Minimax Estimation of a Variance

Thomas S. Ferguson
Mathematics Department
U. C. L. A.
Los Angeles, CA 90024

Lynn Kuo
Statistics Department
University of Connecticut
Storrs, CT 06269

Abstract: The nonparametric problem of estimating a variance based on a sample of size n from a univariate distribution which has a known bounded range but is otherwise arbitrary is treated. For squared error loss, a certain linear function of the sample variance is seen to be minimax for each n from 2 through 13, except $n = 4$. For squared error loss weighted by the reciprocal of the variance, a constant multiple of the sample variance is minimax for each n from 2 through 11. The least favorable distribution for these cases gives probability one to the Bernoulli distributions.

AMS 1980 subject classifications. Primary 62C20; Secondary 62G05.

Key words and phrases. Admissible, minimax, nonparametric, linear estimator, moment conditions.

1. Introduction and Summary.

We study the problem of finding nonparametric minimax estimates of the variance of an unknown distribution F on the real line, based on a sample from F, similar to the treatment of Hodges and Lehmann (1950) for the problem of estimating the mean of F.

We first review the problem of estimating the mean nonparametrically. Let $X_1, \ldots, X_n$ be a sample from a distribution F with finite mean, μ, and consider the problem of estimating μ with squared error loss, $L(F, \hat{\mu}) = (\mu - \hat{\mu})^2$. To rule out the possibility that every estimator of μ have infinite maximum risk, Hodges and Lehmann consider two possible restrictions on F: (i) bounded variance, say $\mathrm{Var}(X_i) \leq 1$; (ii) bounded range, say $0 \leq X_i \leq 1$. Under (i) the sample mean $\overline{X}_n$ is minimax, and the normal distributions with variance 1 form a least favorable class. Under (ii), the Bernoulli distributions on $\{0,1\}$ are least favorable, and the estimate

$$d(\overline{X}_n) = \frac{\sqrt{n}}{1+\sqrt{n}}\,\overline{X}_n + \frac{1}{1+\sqrt{n}}\,\frac{1}{2}$$

is minimax. Meeden, Ghosh, and Vardeman (1985) show that $\overline{X}_n$ is admissible in case (i). Applying their examples 1 and 2 (with $M = \sqrt{n}$ and $\mu^* = 1/2$), one can show that $\overline{X}_n$ and $d(\overline{X}_n)$ are admissible in case (ii).

A third case arises when the loss is taken to be the scale invariant loss function, $L_s(F, \hat{\mu}) = (\mu - \hat{\mu})^2/\sigma^2$. In this case, we need only restrict the parameter space to those F with finite positive variance. Then $\overline{X}_n$ is again minimax, since it is an equalizer rule and an extended Bayes rule with respect to the priors concentrated on normal distributions, $F = N(\mu, 1)$, where μ is $N(0, \tau^2)$ with τ^2 known and large. $\overline{X}_n$ is also admissible for $L_s(F, \hat{\mu})$ since it is admissible for $L(F, \hat{\mu})$.

Consider now the corresponding problems of estimating a variance. Throughout this paper we use θ instead of σ^2 to represent the variance of the unknown distribution, F. We consider three loss functions: squared error loss, $L_1(F, \hat{\theta}) = (\theta - \hat{\theta})^2$, a weighted squared error loss, $L_2(F, \hat{\theta}) = (\theta - \hat{\theta})^2/\theta$, and the scale invariant loss, $L_3(F, \hat{\theta}) = (\theta - \hat{\theta})^2/\theta^2$. We denote the risk function for the loss L_i by R_i, that is, $R_i(F, \delta) = E_F L_i(F, \delta(\mathbf{X}))$, where $\mathbf{X} = (X_1, \ldots, X_n)$.

In some cases, the minimax estimate turns out to be degenerate. For the scale invariant loss, L_3, where we restrict the parameter space to be distributions with finite positive fourth moment, the degenerate estimate $\hat{\theta}_0 \equiv 0$ is minimax for any sample size. This may be seen as follows. Clearly, $R_3(F, \hat{\theta}_0) \equiv 1$, so it suffices to show that $\sup_F R_3(F, d) \geq 1$ for any decision rule (estimate), d. In fact, $\sup_{F \in \mathcal{G}} R_3(F, d) \geq 1$ for all d, where $\mathcal{G}$ is the set of all distributions G_p for $0 < p \leq 1/2$, where G_p gives mass p to $+1$ and -1, and mass $1-2p$ to 0. This holds since the probability that all observations are zero is $P(\mathbf{X}=\mathbf{0})=(1-2p)^n$, so that

$$R_3(G_p,d)\geq(1-2p)^n\left(1-\frac{d(\mathbf{0})}{2p}\right)^2,$$

and as $p\to 0$, this quantity tends to ∞ if $d(\mathbf{0})\neq 0$, and to 1 if $d(\mathbf{0})=0$.

A similar analysis shows that in case (i) above with squared error loss and with variance at most 1, the degenerate rule, $\hat{\theta}_1\equiv 1/2$ is minimax for any sample size. Here, we have $R_1(F,\hat{\theta}_1)=(\theta-1/2)^2\leq 1/4$ for all distributions with $0\leq\theta\leq 1$. Yet for the class, $\mathcal{G}_\infty$, consisting of the distribution G_∞, degenerate at zero, and of the distributions, G_a for $a\geq 1$, where G_a gives mass $1/(2a^2)$ to both $+a$ and $-a$, and mass $1-(1/a^2)$ to 0, we have for any decision rule, d,

$$R_1(G_a,d)\geq(1-\frac{1}{a^2})^n(1-d(\mathbf{0}))^2$$

so that

$$\sup_{G\in\mathcal{G}_\infty}R_1(G,d)\geq\max(d(\mathbf{0})^2,\sup_a(1-\frac{1}{a^2})^n(1-d(\mathbf{0}))^2)\geq\frac{1}{4}.$$

For the weighted squared error loss, $L_2(F,\theta)$, and with variance at most one, a similar argument gives 1 as the minimax risk achieved at the degenerate rule, $\hat{\theta}_0\equiv 0$.

