The Annals of Mathematical Statistics
1968, Vol. 39, No. 2, 492–501

MINIMAX ESTIMATION OF A RANDOM PROBABILITY WHOSE FIRST N MOMENTS ARE KNOWN[1]

BY MORRIS SKIBINSKY

Brookhaven National Laboratory

1. Summary. Let N be a positive integer. In Section 2 an expository account in terms of moment space dependence is given of the Bayes estimate of a random probability Θ, relative to squared difference loss, from an observable X which given Θ is conditionally binomial (N, Θ). The risk and Bayes envelope functional are also considered in these terms. In Section 3 an explicit formulation is given for the minimax estimate of Θ when its first N moments are known. Theorem 2 characterizes the condition that a Bayes estimate have constant risk over the class of all "priors" which yield these moments. In Section 4, a transformation is introduced which puts the interior of the space of the first N moments for distributions on [0, 1] in one–one correspondence with the interior of the N-dimensional unit cube. This transformation is used to show that the supremum of the difference between minimax and Bayes risks over the class of all prior distributions is bounded above by 2^{-N}. Examples for $N = 1, 2$, and 3 in terms of the above transformation are considered in Section 5.

2. Introduction. Let Θ and X be random variables defined on a measurable space $(\Omega, \mathcal{A})$; the former distributed on the unit interval [0, 1], the latter on the integers $0, 1, \cdots, N$, where N is fixed. Let $\mathcal{P}$ denote the class of all probability measures on $\mathcal{A}$ which yield the above structure and are such that the conditional distribution of X given $\Theta = \theta$ is binomial with parameters N and θ. Let

$$m_i(P) = E_P\Theta^i = \int_\Omega \Theta^i \, dP, \qquad i = 0, 1, \cdots, \qquad P \in \mathcal{P},$$

and take

$$\mathbf{m} = (m_1, m_2, \cdots, m_N), \qquad \mu = m_{N+1}.$$

Let τ_N denote the class of all functions on $\{0, 1, \cdots, N\}$ to [0, 1]. The elements of τ_N (also their composite with X) will be referred to as estimates of Θ. We define the *risk* of an estimate t in τ_N under a probability measure P in $\mathcal{P}$ by

$$R_N(t, P) = E_P(tX - \Theta)^2.$$

Note that the risk depends upon P only through the distribution of Θ that is induced by P. Preliminary to consideration of the results of main concern to this paper, to motivate their development and introduce an appropriate notation, we make a series of seven expository remarks concerning some known properties of this risk and related functionals. The substance of Remarks 1, 2, 3, and 6

Received 7 June 1967; revised 18 October 1967.

[1] Work performed under the auspices of the U. S. Atomic Energy Commission.

together with proofs thereof (albeit in somewhat different notation and context) may be found in either [2] or [3].

Let $\mathcal{Q}$ denote the class of all probability measures on the Borel subsets of [0, 1]. For each positive integer n, let

$$M_n = \{(c_1, c_2, \cdots, c_n): c_i = \int_{[0,1]} x^i \, dQ(x), i = 1, 2, \cdots, n, Q \in \mathcal{Q}\}.$$

REMARK 1. For each positive integer n, M_n is convex, closed, bounded, and n-dimensional; the convex hull of $\{(a, a^2, \cdots, a^n): 0 \leqq a \leqq 1\}$. Moreover, a point $(c_1, c_2, \cdots, c_n)$ is interior to M_n if and only if $(c_1, c_2, \cdots, c_i)$ is interior to M_i, $i = 1, 2, \cdots, n$.

For $j = 0, 1, \cdots, n$, and $n = 1, 2, \cdots$, define ξ_{nj} on M_n by

$$\xi_{nj}(c_1, c_2, \cdots, c_n) = \sum_{k=j}^{n} (-1)^{j+k} \binom{n-j}{n-k} c_k. \tag{1}$$

REMARK 2.

$$P(X = j) = \binom{N}{j} \xi_{Nj}(\mathbf{m}(P)), \quad j = 0, 1, \cdots, N, \quad P \in \mathcal{P}.$$

Thus the distribution of X depends upon P precisely through the first N moments of Θ under P.

