MINIMAX SOLUTION OF STATISTICAL DECISION PROBLEMS BY ITERATION[1]

BY WAYNE NELSON

University of Illinois and General Electric Research & Development Center

1. Introduction and summary. For the general statistical decision problem, Wald suggested the minimax solution as one possible choice of an optimum solution. Sufficient conditions for the existence of minimax solutions are well known. However, minimax solutions have been calculated only for isolated problems. A drawback of the methods that have been used to obtain minimax solutions is that there is no guarantee that they will yield a minimax solution when they are applied to a specific new problem. An exception to this is M. N. Ghosh's method [2] of approximating minimax estimators for a certain class of problems. Since his method differs from the method presented here, no further mention of his work will be made.

Presented in Section 3 is an iterative method of calculating minimax solutions that is applicable to a class of problems. The method is based on the result that, under certain conditions, if, for a sequence of *a priori* distributions on the parameter space, the corresponding sequence of Bayes risks converges to the supremum of all Bayes risks, then the corresponding sequence of Bayes decision functions converges to the minimax decision function and the corresponding sequence of risk functions converges uniformly to the minimax risk function. A method for iteratively constructing such a sequence of *a priori* distributions is given.

The derivation in Section 3 of the iterative method of calculating minimax solutions depends on some well-known results of Wald [5]. These results are stated in Section 2 for convenience of reference and are proved for the sake of completeness, although many of the proofs are available in [5] or elsewhere. Also, previously no exposition of the results of Section 2 was available in the literature with the mathematical presentation used here. The assumptions used here to obtain Wald's results are weaker than those used in [5] and are similar to Wald's Assumptions 5.1 through 5.6 in [6]. However, in [6] Wald did not give all of the results of Section 2. Also, the results hold for a larger class of problems than Wald indicated. Wald's restriction to fixed sample size problems is unnecessary; the results apply to sequential problems with prescribed sequential experimentation rule of the type considered by Wald in [6]. In [6], Wald treats discrete and continuous random variables separately. Here the results are given in a generality that includes discrete and continuous random variables as special cases.

Received 7 February 1966; revised 27 July 1966.

[1] This paper is part of doctoral dissertation submitted at the University of Illinois, September, 1965. The research was supported in part by the National Science Foundation Fellowhip and Grant GP-3814.

2. Assumptions and preliminary results. Let X be a chance variable with outcomes x in a sample space $\mathfrak{X}$ and let $B(\mathfrak{X})$ be a σ-field of subsets of $\mathfrak{X}$. The outcomes are not necessarily numerical. For example, for an experiment with fixed sample size n, an outcome is an n-tuple of observations, and, for a sequential experiment with a choice of experiments at each stage, an outcome is a sequence of experiments, each with a corresponding observation.

ASSUMPTION 1. The unknown distribution of X is a member of a given parametric class $\{P_\omega : \omega \text{ in } \Omega\}$ where Ω is a compact subset of a finite dimensional Euclidean space.

Let Ξ denote the class of all *a priori* distributions ξ on the σ-field $B(\Omega)$ of Borel subsets of Ω.

ASSUMPTION 2. (i) There is a σ-finite measure μ on $B(\mathfrak{X})$ such that, for every ω in Ω, P_ω has a density $p(x \mid \omega)$ with respect to μ. (ii) $p(\cdot \mid \cdot)$ is jointly measurable on the product σ-field $B(\mathfrak{X}) \times B(\Omega)$.

(For discrete random variables, μ is counting measure, and for continuous random variables with Euclidean sample space, μ is Lebesgue measure.)

(iii) For each x in $\mathfrak{X}$, $p(x \mid \cdot)$ is continuous on Ω.

This and the compactness of Ω imply that $p(x \mid \cdot)$ is bounded and uniformly continuous on Ω for each x in $\mathfrak{X}$.

ASSUMPTION 3. The set D of allowed terminal decisions d is a compact subset of a finite dimensional Euclidean space.