A perhaps more direct analogy with the case (i) problem of estimating a mean would be to restrict the distributions to have bounded fourth central moment, say $\mu_4\leq 1$. The analysis of the above paragraph does not work because the distribution G_a has fourth central moment tending to infinity as $a\to\infty$. We do not know the minimax estimate of the variance for this problem.

For case (ii) above, the minimax estimate turns out to be nontrivial and we are successful in finding it only for certain values of n for squared error loss, L_1, and weighted squared error loss, L_2. We restrict F to be in the class, $\mathcal{F}_{[0,1]}$, of distributions with support in $[0,1]$, and for the L_2 loss function we assume that the variance of F is positive. Let $\tilde{\theta}_n$ denote the unbiased estimate of θ,

$$\tilde{\theta}_n=\sum_{i=1}^n(X_i-\overline{X}_n)^2/(n-1). \tag{1.1}$$

The method of attack will proceed along these lines. First, we make a conjecture, eventually shown to be correct for some values of n, that the least favorable distribution gives weight only to the class, $\mathcal{F}_{\{0,1\}}$, of Bernoulli distributions B_p for $p\in[0,1]$, where B_p gives mass p to 1 and mass $1-p$ to 0. Under this conjecture, the problem reduces to finding minimax rules for estimating the variance, $\theta=p(1-p)$, of the Bernoulli distribution B_p. We therefore search for equalizer rules for both loss functions, L_1 and L_2. In Section

2 we find linear functions of $\tilde{\theta}_n$ that are equalizer rules for estimating $\theta = p(1-p)$. Therefore in the following, it suffices to restrict attention to linear estimators.

Second, we show that the supremum of the risk of linear estimators over $\mathcal{F}_{[0,1]}$ is attained at the Bernoulli distributions. This is done in Section 3.

Thus it is sufficient to show these equalizer rules are minimax for the estimation of the variance of the Bernoulli distribution. This we attempt in Section 4. In Section 4.1 we show that the equalizer rules are minimax within the class of linear functions of $\tilde{\theta}_n$. In Section 4.2, we show that the equalizer estimators are admissible and minimax among all estimators under L_1 for values of $n = 3, 5, 6, 7, \ldots, 13$, and under L_2 for $n = 2, 3, 4, \ldots, 11$. For the loss L_1 and $n = 4$, we find the minimax estimator by numerical methods; whether this estimator is also minimax for the nonparametric problem is still unknown.

We are led to believe that the minimax property is a very delicate one. The equalizer rules seem to be very good in any case (for $n = 4$ and L_1 loss, the minimax rule improves on the equalizer rule by only .00000047), so whether or not the equalizer rule is minimax is much a matter of chance. For large n, there is a much greater possibility of having a complex estimator uniformly improve on the equalizer rule. What is perhaps surprising is that, except for $n = 4$ and loss L_1, there seems to be a sharp cutoff for n at which the equalizer rule is minimax: 13 for L_1 and 11 for L_2.

Brown, Chow and Fong (1992) have shown that the maximum likelihood estimator of the variance of a binomial distribution under squared error loss is admissible for $n \le 5$ and inadmissible for $n \ge 6$. The admissibility of $\tilde{S}_n^2 = (n+1)^{-1}(n-1)\tilde{\theta}_n$ for the L_1, L_2 and L_3 loss functions for all F is established by Meeden, Ghosh and Vardeman (1985). Other papers such as Aggarwal (1955), Phadia (1973), Cohen and Kuo (1985), Brown (1988), and Yu (1989), study the nonparametric estimation of a distribution function from a decision theoretic point of view.

2. Equalizer Rules for the Bernoulli Distributions, $n \ge 2$.

In this section, we restrict attention to the Bernoulli distributions and find constant risk decision rules for both loss functions, L_1 and L_2 for sample of size $n \ge 2$. For later use, we first give a formula for the risk function under squared error loss, L_1, of an arbitrary linear function of $\tilde{\theta}_n$, for arbitrary distributions F having finite fourth moment.

Lemma 2.1. *Let μ_4 represent the fourth moment of F about the mean. Then,*

$$R_1(F, a\tilde{\theta}_n + b) = \frac{a^2}{n}\mu_4 + \left((1-a)^2 - \frac{(n-3)a^2}{n(n-1)}\right)\theta^2 - 2b(1-a)\theta + b^2. \tag{2.1}$$

Proof.

$$\begin{aligned} R_1(F, a\tilde{\theta}_n + b) &= EL_1(F, a\tilde{\theta}_n + b) \\ &= E(a\tilde{\theta}_n + b - \theta)^2 \\ &= a^2 \mathrm{Var}(\tilde{\theta}_n) + (b - \theta(1-a))^2. \end{aligned}$$

The formula follows using the expression,

$$\mathrm{Var}(\tilde{\theta}_n) = \frac{\mu_4}{n} - \frac{(n-3)\theta^2}{n(n-1)},$$

(see, for example, S. S. Wilks (1962), p. 199) and collecting terms in μ_4, θ^2 and θ. ∎

For the Bernoulli distributions, B_p, the sample variance takes on the simple form, $\tilde{\theta}_n = W_n(n - W_n)/(n(n-1))$, where $W_n = \sum_{i=1}^n X_i$ is the number of ones in the sample. The variance of B_p is $\theta = p(1-p)$, and the fourth moment about the mean is $\mu_4 = p(1-p)^4 + (1-p)p^4 = p(1-p)(1-3p+3p^2) = \theta(1-3\theta)$. Substituting this into (2.1) and collecting terms gives the following corollary to Lemma 2.1.

