^2)$. As

$$\sigma \to \infty, \qquad \delta_\sigma \to \delta = (X_1 - 2^{-\frac{1}{2}}\nu((Y_1 - X_1)/2^{\frac{1}{2}}), Y_1 + 2^{-\frac{1}{2}}\nu((Y_1 - X_1)/2^{\frac{1}{2}})).$$

Now suppose δ is not admissible. Then there exists an estimator δ^+ such that $\rho_{\delta+}(\theta) \leqq \rho_\delta(\theta)$ for all $\theta \varepsilon \Omega$, and strict inequality for some θ. Consider the quantity

$$[r(\delta) - r(\delta^+)]/[r(\delta) - r(\delta_\sigma)] \tag{9}$$

where $r(\delta)$, $r(\delta^+)$ and $r(\delta_\sigma)$ are the average risks with respect to $\lambda_\sigma(\theta)$ of the estimators δ, δ^+, and δ_σ respectively. The numerator of (9) is non-negative. Using the method of Blyth [2], it can be shown that the ratio (9) > 1, for σ sufficiently large. This implies $r(\delta_\sigma) > r(\delta^+)$ which is false. This contradiction proves admissibility.

Further the differential inequality procedure of Hodges and Lehmann [3] shows that a minimax estimator of θ has risk $\leqq 2$. The risk of δ is

$$2 - (\theta_2 - \theta_1)/2^{\frac{1}{2}} \cdot E\{\nu((Y_1 - X_1)/2^{\frac{1}{2}})\} \leqq 2.$$

Following [4] we have

$$2 - (\theta_2 - \theta_1)/2^{\frac{1}{2}} \cdot E\{\nu((Y_1 - X_1)/2^{\frac{1}{2}})\} \to 2$$

as $\theta_2 - \theta_1 \to \infty$. Hence δ is minimax.

The estimator (8) could be used in the ordered binomial estimation problem, if n is large enough to permit a normal approximation. For properties of the function ν, the reader is referred to [4].

REFERENCES

[1] AYER, MIRIAM, BRUNK, H. D., EWING G. M., REID, W. T., and SILVERMAN, EDWARD (1955). An empirical distribution function for sampling with incomplete information. *Ann. Math. Statist.* **26** 641–647.

[2] BLYTH, COLIN R. (1951). On minimax statistical decision procedures and their admissibility. *Ann. Math. Statist.* **22** 22–42.

[3] HODGES, J. L., JR. and LEHMANN, E. L. (1951). Some applications of the Cramér-Rao inequality. *Proc. Second Berkeley Symp. Math. Statist. Prob.* 13–22. Univ. of California Press.

[4] KATZ, MORRIS W. (1961). Admissible and minimax estimates of parameters in truncated spaces. *Ann. Math. Statist.* **32** 136–142.

[5] KENDALL, MAURICE G. (1943). *The Advanced Theory of Statistics*, **1** Griffin, London.