D and Ω are not necessarily the same set, nor do they necessarily have the same dimension.

ASSUMPTION 4. The loss function $L(\cdot, \cdot)$ is real-valued, non-negative, and jointly continuous on the Cartesian product $D \times \Omega$. $L(d, \omega)$ is the loss incurred if decision d is taken and the distribution of X has parameter ω.

This and the compactness of $D \times \Omega$ imply that $L(\cdot, \cdot)$ is bounded and uniformly continuous on $D \times \Omega$.

ASSUMPTION 5. The class Δ of allowed terminal decision functions δ consists of all measurable functions mapping $(\mathfrak{X}, B(\mathfrak{X}))$ into $(D, B(D))$ where $B(D)$ is the σ-field of Borel subsets of D.

ASSUMPTION 6. Let ξ be any *a priori* distribution in Ξ. Then, for any outcome x in $\mathfrak{X}$, except possibly on a set of μ-measure zero, there is at most one decision d in D that minimizes $\int L(d, \omega)p(x \mid \omega)\, d\xi(\omega)$.

Because of this assumption, randomized terminal decision functions need not be considered. Some well-known conditions that are sufficient for this assumption and that are satisfied in many estimation problems are: D is a convex set, $L(\cdot, \omega)$ is a strictly convex function on D for each ω in Ω, and, for almost all x, $p(x \mid \cdot) > 0$.

Ignoring for the moment all questions of existence, measurability, and integrability, we introduce the terminology and notation.

To judge a decision function δ, we use its *risk function* (expected loss) defined by

$$R_\delta(\omega) = \int L(\delta(x), \omega)p(x \mid \omega)\, d\mu(x)$$

for each ω in Ω. Let δ be a decision function. Then, if there is a decision function δ' such that $R_{\delta'}(\omega) \leqq R_\delta(\omega)$ for all ω in Ω with strict inequality for at least one ω, we say that δ is *inadmissible*. If no such δ' exists, we say that δ is *admissible* In view of wanting to minimize the risk function, we require that only admissible decision functions be used.

A decision function δ^* is called *minimax* if, for any decision function δ,

$$\sup_\omega R_{\delta^*}(\omega) \leqq \sup_\omega R_\delta(\omega).$$

$\int R_\delta(\omega)\, d\xi(\omega)$ is called the *average risk* of the decision function δ with respect to the *a priori* distribution ξ. A decision function δ_ξ satisfying

$$\int R_{\delta_\xi}(\omega)\, d\xi(\omega) = \inf_\delta \int R_\delta(\omega)\, d\xi(\omega)$$

is called *Bayes relative to* ξ, or, simply Bayes(ξ). R_{δ_ξ} is called a *Bayes*(ξ) *risk function* and is also denoted by R_ξ . $r(\xi) = \int R_{\delta_\xi}(\omega)\, d\xi(\omega) = \inf_\delta \int R_\delta(\omega)\, d\xi(\omega)$ is called the *Bayes risk for* ξ. $r^* = \sup_\xi r(\xi)$ is called the *maximum Bayes risk*. ξ^* is called a *least favorable a priori* distribution if $r(\xi^*) = r^* \equiv \sup_\xi r(\xi)$.

The main results of this section are that, under the assumptions, a unique and hence admissible minimax decision function exists, and it is Bayes relative to any least favorable *a priori* distribution. We now prove the preliminary results.

LEMMA 1. *The family of all risk functions* $\{R_\delta(\cdot): \delta$ *in* $\Delta\}$ *is equicontinuous on* Ω.