Lemma 2.2.

$$R_1(B_p, a\tilde{\theta}_n + b) = \left((1-a)^2 - \frac{(4n-6)a^2}{n(n-1)}\right)\theta^2 + \left(\frac{a^2}{n} - 2b(1-a)\right)\theta + b^2. \tag{2.2}$$

We may use this formula to derive the equalizer rules, to be denoted by δ_n,

$$\delta_n = a_n\tilde{\theta}_n + b_n, \tag{2.3}$$

by equating the coefficients of θ and θ^2 to zero. This leads to the equations,

$$2b_n(1 - a_n) = a_n^2/n \tag{2.4}$$

and

$$1 - 2a_n + z_n a_n^2 = 0, \tag{2.5}$$

where

$$z_n = \frac{(n-2)(n-3)}{n(n-1)}. \tag{2.6}$$

Also note that the constant risk of the rule, δ_n, is

$$R_1(B_p, \delta_n) = b_n^2.$$

For $n = 2$ and $n = 3$, (2.5) is linear in a_n, and the equations (2.4) and (2.5) have a unique solution,

$$\begin{aligned} a_2 &= \frac{1}{2} \qquad b_2 = \frac{1}{8} \\ a_3 &= \frac{1}{2} \qquad b_3 = \frac{1}{12}. \end{aligned} \tag{2.7}$$

For $n \geq 4$, (2.5) has two roots and the equations have two solutions. We choose as δ_n the solution with the smaller risk, b_n^2, namely, $\delta_n = a_n\tilde{\theta}_n + b_n$, where,

$$a_n = \frac{1-\sqrt{1-z_n}}{z_n} \quad \text{and} \quad b_n = \frac{a_n^2}{2n(1-a_n)}. \tag{2.8}$$

Under the weighted squared error loss function, L_2, the risk function of the rule $a\tilde{\theta}_n + b$ is found by dividing (2.2) by θ,

$$R_2(B_p, a\tilde{\theta}_n + b) = \left((1-a)^2 - \frac{(4n-6)a^2}{n(n-1)}\right)\theta + \left(\frac{a^2}{n} - 2b(1-a)\right) + b^2/\theta. \tag{2.9}$$

For this to be constant, the first and the last coefficients must vanish. This leads to equalizer rules, denoted by d_n, which differ from δ_n by the removal of the term b_n,

$$d_n = a_n\tilde{\theta}_n, \tag{2.10}$$

where the a_n are as given in (2.5) and (2.6). The constant risk of these decision rules is

$$R_2(B_p, d_n) = a_n^2/n.$$

3. Reduction to the Bernoulli Case for Linear Estimates.

In this section, we show that in the nonparametric problem of estimating a variance of a distribution on $[0,1]$ by a linear function of $\tilde{\theta}_n$, the worst case distribution is Bernoulli. The proof is based on the following lemma of independent interest. For the remarkably simple proof of this lemma, we are indebted to Thomas Liggett.

Lemma 3.1. *If $X \in [0,1]$, then,*

$$\mu_4 + 3\sigma^4 \leq \sigma^2,$$

with equality if and only if X is Bernoulli or degenerate.

Proof. Let X and Y be i.i.d. on $[0,1]$. Then $(X-Y)^2 \in [0,1]$, so that $\mathrm{E}(X-Y)^4 \leq \mathrm{E}(X-Y)^2$ with equality if and only if X is Bernoulli or degenerate. This inequality is equivalent to

$$\mathrm{E}((X-\mu)-(Y-\mu))^4 \leq \mathrm{E}((X-\mu)-(Y-\mu))^2$$

which reduces to

$$2\mu_4 + 6\sigma^4 \leq 2\sigma^2. \quad \blacksquare$$

From this, the main result follows easily.

Theorem 3.1. *For every a and b, and for every F in $\mathcal{F}_{[0,1]}$, there exists a Bernoulli distribution, B_p in $\mathcal{F}_{\{0,1\}}$ such that*

$$R_1(F, a\tilde{\theta}_n + b) \leq R_1(B_p, a\tilde{\theta}_n + b).$$

Similarly for the risk function, R_2.

Proof. The variance of any $F \in \mathcal{F}_{[0,1]}$ satisfies $0 \leq \theta \leq 1/4$ so we can find $B_p \in \mathcal{F}_{\{0,1\}}$ with the same variance. The substitution of μ_4 with $\theta - 3\theta^2$ in the $R_1(F, a\tilde{\theta}_n + b)$ of Lemma 2.1 results in the $R_1(B_p, a\tilde{\theta}_n + b)$ of Lemma 2.2 with this B_p. Lemma 3.1 shows that the substitution results in an increase. Since R_2 is equal to R_1 divided by θ, the same result holds for R_2 as well. ■

4. Minimax Estimator of the Variance of the Binomial Distribution.

Minimax estimation of the variance of the restricted family $\mathcal{F}_{\{0,1\}}$ is studied in this section. Thus, we deal with the sufficient statistic, W_n, the number of 1's in the sample, which has a binomial distribution, $\mathcal{B}(n,p)$. The minimum variance unbiased estimate (1.1) of the variance, $\theta = p(1-p)$, reduces to $\tilde{\theta}_n = W_n(n - W_n)/(n(n-1))$. In Section 4.1, we find the minimax estimator of θ within the class of linear functions of $\tilde{\theta}_n$, and in Section 4.2, we show this estimator is minimax and admissible within the class of all estimators for certain values of n.