REMARK 3.

$$E_P\left(\binom{X}{1}\Big/\binom{N}{1}, \binom{X}{2}\Big/\binom{N}{2}, \cdots, \binom{X}{N}\Big/\binom{N}{N}\right) = \mathbf{m}(P), \qquad P \in \mathcal{P}. \tag{2}$$

If we let S_N denote the N-dimensional simplex generated by the $N + 1$ possible realizations of the random vector whose expectation is taken on the left-hand side of (2), we have that $M_N \subset S_N$. Moreover, the barycentric coordinates of a point $\mathbf{c} = (c_1, c_2, \cdots, c_N)$ in M_N relative to the simplex S_N are given by

$$\left(\binom{N}{0}\xi_{N0}(\mathbf{c}), \binom{N}{1}\xi_{N1}(\mathbf{c}), \cdots, \binom{N}{N}\xi_{NN}(\mathbf{c})\right).$$

It follows that

$$\xi_{Nj}(\mathbf{c}) > 0, \qquad j = 0, 1, \cdots, N, \qquad \forall \mathbf{c} \text{ interior to } M_N. \tag{3}$$

REMARK 4.

$$R_N(t, P) = f_N(t; m(P), \mu(P)), \qquad t \in \tau_N, \quad P \in \mathcal{P}$$

where for all t in τ_N and all $(\mathbf{c}, d)$ in M_{N+1},

$$f_N(t; \mathbf{c}, d) = c_2 + \sum_{j=0}^{N} \binom{N}{j} [\xi_{Nj}(\mathbf{c}) t^2(j) - 2\xi_{N+1,j+1}(\mathbf{c}, d) t(j)].$$

Hence, for each t in τ_N, the risk depends upon P precisely through the first $N + 1$ moments of Θ under P.

We shall say that $\hat{t}$ in τ_N is a *Bayes estimate* of Θ relative to a probability measure P in $\mathcal{P}$ if

$$R_N(\hat{t}, P) = \inf_{t \in \tau_N} R_N(t, P) = R_N(P), \qquad \text{say.}$$

The right-hand side above is the value at P of the *Bayes envelope functional.*

For each point $(\mathbf{c}, d)$ in M_{N+1} and for $j = 0, 1, \cdots, N$, define

$$(4) \qquad t_{\mathbf{c},d}(j) = \xi_{N+1,j+1}(\mathbf{c}, d)/\xi_{Nj}(\mathbf{c}), \qquad \xi_{Nj}(\mathbf{c}) > 0,$$
$$= 0, \qquad \xi_{Nj}(\mathbf{c}) = 0.$$

(By Remark 2, $\xi_{Nj}(\mathbf{c})$ cannot be negative).

REMARK 5. $t_{\mathbf{m}(P),\mu(P)}$ is in τ_N for each P in $\mathcal{P}$ and is a Bayes estimate of Θ relative to P. If $\mathbf{m}(P)$ is interior to S_N—hence if $\mathbf{m}(P)$ is interior to M_N—$t_{\mathbf{m}(P),\mu(P)}$ is the unique Bayes estimate of Θ relative to P.

Using the above remark, the Bayes envelope functional at P in $\mathcal{P}$ may be written

$$(5) \qquad R_N(P) = g_N(\mathbf{m}(P), \mu(P)),$$

where

$$(6) \qquad g_N(\mathbf{c}, d) = f_N(t_{\mathbf{c},d}\,; \mathbf{c}, d)$$
$$= c_2 - \sum_{j=0}^{N} t^2_{\mathbf{c},d}(j)\binom{N}{j}\xi_{Nj}(\mathbf{c})$$

for each $(\mathbf{c}, d)$ in M_{N+1}.

A probability measure P^* in a subclass $\mathfrak{M}$ of $\mathcal{P}$ will be called *least favorable* in $\mathfrak{M}$ if the Bayes envelope functional attains its supremum over $\mathfrak{M}$ at P^*. We say that an estimate t^* in τ_N is *minimax* in $\mathfrak{M}$ if $\sup_{P\varepsilon\mathfrak{M}} R_N(t, P)$ attains its infimum at $t = t^*$. Robbins in [5] was the first to apply (in broader context) this "generalized" minimax concept. It has been applied in [7], for $N = 1$ in the present context, to the class $\{P \,\varepsilon\, \mathcal{P}: P(a \leqq \Theta \leqq b) \geqq 1 - \alpha\}$ for arbitrary a, b, α. We shall be concerned here with subcollections of $\mathcal{P}$ indexed by points $\mathbf{c}$ in M_N and defined by

$$\mathfrak{M}_N(\mathbf{c}) = \{P \,\varepsilon\, \mathcal{P}: \mathbf{m}(P) = \mathbf{c}\}.$$

REMARK 6. $\mathbf{c}$ is on the boundary of M_N if and only if the class of distributions of Θ induced by P, as P varies over $\mathfrak{M}_N(\mathbf{c})$, contains precisely one member. (See Theorem 20.1, p. 64 of [2]).

REMARK 7. To each $\mathbf{c}$ in M_N, there corresponds an estimate which is minimax, and a probability measure which is least favorable in $\mathfrak{M}_N(\mathbf{c})$. Moreover

$$\sup_{P\varepsilon\mathfrak{M}_N(\mathbf{c})} R_N(t^*, P) = R_N(P^*)$$

whenever t^* is minimax, and P^* is least favorable in $\mathfrak{M}_N(\mathbf{c})$.