PROOF. Following the definition of equicontinuity in [4], we show that, given any ω' in Ω and any $\epsilon > 0$, there exists an $\eta > 0$ such that if $|\omega - \omega'| < \eta$ then $|R_\delta(\omega) - R_\delta(\omega')| < \epsilon$ for all δ in Δ.

$$\begin{aligned}
|R_\delta(\omega) - R_\delta(\omega')| &= \left|\int L(\delta(x), \omega)p(x \mid \omega)\, d\mu(x) - \int L(\delta(x), \omega')p(x \mid \omega')\, d\mu(x)\right| \\
&\leqq \left|\int L(\delta(x), \omega)p(x \mid \omega)\, d\mu(x) - \int L(\delta(x), \omega)p(x \mid \omega')\, d\mu(x)\right| \\
&\quad + \left|\int L(\delta(x), \omega)p(x \mid \omega')\, d\mu(x)\right. \\
&\quad \left. - \int L(\delta(x), \omega')p(x \mid \omega')\, d\mu(x)\right| \\
&\leqq M \int |p(x \mid \omega) - p(x \mid \omega')|\, d\mu(x) \\
&\quad + \int |L(\delta(x), \omega) - L(\delta(x), \omega')| p(x \mid \omega')\, d\mu(x)
\end{aligned}$$

where M is a finite upper bound for $L(\cdot, \cdot)$.

By the uniform continuity of $L(\cdot, \cdot)$ on $D \times \Omega$, there is an $\eta' > 0$ such that if $|\omega - \omega'| < \eta'$, then $|L(\cdot, \omega) - L(\cdot, \omega')| < \epsilon/2$. This implies that the last integral on the right is less than $\epsilon/2$.

For any sequence $\{\omega_i\}$ in Ω such that $\omega_i \to \omega'$, $p(x \mid \omega_i) \to p(x \mid \omega')$ for all x by Assumption 2(iii). Then, by Scheffé's theorem [3], p. 351, $\int |p(x \mid \omega_i) - p(x \mid \omega')|\, d\mu(x) \to 0$. So there is an $\eta'' > 0$ such that if $|\omega - \omega'| < \eta''$ then $\int |p(x \mid \omega) - p(x \mid \omega')|\, d\mu(x) < \epsilon/2M$. Let $\eta = \min(\eta', \eta'')$. Then, if $|\omega - \omega'| < \eta$, $|R_\delta(\omega) - R_\delta(\omega')| < \epsilon$ for all δ in Δ.

LEMMA 2. *Let* ξ *be any* a priori *distribution in* Ξ. *Then, for any* x *in* $\mathfrak{X}$, *there is* *a d in D that minimizes*

$$\int L(d, \omega)p(x \mid \omega)\, d\xi(\omega). \tag{1}$$

Proof. Assumptions 1, 2, and 4 imply that the integrand is measurable and bounded. So (1) exists for any x in $\mathfrak{X}$ and d in D.

We show that, for any x in $\mathfrak{X}$, (1) is continuous in d on the compact set D. This implies there is a d in D that minimizes (1). Let any x in $\mathfrak{X}$ and any $\epsilon > 0$ be given. $p(x \mid \cdot)$ is bounded and $L(\cdot, \cdot)$ uniformly continuous imply there is an $\eta > 0$ such that if $|d - d'| < \eta$ then $|L(d, \omega)p(x \mid \omega) - L(d', \omega)p(x \mid \omega)| < \epsilon$ for all ω in Ω. Hence

$$\left|\int L(d, \omega)p(x \mid \omega)\, d\xi(\omega) - \int L(d', \omega)p(x \mid \omega)\, d\xi(\omega)\right| < \epsilon$$

for $|d - d'| < \eta$, and (1) is continuous in d.

A Bayes(ξ) decision function δ_ξ is said to be *essentially unique* if, for any other Bayes(ξ) decision function δ_ξ', $\delta_\xi' = \delta_\xi$ a.e. μ.

Theorem 1. *For any* a priori *distribution ξ in Ξ, there exists a δ in Δ that is Bayes(ξ). It is essentially unique and hence admissible.*

Proof. For each x in $\mathfrak{X}$, define $\delta(x)$ equal to a d in D that minimizes (1). By Lemma 2, δ exists and, by Assumption 6, is uniquely defined a.e. μ. The definition of δ on the set of μ-measure zero is of no consequence. We show that δ satisfies the theorem.