4.1 Minimax Linear Estimators.

First we show that for squared error loss the equalizer rule, $\delta_n = a_n\tilde{\theta}_n + b_n$ with a_n and b_n given by (2.7) and (2.8), is minimax within the class of all estimators that are linear in $\tilde{\theta}_n$, for all $n \geq 3$. In the proof, we use the principle that if an equalizer rule d is a Bayes rule within a class $\mathcal{C}$ of decision rules, then d is minimax within $\mathcal{C}$. For if $d \in \mathcal{C}$ is not minimax, then there is an $\epsilon > 0$ and a rule $d^* \in \mathcal{C}$ such that $\max_\theta R(\theta, d^*) < \max_\theta R(\theta, d) - \epsilon$, which implies, if d is an equalizer rule, that $R(\theta, d^*) < R(\theta, d) - \epsilon$ for all θ, which in turn implies for any prior that the Bayes risk of d^* is at least ϵ smaller that the Bayes risk of d, so that d cannot be Bayes within $\mathcal{C}$ for any prior.

Theorem 4.1. *For all $n \geq 3$, the equalizer rule, δ_n, is minimax with respect to squared error loss, $L_1(p,a) = (\theta - a)^2$ where $\theta = p(1-p)$, within the class of estimators that are linear functions of $\tilde{\theta}_n$.*

Proof. We take $\mathcal{C}$ to be the class of linear rules, $a\tilde{\theta}_n + b$, and from Lemma 2.2 note that the risk function of such rules may be written as

$$R_1(\theta, a\tilde{\theta}_n + b) = A\theta^2 + B\theta + C,$$

where

$$
\begin{aligned}
A &= a^2 z_n - 2a + 1 = (1-a)^2 - a^2(1-z_n) \\
B &= \frac{a^2}{n} - 2b(1-a) \\
C &= b^2 \\
z_n &= \frac{(n-2)(n-3)}{n(n-1)}.
\end{aligned}
$$

To show that for $n \geq 3$ the equalizer rule, $\delta_n = a_n \tilde{\theta}_n + b_n$ where a_n and b_n are given in (2.7) and (2.8), is minimax within $\mathcal{C}$, it is sufficient to show that there exists a prior distribution, π, for θ in the interval $[0, 1/4]$ such that δ_n is Bayes with respect to π within $\mathcal{C}$.

The Bayes risk of a linear rule with respect to a prior distribution, π, may be written as

$$
\begin{aligned}
r(\pi, a\tilde{\theta}_n + b) &= A\mu_2 + B\mu_1 + C \\
&= \mu_2((1-a)^2 - a^2(1-z_n)) + \mu_1(\frac{a^2}{n} - 2b(1-a)) + b^2
\end{aligned} \tag{4.1}
$$

where $\mu_i = \mathrm{E}_\pi \theta^i$. We are to show that there exists a prior π such that the minimum of (4.1) over all a and b occurs at that a and b that make the coefficients A and B zero. For fixed a, (4.1) is a quadratic function of b with a minimum at $b = \mu_1(1-a)$. With this value of b, the Bayes risk becomes

$$
r(\pi, a\tilde{\theta}_n + b) = \mu_2((1-a)^2 - a^2(1-z_n)) + \mu_1(\frac{a^2}{n} - 2\mu_1(1-a)^2) + \mu_1^2(1-a)^2. \tag{4.2}
$$

If both coefficients A and B are to be zero, then we must have $(1-a)^2 = a^2(1-z_n)$ and $a^2/n = 2\mu_1(1-a)^2$, which, eliminating a, gives

$$
\mu_1 = \frac{1}{2n(1-z_n)} = \frac{n-1}{4(2n-3)}
$$

as a necessary condition for the desired prior. With this value of μ_1, we may write (4.2) as

$$
r(\pi, a\tilde{\theta}_n + b) = (\mu_2 - 2\mu_1^2)((1-a)^2 - a^2(1-z_n)) + \mu_1^2(1-a)^2. \tag{4.3}
$$

We must show that there exists a choice of μ_2 as a second moment of a distribution on $[0, 1/4]$ whose first moment is $\mu_1 = 1/(2n(1-z_n))$, such that the minimum of (4.3) occurs at a point a that satisfies $(1-a)^2 = a^2(1-z_n)$. (4.3) is a quadratic function of a with a minimum at the point

$$
a = \frac{\mu_2 - \mu_1^2}{(\mu_2 - \mu_1^2)z_n + \mu_1^2(1-z_n)}.
$$

The equation, $(1-a)^2 = a^2(1-z_n)$, becomes equivalent to $(\mu_2-\mu_1^2)^2 = (1-z_n)(2\mu_1^2-\mu_2)^2$. We may solve this for μ_2 in the interval $(\mu_1^2, 2\mu_1^2)$ by taking square roots, $\mu_2 - \mu_1^2 = \sqrt{1-z_n}(2\mu_1^2 - \mu_2)$, or equivalently,