Remark 7 is an immediate consequence of standard game theoretic results (e.g. see [1], Theorem 2.51, p. 51 and Corollary 2, p. 53), and obvious properties of τ_N, $\mathfrak{M}_N(\mathbf{c})$, and the risk.

It follows by Remarks 5 and 7 for $\mathbf{c}$ interior to M_N that the Bayes estimate of Θ relative to a least favorable probability measure in $\mathfrak{M}_N(\mathbf{c})$ is minimax in $\mathfrak{M}_N(\mathbf{c})$. For $\mathbf{c}$ on the boundary of M_N the situation is made trivial by Remark 6. For in this case any Bayes estimate relative to a probability measure in $\mathfrak{M}_N(\mathbf{c})$ is of necessity minimax in $\mathfrak{M}_N(\mathbf{c})$.

3. The minimax estimate of Θ in $\mathfrak{M}_N(\mathbf{c})$. To isolate the dependence of the Bayes envelope functional on the $(N + 1)$st moment of Θ under P, we may write

$$(7)\quad \xi_{N+1,j+1}(\mathbf{c}, d) = \gamma_{Nj}(\mathbf{c}) + (-1)^{N+j} d, \quad j = 0, 1, \cdots, N, \quad (\mathbf{c}, d) \,\varepsilon\, M_{N+1},$$

where by (1)

$$\gamma_{Nj}(\mathbf{c}) = \sum_{k=j+1}^{N} (-1)^{j+1+k} \binom{N-j}{N+1-k} c_k, \quad j = 0, 1, \cdots, N - 1; \quad \gamma_{NN}(\mathbf{c}) = 0.$$

Then by (6) and (4), we have for $N = 1, 2, \cdots$, and all points $(\mathbf{c}, d)$ in M_{N+1} such that $\mathbf{c}$ is interior to M_N that

$$(8)\qquad g_N(\mathbf{c}, d) = B_N(\mathbf{c}) - A_N(\mathbf{c})(d - \hat{\nu}_{N+1}(\mathbf{c}))^2$$

where

$$(9)\qquad A_N = \sum_{j=0}^{N} \binom{N}{j} / \xi_{Nj},$$

$$\hat{\nu}_{N+1} = A_N^{-1} \sum_{j=0}^{N-1} (-1)^{N+1+j} \binom{N}{j} \gamma_{Nj} / \xi_{Nj}, \qquad N = 2, 3, \cdots$$

and also B_N is a function of $\mathbf{c}$ only. When $N = 1$, we have

$$(10)\qquad g_1(c_1, d) = \tfrac{1}{4} c_1(1 - c_1) - (c_1(1 - c_1))^{-1}(d - \tfrac{1}{2} c_1(1 + c_1))^2,$$

from which the definitions of A_1, B_1 and $\hat{\nu}_2$ may be abstracted.

Now define $\nu_{N+1}^{\pm}$ on M_N by taking

$$(11)\quad \nu_{N+1}^{-}(\mathbf{c}) = \min \{d: (\mathbf{c}, d) \,\varepsilon\, M_{N+1}\}, \quad \nu_{N+1}^{+}(\mathbf{c}) = \max \{d: (\mathbf{c}, d) \,\varepsilon\, M_{N+1}\}.$$

These minima and maxima exist in view of Remark 1, and may be computed from well known "Hankel" determinants. For example, see Sections 17 and 18 of [2]. Denote by ν_{N+1}^{*} the function on M_N whose value at $\mathbf{c}$ is $\nu_{N+1}^{-}(\mathbf{c})$, $\nu_{N+1}^{+}(\mathbf{c})$, or $\hat{\nu}_{N+1}(\mathbf{c})$, according as $\hat{\nu}_{N+1}(\mathbf{c})$ is less than $\nu_{N+1}^{-}(\mathbf{c})$, greater than $\nu_{N+1}^{+}(\mathbf{c})$, or otherwise. By (5), (8), and Remarks 3, 5 and 7, we then have

THEOREM 1. *Let $\mathbf{c}$ be an arbitrary point in M_N. Then P^* in $\mathfrak{M}_N(\mathbf{c})$ is least favorable in $\mathfrak{M}_N(\mathbf{c})$ if and only if*

$$m_{N+1}(P^*) = \nu_{N+1}^{*}(\mathbf{c}).$$

The estimate $t_{\mathbf{c},\nu_{N+1}^{*}(\mathbf{c})}$ is minimax in $\mathfrak{M}_N(\mathbf{c})$ It is the unique minimax estimate of Θ in $\mathfrak{M}_N(\mathbf{c})$, if and only if $\mathbf{c}$ is interior to S_N—hence, whenever $\mathbf{c}$ is interior to M_N. The supremum of its risk over $\mathfrak{M}_N(\mathbf{c})$ is $g_N(\mathbf{c}, \nu_{N+1}^{*}(\mathbf{c}))$.