We show that δ is measurable, i.e., in Δ. Denote (1) by $f(d, x)$. Since the integrand of (1) is jointly measurable on $B(\mathfrak{X}) \times B(\Omega)$, $f(d, \cdot)$ is $B(\mathfrak{X})$ measurable for each d in D by Tonelli's theorem [4]. Let $\{d_k\}$ be a dense subset of D. By the last proof, for each x in $\mathfrak{X}$, $f(\cdot, x)$ is continuous on D. Thus $\inf_d f(d, x) = \inf_k f(d_k, x)$ for all x in $\mathfrak{X}$. $\inf_d f(d, \cdot) = \inf_k f(d_k, \cdot)$ is $B(\mathfrak{X})$-measurable since $f(d_k, \cdot)$ is measurable for each k. Then $g(d, \cdot) = f(d, \cdot) - \inf_d f(d, \cdot)$ is measurable for each d in D. Also, $g \geqq 0$, $g(\delta(x), x) = 0$ for all x in $\mathfrak{X}$, and, for almost all x, $\delta(x)$ is the unique decision in D that minimizes $g(d, x)$.

To show that δ is measurable, we construct a sequence of measurable functions that converges to δ. Let $\{\epsilon_n\}$ be a sequence of positive numbers such that $\epsilon_n \to 0$ and define $A_{nk} = \{x: g(d_k, x) < \epsilon_n\}$, which is a measurable set for each n and k. For each n, the sets $B_{n1} = A_{n1}$, $B_{n2} = A_{n2} - B_{n1}$, $\cdots$, $B_{nk} = A_{nk} - \bigcup_{j=1}^{k-1} B_{nj}$, $\cdots$ are disjoint and measurable. x is in B_{nk} implies $g(d_k, x) < \epsilon_n$. We show that $\bigcup_k B_{nk} = \mathfrak{X}$ for each n. Let any x in $\mathfrak{X}$ be given. Since $f(\cdot, x)$ is continuous and $\{d_k\}$ is dense in D, there is a d_k such that $g(d_k, x) < \epsilon_n$; so x is in A_{nk}. Hence $\bigcup_k A_{nk} = \mathfrak{X}$ and $\bigcup_k B_{nk} = \bigcup_k A_{nk} = \mathfrak{X}$.

For each n and each x in $\mathfrak{X}$, define $\delta_n(x) = d_k$ if x is in B_{nk}, that is, $\delta_n(\cdot) = \sum_k d_k I_{B_{nk}}(\cdot)$ where I_B is the indicator function of B. Thus δ_n is measurable. Also, for all x in $\mathfrak{X}$, $g(\delta_n(x), x) < \epsilon_n$ for each n. For almost all x, $g(\cdot, x)$ is uniquely minimized by $\delta(x)$. Then it follows from the continuity of $g(\cdot, x)$ on D which is compact and from $g(\delta_n(x), x) \to 0$ that $\delta_n(x) \to \delta(x)$ for almost all x. Hence δ is a measurable function.

For the well-known proof that δ is Bayes(ξ), see the proof of Theorem 4.2 in

[5]. We will denote δ by δ_ξ hereafter. A slight modification of the proof of Theorem 4.2 in [5] yields the essential uniqueness of δ_ξ. Admissibility of an essentially unique Bayes decision function is well known.

Because the Bayes(ξ) decision function is essentially unique, we let δ_ξ be a generic notation for any member of the equivalence class of decision functions equal a.e. μ to the Bayes(ξ) decision function given above.

We say that the sequence $\{\xi_i\}$ of *a priori* distributions on Ω *converges weakly* to the *a priori* distribution ξ_0 if and only if, for every bounded continuous function g on Ω, $\lim \int g(\omega)\, d\xi_i(\omega) = \int g(\omega)\, d\xi_0(\omega)$. We denote this by $\xi_i \to \xi_0$.