$$\mu_2 = \mu_1^2 \frac{1+2\sqrt{1-z_n}}{1+\sqrt{1-z_n}}.$$

It remains to be shown that there is a distribution on $[0, 1/4]$ with μ_1 and μ_2 as the first two moments. For this it is necessary and sufficient that $\mu_1^2 \le \mu_2 \le \mu_1/4 \le 1/16$. The first and third inequalities are clear. To show $\mu_2 \le \mu_1/4$, we replace μ_2 by its value in terms of μ_1 and cancel μ_1 and find that it is equivalent to show

$$\mu_1 \le \frac{1+\sqrt{1-z_n}}{4(1+2\sqrt{1-z_n})}.$$

We replace μ_1 by its value and find it is equivalent to show

$$\frac{n-1}{2n-3} \le \frac{1+\sqrt{1-z_n}}{1+2\sqrt{1-z_n}}.$$

This reduces to $\sqrt{1-z_n} \le n-2$, which is valid for all $n \ge 3$. ■

When $n = 2$, the equalizer rule, $a_2\tilde{\theta}_n + b_2$ is not minimax. In this case, the class of linear estimators coincides with the class of all estimators. In Section 4.2, we show that $\tilde{a}_2\tilde{\theta}_2 + \tilde{b}_2$ is minimax, where $\tilde{a}_2 = 1-(\sqrt{2})^{-1}$ and $\tilde{b}_2 = (\sqrt{2}-1)/4$.

The corresponding result for scaled squared error loss, $L_2(p,a) = (\theta-a)^2/\theta$, is much easier.

Theorem 4.2. *For all $n \ge 2$, the equalizer rule, $d_n = a_n\tilde{\theta}_n$, is minimax with respect to the loss, $L_2(p,a)$, within the class of estimators that are linear functions of $\tilde{\theta}_n$.*

Proof. From (2.9), we see that the risk of $a\tilde{\theta}_n + b$ is of the form $A\theta + B + b^2/\theta$. Therefore, we must have $b = 0$ to have a bounded risk. Then since $R_2(p, a\tilde{\theta}_n)$ is linear in θ, it achieves its maximum at the boundary points $\theta = 0$ and $\theta = 1/4$, corresponding to say $p = 0$ and $p = 1/2$, so that

$$\begin{aligned} \max_{0\le p\le 1} R_2(p, a\tilde{\theta}_n) &= \max\{R_2(0, a\tilde{\theta}_n), R_2(\frac{1}{2}, a\tilde{\theta}_n)\} \\ &= \max\{\frac{a^2}{n}, \frac{a^2 z_n - 2a + 1}{4} + \frac{a^2}{n}\}. \end{aligned} \tag{4.4}$$

The left and right terms of this maximum are equal if a is equal to the a_n given in (2.7) and (2.8), while for $a \le a_n$ the right side is decreasing in a and for $a \ge a_n$ the left side is increasing in a. This implies that the minimum of this maximum occurs at $a = a_n$ with a minimax value of a_n^2/n. ■

4.2 Minimaxity of δ_n and d_n within all Estimators.

In this section, we investigate whether or not the minimax linear estimator of Section 4.1 is minimax overall. We first treat the L_1 loss function. It suffices to show that the equalizer rule for $n \geq 3$ is Bayes with respect to some prior distribution on [0,1]. Unfortunately, the conjugate prior does not work. The Bayes rules with respect to the beta distributions, $\mathcal{B}e(\alpha, \alpha)$, with squared error loss are $d_\alpha(W_n) = (2\alpha+n)^{-1}(2\alpha+n+1)^{-1}\{W_n(n-W_n)+\alpha(\alpha+n)\}$. None of these rules come close to the equalizer rule. Instead, we find necessary and sufficient conditions on the first few moments of the prior distribution in order that the equalizer rule be Bayes with respect to this prior. Then we check whether or not there exists a distribution with these first few moments.

We show that $\delta_n = a_n\tilde{\theta}_n + b_n$ is minimax for $n = 3, 5, 6, \ldots, 13$. For $n = 4$, $a_4\tilde{\theta}_4 + b_4$ is not Bayes with respect to any prior distribution on [0,1], nor is it minimax. We obtain a minimax estimator by direct calculation in this case in Theorem 4.4.

Theorem 4.3. *In the problem of estimating $\theta = p(1-p)$ with loss $L_1(p,a) = (\theta - a)^2$ based on $W \in \mathcal{B}(n,p)$, the equalizer rule, δ_n, is admissible and minimax for $n = 3$ and $5 \leq n \leq 13$.*

Proof. The result for $n = 3$ follows immediately from Theorem 4.1 since $\tilde{\theta}_3$ assumes only two values and any function defined on only two values is linear. Assume then that $n \geq 4$ where (2.8) holds for a_n and b_n. Since the problem is invariant under the transformation $x \to 1-x$, we may restrict attention in our search for a minimax rule to rules depending only on $Y_n = \min\{W_n, n-W_n\}$. The density of Y_n given $\theta = p(1-p)$ is

$$\mathrm{P}(Y_n = y|\theta) = \begin{cases} \binom{n}{y}\theta^y[(1-p)^{n-2y} + p^{n-2y}], & \text{for } 0 \leq y < n/2 \\ \binom{n}{y}\theta^y, & \text{if } n \text{ is even and } y = n/2. \end{cases}$$

Since this depends on p only through θ, we may, with some algebraic manipulations, write this density as

$$\mathrm{P}(Y_n = y|\theta) = \binom{n}{y}\theta^y q_{n-2y}(\theta),$$

where $q_0(\theta) = 1$ and

$$q_k(\theta) = \sum_{j=0}^{\lfloor k/2 \rfloor} \frac{k}{k-j}\binom{k-j}{j}(-\theta)^j, \quad \text{for} \quad k \geq 1,$$

and $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x. The first few functions, q_k, are: $q_1(\theta) = 1$; $q_2(\theta) = 1-2\theta$; $q_3(\theta) = 1-3\theta$; $q_4(\theta) = 1-4\theta+2\theta^2$; $q_5(\theta) = 1-5\theta+5\theta^2$; $q_6(\theta) = 1-6\theta+9\theta^2-2\theta^3$; etc.