An analogue for a standard device by means of which estimates minimax in a class $\mathfrak{M}$ are often derived is to find a P in $\mathfrak{M}$ such that the Bayes estimate of Θ relative to P has constant risk over $\mathfrak{M}$, e.g. see [5], Section 5. Such an estimate is of course automatically minimax in $\mathfrak{M}$. It is of interest to know for which P in $\mathfrak{M}_N(\mathbf{c})$ if any, the Bayes estimate relative to P has constant risk over $\mathfrak{M}_N(\mathbf{c})$.

THEOREM 2. *Let $\mathbf{c}$ be an arbitrary point interior to M_N and let $\hat{P}$ be in $\mathfrak{M}_N(\mathbf{c})$. Then the Bayes estimate of Θ relative to $\hat{P}$ has constant risk over $\mathfrak{M}_N(\mathbf{c})$ if and only if*

$$m_{N+1}(\hat{P}) = \hat{\nu}_{N+1}(\mathbf{c}).$$

Hence whenever $\mathbf{c}$ *is such that either*

$$\hat{\nu}_{N+1}(\mathbf{c}) < \nu_{N+1}^{-}(\mathbf{c}) \quad \text{or} \quad > \nu_{N+1}^{+}(\mathbf{c}) \tag{12}$$

there can exist no $\hat{P}$ *in* $\mathfrak{M}_N(\mathbf{c})$ *such that the Bayes estimate of* Θ *relative to* $\hat{P}$ *has constant risk over* $\mathfrak{M}_N(\mathbf{c})$.

PROOF. By (7) and Remark 4,

$$f_N(t;\, \mathbf{c},\, d) = c_2 + \sum_{j=0}^{N} \binom{N}{j}[\xi_{Nj}(\mathbf{c})t^2(j) - 2\gamma_{Nj}(\mathbf{c})t(j)] - 2h_N(t)\, d,$$

where

$$h_N(t) = \sum_{j=0}^{N} (-1)^{N+j}\binom{N}{j}t(j). \tag{13}$$

Hence an estimate t in τ_N has constant risk over $\mathfrak{M}_N(\mathbf{c})$ if and only if $h_N(t) = 0$. Let $\hat{d} = m_{N+1}(\hat{P})$, then the Bayes estimate of Θ relative to $\hat{P}$, namely $t_{\mathbf{c},\hat{d}}$, has constant risk over $\mathfrak{M}_N(\mathbf{c})$ if and only if $h_N(t_{\mathbf{c},\hat{d}}) = 0$. But the left-hand side, after simple manipulation, may be written $A_N(\mathbf{c})(\hat{d} - \hat{\nu}_{N+1}(\mathbf{c}))$. Q.E.D.

That the set determined by (12) need not be vacuous (it is vacuous for $N = 1$ and 2) is shown by example in Section 5.

4. The difference in risk between Bayes and minimax estimates. The empirical Bayes approach to statistical decision problems introduced and developed by Herbert Robbins [4], [5] considers a sequence $(X_1, \Theta_1), (X_2, \Theta_2), \cdots$ of independent pairs of random variables with common distribution identical to that of (X, Θ). Only the first member of each pair is observable. The problem is to estimate Θ_{n+1} from the corresponding value of X_{n+1} using an "adaptive" sequence of estimates which in some optimal and/or consistent sense approximate to the Bayes estimate of Θ or are otherwise optimal relative to the Bayes envelope functional.

In the case presently under consideration, the distribution of X depends upon P only through $\mathbf{m}(P)$, the first N moments of Θ under P. It follows that no adaptive sequence of estimates of Θ can exist which consistently and uniformly for all P in $\mathcal{P}$ approximates any estimate of Θ that depends upon P other than through $\mathbf{m}(P)$. In particular, a Bayes estimate cannot be so approximated since it depends upon P through the first $N + 1$ moments of Θ. Thus as pointed out in Section 5 of [5] for the case $N = 1$, an "asymptotically optimal" sequence of estimates does not exist. In the absence of prior information, the most we can ultimately hope to learn about P by experience (i.e. observing $X_1, X_2, \cdots$) is the class $\mathfrak{M}_N(\mathbf{m}(P))$ to which it belongs. We may in fact consistently and uniformly for all P in $\mathcal{P}$ approximate the minimax estimate in this class in such a way that the supremum over $\mathfrak{M}_N(\mathbf{m}(P))$ of the risks of the approximating estimates tend in probability (uniformly for all P in $\mathcal{P}$) to the corresponding supremum for the risk of the estimate minimax in $\mathfrak{M}_N(\mathbf{m}(P))$. By Theorem 1 and Remark 3, the sequence of estimates

$$t_{\hat{\mathbf{c}}_j, \nu_{N+1}^{*}(\hat{\mathbf{c}}_j)}, \qquad j = 1, 2, \cdots,$$

where

$$\hat{\mathbf{c}}_j = j^{-1} \sum_{i=1}^{j} \left(\binom{X_i}{1}\Big/\binom{N}{1}, \binom{X_i}{2}\Big/\binom{N}{2}, \cdots, \binom{X_i}{N}\Big/\binom{N}{N}\right), \qquad j = 1, 2, \cdots,$$

is just such an adaptable sequence.