LEMMA 3. *Let $\{f_i\}$ be any sequence of real-valued functions on Ω that converges uniformly to a continuous function f_0. Let $\{\xi_i\}$ be any sequence of* a priori *distributions on Ω that converges weakly to an* a priori *distribution ξ_0. Then*

$$\lim \int f_i(\omega)\, d\xi_i(\omega) = \int f_0(\omega)\, d\xi_0(\omega).$$

PROOF. The proof is immediate and will be omitted.

THEOREM 2. *Let $\{\xi_i\}$ be a sequence of* a priori *distributions on Ω that converges weakly to the* a priori *distribution ξ_0. Then*

(i) $\lim \delta_{\xi_i} = \delta_{\xi_0}$ a.e. μ,

(ii) $\lim R_{\xi_i} = R_{\xi_0}$ *uniformly on Ω, and*

(iii) $\lim r(\xi_i) = r(\xi_0)$.

PROOF. (i) Use proof by contradiction. Assume that for all x in a set M of positive μ-measure the equality does not hold. Let M_i, $i = 0, 1, 2, \cdots$, denote the null set on which δ_{ξ_i} is not given uniquely by Theorem 1. $\bigcup_i M_i$ has μ-measure zero, so $M' = M - \bigcup_i M_i$ has positive μ-measure. Let x be any point in M'. By the assumption above, $\lim\sup_i |\delta_{\xi_i}(x) - \delta_{\xi_0}(x)| > 0$. By the compactness of D, there is a subsequence $\{j\}$ of $\{i\}$ and a d_0 in D such that $\lim \delta_{\xi_j}(x) = d_0 \neq \delta_{\xi_0}(x)$. By Assumption 6,

$$\int L(d_0, \omega)p(x \mid \omega)\, d\xi_0(\omega) > \int L(\delta_{\xi_0}(x), \omega)p(x \mid \omega)\, d\xi_0(\omega). \tag{2}$$

The boundedness of $p(x \mid \cdot)$ and the joint continuity of $L(\cdot, \cdot)$ imply that the sequence $\{L(\delta_{\xi_j}(x), \omega)p(x \mid \omega)\}$ converges uniformly in ω to $L(d_0, \omega)p(x \mid \omega)$ which is continuous in ω. By Lemma 3, this implies

$$\lim \int L(\delta_{\xi_j}(x), \omega)p(x \mid \omega)\, d\xi_j(\omega) = \int L(d_0, \omega)p(x \mid \omega)\, d\xi_0(\omega). \tag{3}$$

Since $L(\delta_{\xi_0}(x), \cdot)p(x \mid \cdot)$ is continuous on Ω,

$$\lim \int L(\delta_{\xi_0}(x), \omega)p(x \mid \omega)\, d\xi_j(\omega) = \int L(\delta_{\xi_0}(x), \omega)p(x \mid \omega)\, d\xi_0(\omega). \tag{4}$$

(2), (3), and (4) yield

$$\lim \left[\int L(\delta_{\xi_j}(x), \omega)p(x \mid \omega)\, d\xi_j(\omega) - \int L(\delta_{\xi_0}(x), \omega)p(x \mid \omega)\, d\xi_j(\omega)\right] > 0.$$

It follows that the expression in brackets is positive for some j. This contradicts the definition of $\delta_{\xi_j}(x)$ and completes the proof of (i).

(ii) By (i), $\lim \delta_{\xi_i} = \delta_{\xi_0}$ a.e. μ. Thus the bounded convergence theorem and the continuity of $L(\cdot, \cdot)$ imply

$$\lim R_{\xi_i}(\omega) = \lim \int L(\delta_{\xi_i}(x), \omega)p(x \mid \omega)\, d\mu(x)$$
$$= \int L(\delta_{\xi_0}(x), \omega)p(x \mid \omega)\, d\mu(x) = R_{\xi_0}(\omega)$$

for each ω in Ω. Since $L(\cdot, \cdot)$ is bounded, all risk functions are uniformly bounded on Ω. This with Lemma 1 and the compactness of Ω implies that the convergence is uniform on Ω by Ascoli's theorem [4].