Thus, $q_k(\theta)$ is a polynomial of degree $\lfloor k/2 \rfloor$ in θ, and $\mathrm{P}(Y_n = y|\theta)$ is a polynomial of degree $\lfloor n/2 \rfloor$ in θ. The risk function depends only on θ so we write it as $R(\theta, d)$ and evaluate it as

$$R(\theta, d) = \sum_{y=0}^{\lfloor n/2 \rfloor} \mathrm{P}(Y_n = y|\theta)(\theta - d(y))^2.$$

This is a polynomial of degree $\lfloor n/2 \rfloor + 1$. (The coefficient of the highest term gets cancelled.)

To find a Bayes rule with respect to a given prior, π, we minimize $\mathrm{E}_\pi R(\theta, d)$ separately for each $d(y)$. Thus, the Bayes rule is

$$d(y) = \frac{\mathrm{E}_\pi \theta^{y+1} q_{n-2y}(\theta)}{\mathrm{E}_\pi \theta^y q_{n-2y}(\theta)}, \qquad \text{for} \quad y = 0, 1, \ldots, \lfloor n/2 \rfloor.$$

The question is: Does there exist a prior distribution, $\pi(\theta)$ for θ in $[0, 1/4]$, such that

$$\text{(4.5)} \qquad \frac{\mathrm{E}_\pi \theta^{y+1} q_{n-2y}(\theta)}{\mathrm{E}_\pi \theta^y q_{n-2y}(\theta)} = \delta_n(y) = a_n \frac{y(n-y)}{n(n-1)} + b_n \quad \text{for all} \quad y = 0, 1, \ldots, \lfloor n/2 \rfloor?$$

If so, the equalizer rule, being Bayes, will be minimax. Moreover, being unique Bayes, it is admissible among invariant rules and hence admissible (Theorem 4.3.2 of Ferguson (1967)).

Equation (4.5) is essentially a linear system of $\lfloor n/2 \rfloor + 1$ equations in the first $\lfloor n/2 \rfloor + 1$ moments of the prior distribution. Let $\mu_i = \mathrm{E}_\pi \theta^i$. For fixed n, we can solve these equations for $\mu_1, \ldots, \mu_{\lfloor n/2 \rfloor + 1}$, and then check whether or not there exists a distribution on $[0, 1/4]$ having these moments.

These equations may be written $\mathbf{A}\mu = \mathbf{b}$, where $\mu = (\mu_1, \ldots, \mu_{\lfloor n/2 \rfloor + 1})^T$, where $\mathbf{b} = (-b_n, 0, \ldots, 0)^T$, and where $\mathbf{A}$ is the $(\lfloor n/2 \rfloor + 1) \times (\lfloor n/2 \rfloor + 1)$ matrix whose i, j^{th} entry is given by

$$A_{ij} = (-1)^{j-i+1}(\delta_n(i-1) b_{ij} + c_{ij}),$$

where

$$b_{ij} = \begin{cases} \frac{(n-2i+2)(n-i-j)!}{(j-i+1)!(n-2j)!}, & \text{for } i = 1, \ldots, \lfloor n/2 \rfloor,\ j = i-1, \ldots, \lfloor n/2 \rfloor \\ 1 & \text{for } i = \lfloor n/2 \rfloor + 1,\ j = \lfloor n/2 \rfloor \\ 0 & \text{otherwise,} \end{cases}$$

$$c_{ij} = \begin{cases} \frac{(n-2i+2)(n-i-j+1)!}{(j-i)!(n-2j+2)!}, & \text{for } i = 1, \ldots, \lfloor n/2 \rfloor,\ j = i, \ldots, \lfloor n/2 \rfloor + 1 \\ 1 & \text{for } i = j = \lfloor n/2 \rfloor + 1 \\ 0 & \text{otherwise.} \end{cases}$$

The solution to the problem of determining whether a given sequence of numbers, $m_1, \ldots, m_k$, can be the first k moments of a distribution on $[0, 1]$ is a well-known result in the theory of moments. See Chapter III of Shohat and Tamarkin (1943) or Chapter IV of Karlin and Shapley (1953). We give a brief description of these results in terms of the matrices $\underline{\triangle}_k$ and $\overline{\triangle}_k$ defined as follows. For k even, let $\underline{\triangle}_k$, (resp. $\overline{\triangle}_k$), be the $(\frac{k}{2}+1) \times (\frac{k}{2}+1)$ matrix, (resp. $\frac{k}{2} \times \frac{k}{2}$ matrix), whose ij^{th} element is m_{i+j-2}, (resp. $m_{i+j-1} - m_{i+j}$). For k odd, let $\underline{\triangle}_k$, (resp. $\overline{\triangle}_k$), be the $(\frac{k+1}{2}) \times (\frac{k+1}{2})$ matrix whose ij^{th} element is m_{i+j-1}, (resp. $m_{i+j-2} - m_{i+j-1}$), where m_0 is defined to be 1. Then, a necessary and sufficient condition for the existence of a distribution on $[0, 1]$ with moments $m_1, \ldots, m_k$ is that the determinants of $\underline{\triangle}_1, \overline{\triangle}_1, \ldots, \underline{\triangle}_k, \overline{\triangle}_k$ be nonnegative.