Thus for each P in $\mathcal{P}$, the quantity

$$(14) \qquad \inf_{t\varepsilon\tau_N} \sup_{P'\varepsilon\mathfrak{M}_N(\mathrm{m}(P))}, R_N(t, P') - R_N(P) = W_N(P), \qquad \text{say},$$

represents the difference between a minimax risk ultimately uniformly attainable in probability (e.g. by means of observations on X_1, X_2, $\cdots$ in an empirical Bayes framework), and the value of the Bayes envelope functional at P which is not so attainable. The risk associated with the standard estimate X/N is easily seen to be bounded above by $1/4N$ for all P in $\mathcal{P}$ and as an immediate consequence, $W_N(P)$ which is of course non-negative, has this bound as well. It therefore tends to zero uniformly in P as N tends to infinity. In fact, it tends to zero as N tends to infinity much more rapidly than $1/4N$ as the following theorem indicates.

THEOREM 3. *For each positive integer* N,

$$\sup_{P\varepsilon\mathcal{P}} W_N(P) \leqq 2^{-N}.$$

PROOF. By (5), (6), (14), and Theorem 1,

$$W_N(P) = D_N(\mathrm{m}(P), m_{N+1}(P)), \qquad \forall P \varepsilon \mathcal{P},$$

where

$$(15) \qquad D_N(\mathbf{c}, d) = g_N(\mathbf{c}, \nu^*_{N+1}(\mathbf{c})) - g_N(\mathbf{c}, d), \qquad \forall (\mathbf{c}, d) \varepsilon M_{N+1}.$$

By Remark 6, this is equal to zero whenever $\mathbf{c}$ is on the boundary of M_N. Suppose now that $\mathbf{c}$ is interior to M_N. By (6),

$$D_N(\mathbf{c}, d) = \textstyle\sum_{j=0}^N \binom{N}{j}[t^2_{\mathbf{c},d}(j) - t^2_{\mathbf{c},\nu^*_{N+1}(\mathbf{c})}(j)]\xi_{Nj}(\mathbf{c}).$$

By (3), (4), and (7)

$$t_{\mathbf{c},d}(j) - t_{\mathbf{c},\nu^*_{N+1}(\mathbf{c})}(j) = (-1)^{N+j}(d - \nu^*_{N+1}(\mathbf{c}))/\xi_{Nj}(\mathbf{c}).$$

Hence

$$(16) \quad D_N(\mathbf{c}, d) = (d - \nu^*_{N+1}(\mathbf{c})) \textstyle\sum_{j=0}^N (-1)^{N+j}\binom{N}{j}[t_{\mathbf{c},d}(j) + t_{\mathbf{c},\nu^*_{N+1}(\mathbf{c})}(j)]$$
$$\leqq (\nu^+_{N+1}(\mathbf{c}) - \nu^-_{N+1}(\mathbf{c})) \textstyle\sum_{j=0}^N \binom{N}{j} [1 + (-1)^{N+j}].$$

Elsewhere [6], we have shown that everywhere interior to M_N

$$(17) \qquad \nu^+_{N+1} - \nu^-_{N+1} = \textstyle\prod_{j=1}^N p_j q_j = r_N, \qquad \text{say},$$

where

$$(18) \qquad p_j = 1 - q_j = (\nu_j - \nu_j^-)/(\nu_j^+ - \nu_j^-),$$

ν_j is a function which assigns to any moment sequence, its jth coordinate; and $\nu_j^{\pm}$ are defined as in (11) with N replaced by $j - 1$ and $\mathbf{c} = (c_1, c_2, \cdots, c_{j-1})$. (If we define the right-hand side of (17) to be zero on the boundary of M_N, then (17) holds there as well). It follows from (17) that

$$\nu^+_{N+1}(\mathbf{c}) - \nu^-_{N+1}(\mathbf{c}) \leqq 2^{-2N}, \qquad \forall \mathbf{c} \varepsilon M_N,$$

which bound is attained when $p_j = \frac{1}{2}, j = 1, 2, \cdots, N$. That the width of M_n in the c_n direction is 2^{-2n+2} was first discovered by Karlin and Shapely [2]. On the other hand, the sum which appears on the right-hand side of (16) is precisely 2^N. Q.E.D.