(iii) This follows from (ii), the continuity of R_{ξ_0}, and Lemma 3, since

$$\lim r(\xi_i) = \lim \int R_{\xi_i}(\omega)\, d\xi_i(\omega)$$
$$= \int R_{\xi_0}(\omega)\, d\xi_0(\omega) = r(\xi_0).$$

THEOREM 3. *There exists a least favorable* a priori *distribution.*

PROOF. Since $L(\cdot, \cdot)$ is bounded, $r^* = \sup_\xi r(\xi)$ is finite, and there is a sequence $\{\xi_i\}$ of *a priori* distributions such that $\lim r(\xi_i) = r^*$. By Helly's weak compactness theorem and the compactness of Ω, there is a subsequence $\{\xi_j\}$ of $\{\xi_i\}$ and an *a priori* distribution ξ^* such that $\xi_j \to \xi^*$. By Theorem 2(iii)

$$r(\xi^*) = \lim r(\xi_j) = r^*.$$

Hence ξ^* is least favorable. It is not necessarily unique.

THEOREM 4. (i) *An* a priori *distribution ξ^* is least favorable if and only if* $\max_\omega R_{\xi^*}(\omega) = r(\xi^*)$.

(ii) *If ξ^* is least favorable, then δ^* is minimax if and only if δ^* is Bayes(ξ^*).*

PROOF. By the compactness of Ω and the continuity of R_ξ for any ξ, $\max_\omega R_\xi(\omega)$ exists. Let ξ be any *a priori* distribution. Then the "if" part of (i) follows from

$$r(\xi) = \inf_\delta \int R_\delta(\omega)\, d\xi(\omega) \leqq \int R_{\xi^*}(\omega)\, d\xi(\omega)$$
$$\leqq \max_\omega R_{\xi^*}(\omega) = r(\xi^*).$$

For the "only if" part of (i), use proof by contradiction. For any ξ,

$$\max_\omega R_\xi(\omega) \geqq \int R_\xi(\omega)\, d\xi(\omega) = r(\xi).$$

Let ξ be least favorable and assume that

$$\max_\omega R_\xi(\omega) > r(\xi).$$

Above we saw that there is an ω^* in Ω that maximizes R_ξ. Let ζ denote the *a priori* distribution that puts probability one on ω^*. For $0 \leqq \lambda \leqq 1$, the mixture $\xi_\lambda(\cdot) = \lambda\zeta(\cdot) + (1 - \lambda)\xi(\cdot)$ is an *a priori* distribution on Ω, and $\xi_\lambda \to \xi$ as $\lambda \to 0^+$. We show that there is a λ^* such that

$$r(\xi_{\lambda^*}) > r(\xi), \tag{5}$$

which contradicts that ξ is least favorable and completes the proof. We prove

$$\lim_{\lambda \to 0^+} \{[r(\xi_\lambda) - r(\xi)]/\lambda\} > 0 \tag{6}$$

which implies (5) holds for some positive λ^*.

$$r(\xi_\lambda) = \int R_{\xi_\lambda}(\omega)[\lambda\, d\zeta(\omega) + (1 - \lambda)\, d\xi(\omega)]; \text{ so}$$

$$[r(\xi_\lambda) - r(\xi)]/\lambda = \{\int R_{\xi_\lambda}(\omega)\, d\zeta(\omega) - \int R_{\xi_\lambda}(\omega)\, d\xi(\omega)\} + \lambda^{-1}\{\int R_{\xi_\lambda}(\omega)\, d\xi(\omega) - r(\xi)\}.$$

The second expression in braces is non-negative for $0 \leqq \lambda \leqq 1$, since $r(\xi) = \inf_\delta \int R_\delta(\omega)\, d\xi(\omega)$. $\xi_\lambda \to \xi$ as $\lambda \to 0^+$ implies, by Theorem 2(ii), $\lim_{\lambda\to 0^+} R_{\xi_\lambda} = R_\xi$ uniformly on Ω. This and the continuity of R_ξ imply that the first expression in braces goes to $\{R_\xi(\omega^*) - r(\xi)\} > 0$ as $\lambda \to 0^+$. Hence (6) is true, and the "only if" part of (i) is proved.