Using this result, we may check whether the μ_i found by solving $\mathbf{A}\mu = \mathbf{b}$, is a moment sequence of a distribution on $[0, 1/4]$ by checking whether $m_i = 4^i \mu_i$ is a moment sequence on the interval $[0, 1]$. This was carried out on a computer for values of n from 3 to 33 inclusive, where it was determined that for $n = 3$ and $5 \leq n \leq 13$ the sequence μ_i is indeed a moment sequence. ■

For $n = 4$ and $14 \leq n \leq 33$, it was determined that the μ_i is not a moment sequence, so that the corresponding equalizer rule is not a Bayes rule.

For $n = 4$, it is of interest to find how much smaller we can make the minimax risk than the value $.00512425\cdots$ achieved by the equalizer rule. The risk function of a rule d as a function of $y = \min(W_4, 4 - W_4)$ is

$$R_1(\theta, d) = (1 - 4\theta + 2\theta^2)(\theta - d(0))^2 + 4\theta(1 - 2\theta)(\theta - d(1))^2 + 6\theta^2(\theta - d(2))^2,$$

which is a cubic function of θ. The maximum of this risk can be found at the boundary or at the roots of $(\partial/\partial\theta)R_1(\theta, d) = 0$. One may then use a numerical procedure, such as the Nelder-Meade downhill simplex method, to find the minimum of $\max_\theta R_1(\theta, d)$ as a function of the three variables, $d(0)$, $d(1)$ and $d(2)$. Algebraically, the least favorable distribution is determined by two equations in two unknowns, π and z. The first equation is that the risk function of the Bayes rule with respect to the prior have a local maximum at $\theta = z$. The other is that the value of the risk function at $\theta = 1/4$ be equal to the value of the risk function at $\theta = z$. We find

Theorem 4.4. *For $n = 4$ and squared error loss, the minimax rule is given by $d(0) = .07151065\cdots$, $d(1) = .2024725\cdots$ and $d(2) = .2443337\cdots$, and the minimax value is $.00512378\cdots$. This rule is Bayes with respect to the prior distribution giving probability $\pi = .5138768\cdots$ to the point $\theta = .25$ and probability $1 - \pi$ to the point $\theta = z = .04313538\cdots$.*

This gives an improvement to the minimax value of only $.00000047\cdots$ over the equalizer rule. Next, we consider the case $n = 2$.

Theorem 4.5. *For $n = 2$ and squared error loss, the invariant rule defined by $d(0) = (\sqrt{2} - 1)/4$ and $d(1) = 1/4$ is admissible and minimax. This rule has the form $\tilde{a}_2\tilde{\theta}_2 + \tilde{b}_2$, where $\tilde{a}_2 = 1 - (\sqrt{2}/2)$ and $\tilde{b}_2 = (\sqrt{2} - 1)/4$.*

Proof. The risk is

$$\begin{aligned} R_1(\theta, d) &= (1 - 2\theta)(\theta - d(0))^2 + 2\theta(\theta - d(1))^2 \\ &= \theta^2(1 + 4d(0) - 4d(1)) + \theta(-2d(0) - 2d(0)^2 + 2d(1)^2) + d(0)^2. \end{aligned}$$

Since $0 \leq \theta \leq 1/4$, we may take $0 \leq d(0) \leq 1/4$ and $0 \leq d(1) \leq 1/4$ without loss of generality. Therefore, the coefficient of θ^2 is $1 + 4d(0) - 4d(1) \geq 0$. Hence $R_1(\theta, d)$ is convex in θ, and

$$\begin{aligned} \max_{0 \leq \theta \leq 1/4} R_1(\theta, d) &= \max\{R_1(0, d), R_1(1/4, d)\} \\ &= \max\{d(0)^2, \frac{1}{2}d(0)^2 - \frac{1}{4}d(0) + \frac{1}{2}d(1)^2 - \frac{1}{4}d(1) + \frac{1}{16}\}. \end{aligned}$$

The minimum of this function over $d(1)$ occurs at $d(1) = 1/4$. Thus,

$$\min_{d(1)} \max_{\theta} R_1(\theta, d) = \max(d(0)^2, \frac{1}{2}d(0)^2 - \frac{1}{4}d(0) + \frac{1}{32}). \tag{4.6}$$

Equation (4.6) has a minimum over $d(0)$ at $d(0)^2 = d(0)^2/2 - d(0)/4 + 1/32$, i.e. $d(0) = (\sqrt{2}-1)/4$. The minimum value is $\min_{d(0),d(1)} \max_{\theta} R_1(\theta, d) = (\sqrt{2}-1)^2/16$, which is achieved at $d(0) = (\sqrt{2}-1)/4$ and $d(1) = 1/4$. This shows that the stated rule is minimax. Since it is a unique Bayes rule among invariant rules, it is admissible among invariant rules and hence admissible. ■

We remark that this rule is not an equalizer rule. It has smaller maximum risk than the equalizer rule. The least favorable distribution is concentrated at $\theta = 0$ and $1/4$ with probabilitities π and $1 - \pi$, where $\pi = \sqrt{2} - 1$.