The transformations (18) which put the interior of M_N in one–one correspondence with the interior of the N-dimensional unit cube may be used to obtain further insight into the nature of the Bayes envelope functional and the "least favorable" $(N+1)$st moment. Define

$$(19) \qquad t_{\mathbf{c}}^{\pm} = t_{\mathbf{c},\nu_{N+1}^{\pm}(\mathbf{c})}, \qquad g_N^{\pm}(\mathbf{c}) = g_N(\mathbf{c}, \nu_{N+1}^{\pm}(\mathbf{c})), \qquad \forall\, \mathbf{c} \,\varepsilon\, M_N .$$

THEOREM 4. *On the interior of* M_N, *let*

$$(20) \qquad \hat{p}_{N+1} = 1 - \hat{q}_{N+1} = (\nu_{N+1} - \nu_{N+1}^{-})/(\nu_{N+1}^{+} - \nu_{N+1}^{-})$$

then $\hat{p}_2$, $\hat{q}_2$ *are identically* $\frac{1}{2}$;

$$(21) \qquad \hat{p}_{N+1} = -h_N(t_{\cdot}^{-})/A_N r_N, \qquad \hat{q}_{N+1} = h_N(t_{\cdot}^{+})/A_N r_N, \qquad N = 2, 3, \cdots;$$

(h_N *defined by* (13); A_N, by (9)); *and*

$$(22) \qquad \begin{aligned} g_N &= g_N^{-} + A_N r_N^2[\hat{p}_{N+1}^2 - (p_{N+1} - \hat{p}_{N+1})^2], \\ &= g_N^{+} + A_N r_N^2[\hat{q}_{N+1}^2 - (q_{N+1} - \hat{q}_{N+1})^2], \end{aligned}$$

at each point $(\mathbf{c}, d)$ *of* M_{N+1} *such that* $\mathbf{c}$ *is interior to* M_N.

PROOF. Simple manipulations of (10) yield the theorem for $N = 1$. (Observe that $g_1^{\pm}$ are identically 0 on M_1). Now for $j = 0, 1, \cdots, N$,

$$(23) \qquad \xi_{N+1,j+1} = \xi_{N+1,j+1}^{-} + (-1)^{N+j} r_N p_{N+1} = \xi_{N+1,j+1}^{+} - (-1)^{N+j} r_N q_{N+1} .$$

where $\xi_{N+1,j+1}^{\pm}$ are defined on M_N analogously to (19). Hence using (6), (4), and (18), we have

$$\begin{aligned} g_N &= g_N^{-} - 2r_N p_{N+1} \textstyle\sum_{j=0}^{N} (-1)^{N+j}\binom{N}{j} t_{\cdot}^{-}(j) - A_N r_N^2 p_{N+1}^2 , \\ &= g_N^{+} + 2r_N p_{N+1} \textstyle\sum_{j=0}^{N} (-1)^{N+j}\binom{N}{j} t_{\cdot}^{+}(j) - A_N r_N^2 q_{N+1}^2 . \end{aligned}$$

By (13) and completing the square, the result follows. Q.E.D.

COROLLARY. *If we exclude* $(\mathbf{c}, d)$ *in* M_{N+1} *for which* $\mathbf{c}$ *is on the boundary of* M_N,

$$(24) \qquad \begin{aligned} D_N &\leqq \tfrac{1}{4} A_N r_N^2 (1 + |\hat{p}_{N+1} - \hat{q}_{N+1}|)^2, && 0 \leqq \hat{p}_{N+1} \leqq 1 \\ &\leqq A_N r_N^2\, |\hat{p}_{N+1} - \hat{q}_{N+1}|, && \text{otherwise.} \end{aligned}$$

This bound is attained for each $\mathbf{c}$ *interior to* M_N *at one or the other extreme value of the* $(N+1)$st *moment corresponding to* $\mathbf{c}$.

PROOF. The right-hand side of (24) is just max (D_N^{-}, D_N^{+}), where $D_N^{\pm}$ are defined analogously to (19). That this in fact is an upper bound may be seen from (15) and (8). Its form is easily derived from (22). As pointed out in the proof of Theorem 3, D_N is zero at points excluded from the corollary.

5. Examples: $N = 1, 2$ and 3. If we invert the transformation (18) for $N = 1, 2, 3, 4$ (keeping to the interior of M_{N-1} for $N = 2, 3, 4$), we find

$$\nu_1 = p_1, \qquad \nu_2 = p_1^2 + r_1 p_2, \qquad \nu_3 = p_1(p_1 + q_1 p_2)^2 + r_2 p_3$$

$$\nu_4 = p_1^2(p_1 + q_1 p_2)^2 + r_1 p_2(p_1 + q_1 p_2 + q_2 p_3)^2 + r_3 p_4 .$$

By (18), ν_N^- and ν_N^+ may be obtained from these by setting $p_N = 0$ and 1, respectively. When $N = 1$, the Bayes estimate of Θ, by (4) and (1), takes the values $p_1 q_2$ and $1 - q_1 q_2$ at 0 and 1, respectively, whereas the minimax estimate has corresponding values $\frac{1}{2}p_1$ and $1 - \frac{1}{2}q_1$. (c.f., Section 5 of [5]). The Bayes envelope functional value at any P in $\mathcal{P}$ is in this case precisely the range of the third moment in M_3 which corresponds to the first two moments of Θ under P, i.e., manipulating (10), we obtain

$$g_1 = p_1 q_1 p_2 q_2 = r_2 = \nu_3^+ - \nu_3^-.$$

As may be seen from (10), the least favorable second moment is given by

$$\nu_2^* = \mathfrak{p}_2 = p_1(1 + p_1)/2$$

so that as already pointed out in Theorem 4, $\hat{p}_2 = \hat{q}_2 = \frac{1}{2}$. The difference D^1 takes the form $r_1(p_2 - \frac{1}{2})^2$. Hence

$$\sup_{P \epsilon \mathcal{P}} W_1(P) = \tfrac{1}{16}.$$

When $N = 2$, the Bayes estimate of Θ takes values

$$p_1 q_2(q_1 q_2 + p_2 p_3)/(1 - p_1 q_2), \quad p_1 q_2 + p_2 q_3, \quad 1 - q_1 q_2(p_1 q_2 + p_2 q_3)/(1 - q_1 q_2)$$

the minimax,

$$p_1 q_2/(1 + p_2), \qquad (1 - q_1 q_2)/(1 + p_2), \qquad 1 - q_1 q_2/(1 + p_2),$$

respectively at 0, 1, and 2. The Bayes envelope functional value at P in $\mathcal{P}$ now depends on P through the first three moments of Θ under P. It takes the form given by

$$g_2 = r_2(1 + p_2)^{-1}[1 - p_2(\hat{p}_3\hat{q}_3)^{-1}(p_3 - \hat{p}_3)^2],$$

with

$$\hat{p}_3 = (1 - q_1 q_2)/(1 + p_2) = \tfrac{1}{2} + (p_1 - q_1)q_2/2(1 + p_2).$$

The least favorable third moment itself is

$$\mathfrak{p}_3 = p_1(1 - q_1 q_2)(1 - q_1 q_2/(1 + p_2)).$$

It is somewhat interesting to note that the values of the minimax estimate given above may be written

$$\hat{p}_3 - p_2/(1 + p_2), \qquad \hat{p}_3, \qquad \hat{p}_3 + p_2/(1 + p_2).$$

Finally, we have

$$D_2 = r_2 p_2(p_3 - \hat{p}_3)^2/(1 + p_2)\hat{p}_3\hat{q}_3$$

$$\leqq r_2(1 + p_2)^{-1} p_2 \max(\hat{p}_3/\hat{q}_3, \hat{q}_3/\hat{p}_3) \leqq r_2/(1 + p_2)$$

so that $\sup_{P \epsilon \mathcal{P}} W_2(P) \leqq (3 - 2(2)^{\frac{1}{2}})/4 \cong .04$.

From the above considerations it is easily seen that the set determined by (12) is empty when $N = 1$ and 2. We show below that such is not the case when $N = 3$. After investigation, we find that when $N = 3$, the Bayes estimate of Θ may be written in the form

$$
\begin{aligned}
&t_{(\cdot,\cdot)}(j) \\
&= p_1q_2 + p_2q_3 - p_2q_3(1 - p_1p_2p_3q_4)/[(1 - p_1q_2)^2 + p_1p_2q_2q_3], && j = 0,\\
(25)\quad &= p_1q_2 + p_2q_3 - p_2p_3q_3q_4/(q_1q_2 + p_2p_3), && j = 1,\\
&= p_1q_2 + p_2q_3 + p_2p_3q_3q_4/(p_1q_2 + p_2q_3), && j = 2,\\
&= p_1q_2 + p_2q_3 + p_2p_3(1 - q_1q_2q_3q_4)/[(1 - q_1q_2)^2 + q_1p_2q_2p_3], && j = 3.
\end{aligned}
$$

From this we may obtain $t.^+$ (defined in (19)) by setting $q_4 = 0$. We then find that

$$(26)\quad h_3(t.^+) = p_2\{q_3/[(1 - p_1q_2)^2 + p_1p_2q_2q_3] + p_3/[(1 - q_1q_2)^2 + q_1p_2q_2p_3]\}.$$