Theorem 4.6. *Consider the problem of estimating θ with loss $L_2(p, a) = (\theta - a)^2/\theta$ where $\theta = p(1-p)$, based on $W \in \mathcal{B}(n, p)$. The estimate $d_n = a_n\tilde{\theta}_n$ is minimax and admissible when $2 \leq n \leq 11$.*

Proof. We follow the proof of Theorem 4.3 but we must modify the argument. The result for $n = 2$ and $n = 3$ follows directly from Theorem 4.2 since then $\tilde{\theta}_n$ takes on only two values and any function defined on two values is linear. We restrict attention to $n \geq 4$ and to rules that are functions of $Y_n = \min\{W_n, n - W_n\}$.

First note that it is sufficient to show that d_n is minimax within the class $\mathcal{C}_0$ of rules d that have $d(0) = 0$. For if $d(0) > 0$, the risk function is unbounded above, and if $d(0) < 0$, we may replace $d(0)$ with 0 and obtain an everywhere smaller risk. Second, note by the principle stated before Theorem 4.1 that it is sufficient to show that d_n is Bayes within the class $\mathcal{C}_0$.

Analogous to (4.5) we have d_n is Bayes within $\mathcal{C}_0$ with respect to a prior π if

$$\frac{\mathrm{E}_\pi \theta^y q_{n-2y}(\theta)}{\mathrm{E}_\pi \theta^{y-1} q_{n-2y}(\theta)} = d_n(y) = a_n \frac{y(n-y)}{n(n-1)} \quad \text{for all} \quad y = 1, \ldots, \lfloor n/2 \rfloor. \tag{4.7}$$

This is a linear system of $\lfloor n/2 \rfloor$ equations in the first $\lfloor n/2 \rfloor$ moments of π, which may be written as $\mathbf{A}\mu = \mathbf{b}$, where $\mu = (\mu_1, \ldots, \mu_{\lfloor n/2 \rfloor})^T$, where $\mathbf{b} = (-a_n, 0, \ldots, 0)^T$, and where is the $\lfloor n/2 \rfloor \times \lfloor n/2 \rfloor$ matrix whose i, j^{th} entry is given by

$$A_{ij} = (-1)^{j-i+1}(d_n(i)b_{ij} + c_{ij}),$$

where

$$b_{ij} = \begin{cases} \frac{(n-2i)(n-i-j-2)!}{(j-i+1)!(n-2j-2)!}, & \text{for } i = 1, \ldots, \lfloor n/2 \rfloor - 1,\ j = i-1, \ldots, \lfloor n/2 \rfloor - 1 \\ 1 & \text{for } i = \lfloor n/2 \rfloor,\ j = \lfloor n/2 \rfloor - 1 \\ 0 & \text{otherwise,} \end{cases}$$

$$c_{ij} = \begin{cases} \frac{(n-2i)(n-i-j-1)!}{(j-i)!(n-2j)!}, & \text{for } i = 1, \ldots, \lfloor n/2 \rfloor - 1,\ j = i, \ldots, \lfloor n/2 \rfloor \\ 1 & \text{for } i = \lfloor n/2 \rfloor,\ j = \lfloor n/2 \rfloor - 1 \\ 0 & \text{otherwise.} \end{cases}$$

As in Theorem 4.3, by computing the determinants of $\underline{\triangle}_1, \overline{\triangle}_1, \ldots, \underline{\triangle}_{\lfloor n/2 \rfloor}, \overline{\triangle}_{\lfloor n/2 \rfloor}$ we may check whether or not the resulting sequence $\mu_1, \ldots, \mu_{\lfloor n/2 \rfloor}$ of numbers forms a moment sequence. When this is done it is found that this is a sequence of moments for $n = 4, \ldots, 11$, but not for $n = 12, \ldots, 31$. ■

Acknowledgement: The second author would like to thank the members of the Statistics Department at Stanford University for their hospitality and Professor Herman Chernoff for advice on the S program used in the computations described in this paper.

References.

Aggarwal, O. P. (1955) Some minimax invariant procedures for estimating a cumulative distribution function. *Ann. Math. Statist.* **26**, 450-462.

Brown, L. D. (1988) Admissibility in discrete and continuous invariant nonparametric estimation problems and their multivariate analogs. *Ann. Statist.* **16**, 1567-1593.

Brown, L. D., Chow, Mosuk and Fong, D. K. H. (1992) On the admissibility of the maximum likehood estimator of the binomial variance. *Canadian J. Statist.*, (to appear).

Cohen, M. P. and Kuo, L. (1985) The admissibility of the empirical distribution function. *Ann. Statist.* **13**, 262-271.

Ferguson, Thomas S. (1967) *Mathematical Statistics: A Decision Theoretic Approach.* Academic Press, New York.and London.

Hodges, J. L. and Lehmann, E. L. (1950) Some problems in minimax point estimation. *Ann. Math. Statist.* **21**, 182-197.

Karlin, S. and Shapley, L. S. (1953) *Geometry of Moment Spaces.* Memoire of the American Mathematical Society, No. 12, *Amer. Math. Soc.*, Providence, RI.

Lehmann, E. L. (1983) *Theory of Point Estimation.* John Wiley & Sons, New York.

Meeden, G., Ghosh, M. and Vardeman, S. (1985) Some admissible nonparametric and related finite population sampling estimators. *Ann. Statist.* **13**, 811-817.

Phadia, E. G. (1973) Minimax estimation of a cumulative distribution function. *Ann. Statist.* **1**, 1149-1157.

Shohat, J. A. and Tamarkin, J. D. (1943) *The Problem of Moments. Amer. Math. Soc.* New York.

Wilks, Samuel S. (1962) *Mathematical Statistics.* John Wiley & Sons, New York.

Yu, Q. (1989) Inadmissibility of the empirical distribution function in continuous invariant problems. *Ann. Statist.* **17**, 1347-1359.