With the exception of points (c_1, c_2, c_3) in M_3 such that (c_1, c_2) is on the boundary of M_2 (the right-hand side is undefined at these points), this is non-negative everywhere on M_3. The left-hand side is defined everywhere on M_3 by (13), (19), and (4) and takes the value 0 or 1 at these exceptional points. It follows from (20) and (21) that $\hat{\nu}_4 \leqq \nu_4^+$ everywhere on M_3. On the other hand for p_1, p_2 arbitrary fixed numbers in the open interval (0, 1), i.e. for (c_1, c_2) arbitrary, fixed interior to M_2, and p_3 tending either to 0 or 1, (26) tends to a positive constant. Moreover, $\xi_{3j}^{\pm}$, $j = 0, 1, 2, 3$, are all everywhere positive interior to M_2 so that by (23) and (9), A_3 is positive and bounded as p_3 tends to either of its extremes. Since r_3 tends to zero in the same circumstances, $\hat{q}_4$ tends to infinity. By (20) and (21), it follows that $\hat{\nu}_4(c_1, c_2, c_3) < \nu_4^-(c_1, c_2, c_3)$ for arbitrarily selected (c_1, c_2) interior to M_2 whenever c_3 is sufficiently close to either of its extremes.

Let us now suppose for additional simplicity that $p_1 = p_2 = \frac{1}{2}$ i.e. that $(c_1, c_2) = (\frac{1}{2}, \frac{3}{8})$. In this particular case, the range of the 4th moment is

$$
\begin{aligned}
r_3 &= p_3q_3/16,\\
\nu_4^- &= 2^{-7}(27 + 16p_3 - 4p_3q_3),\\
\hat{\nu}_4 &= 2^{-7}(72 + 16p_3 - 18(99 + 4p_3q_3)/(39 + 4p_3q_3))
\end{aligned}
$$

and

$$\hat{\nu}_4 < \text{ or } \geqq \nu_4^- \quad \text{according as} \quad 4p_3q_3 < \text{ or } \geqq a_0 = 6(31)^{\frac{1}{2}} - 33 \cong .4.$$

To put this in terms of c_3, we have

$$\hat{\nu}_4(\tfrac{1}{2}, \tfrac{3}{8}, c_3) \geqq \nu_4^-(\tfrac{1}{2}, \tfrac{3}{8}, c_3), \qquad 9.77/32 \leqq c_3 \leqq 10.23/32.$$

For other c_3 in the allowable interval $[\frac{9}{32}, \frac{11}{32}]$, the opposite inequality holds. From the above formulae we easily find that when $p_1 = p_2 = \frac{1}{2}$,

$$\hat{q}_4 = \tfrac{1}{2} + 27(1 - 4p_3q_3)/8p_3q_3(39 + 4p_3q_3).$$

It is of interest to note that this is $\geqq \frac{1}{2}$, $0 < p_3 < 1$. The minimax estimate of Θ may be obtained by appropriate substitution in (25). i.e. $p_1 = p_2 = \frac{1}{2}$, and $\hat{q}_4$ or 1 for q_4 according as $4p_3q_3 \geqq$ or $< a_0$.

By (24), taking $y = 4p_3q_3$, and noting that in the present case,

$$A_3 = 2^{10}(39 + y)/(3 + y)(99 + y).$$

D_3 is sharply bounded above by

$$(3 + y)(9 + y)^2/16(39 + y)(99 + y) \quad \text{or} \quad 27y(1 - y)/4(3 + y)(99 + y),$$

according as $y \geqq$ or $< a_0$. This bound is attained when c_4 equals one or the other of its two extreme values. Thus, after some simple arithmetic, we find that $\sup W_3(P)$ (over all P in $\mathcal{P}$ such that $E_P\Theta = \frac{1}{2}$, $E_P\Theta^2 = \frac{3}{8}$) is less than 1/150.

6. Acknowledgment and conjecture. I am indebted to Herbert Robbins for suggesting the minimax problem that has been treated in this paper. A related conjecture due to Robbins (oral communication to the author) is that

$$\lim_{N\to\infty} \sup_{P\in\mathcal{P}} (W_N(P)/R_N(P)) = 0.$$

We agree with this conjecture, though we have as yet found no proof. We note here as a curiosity however that using the above developed technique, it is easily shown that the above ratio is unbounded on $\mathcal{P}$ for $N = 1$ and 2. This does not appear to be the case for $N = 3$.

REFERENCES

[1] Blackwell, D. and Girshick, M. A. (1954). *Theory of Games and Statistical Decisions*. Wiley, New York.

[2] Karlin, S. and Shapley, L. S. (1953). *Geometry of Moment Spaces*. Memoirs of the American Mathematical Society. Number 12, Providence.

[3] Karlin, S. and Studden, W. J. (1966). *Tchebycheff Systems: With Applications in Analysis and Statistics*. Interscience, New York.

[4] Robbins, H. (1955). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1** 157–164. Univ. of California Press.

[5] Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35** 1–20.

[6] Skibinsky, M. (1967). The range of the $(N + 1)$st moment for distributions on [0, 1]. *J. Appl. Prob.* **4**.

[7] Skibinsky, M. (1968). Minimax estimation of a random probability. *SIAM J. Appl. Math.* **16**