Use proof by contradiction for the "if" part of (ii). Assume that δ^* is not minimax. Then there is a decision function δ such that

$$\max_\omega R_\delta(\omega) < \max_\omega R_{\delta^*}(\omega). \tag{7}$$

Since δ^* is Bayes(ξ^*) and ξ^* is least favorable, (i) implies

$$\max_\omega R_{\delta^*}(\omega) = r(\xi^*) = \int R_{\delta^*}(\omega)\, d\xi^*(\omega).$$

Thus the support of ξ^* consists only of ω points that maximize R_{δ^*}. This and (7) imply $\int R_\delta(\omega)\, d\xi^*(\omega) < \int R_{\delta^*}(\omega)\, d\xi^*(\omega)$. This contradicts that δ^* is Bayes (ξ^*) and completes the proof.

Use proof by contradiction for the "only if" part of (ii). Assume that δ^* is not Bayes(ξ^*). Then $r(\xi^*) < \int R_{\delta^*}(\omega)\, d\xi^*(\omega) \leqq \max_\omega R_{\delta^*}(\omega)$. Let δ_{ξ^*} be Bayes(ξ^*). Then by (i) $\max_\omega R_{\delta_{\xi^*}}(\omega) = r(\xi^*) < \max_\omega R_{\delta^*}(\omega)$. This contradicts that δ^* is minimax and completes the proof.

COROLLARY. *ξ^* is least favorable and δ_{ξ^*} is minimax if $R_{\delta_{\xi^*}}$ is constant.*

PROOF. This follows from Theorem 4, since

$$\max_\omega R_{\delta_{\xi^*}}(\omega) = \int R_{\delta_{\xi^*}}(\omega)\, d\xi^*(\omega) = r(\xi^*).$$

THEOREM 5. *There exists a minimax decision function; it is essentially unique and admissible. The minimax risk function R^* is unique and* $\max_\omega R^*(\omega) = r^* \equiv \sup_\xi r(\xi)$.

In view of this result, we call r^* the *minimax risk.*

PROOF. Existence and admissibility of a minimax decision function follow from Theorems 1, 3, and 4(ii). To prove essential uniqueness, assume that δ^* and δ^{**} are minimax. Let ξ^* be a least favorable *a priori* distribution. Then, by Theorem 4(ii), δ^* and δ^{**} are Bayes(ξ^*). Therefore, by Theorem 1, $\delta^* = \delta^{**}$ a.e. μ. The uniqueness of the minimax risk function $R^* = R_{\delta^*}$ follows from the essential uniqueness of δ^*. $\max_\omega R^*(\omega) = r^*$ follows from Theorem 4.

3. The iterative method. The following convergence theorem provides an iterative method of obtaining a minimax decision function for the statistical decision problem of Section 2.

THEOREM 6. *If $\xi_1, \xi_2, \cdots$ is a sequence of* a priori *distributions on Ω such that the corresponding sequence $r(\xi_1), r(\xi_2), \cdots$ of Bayes risks converges to the minimax* *risk r^*, then the corresponding sequence R_{ξ_1}, R_{ξ_2}, $\cdots$ of Bayes risk functions converges uniformly on Ω to the unique minimax risk function R^*.*

Later we show how to construct such a sequence of *a priori* distributions. An immediate consequence of this theorem is the following corollary which is implied by the uniform convergence of the sequence of risk functions.

COROLLARY. *If ξ_1, ξ_2, $\cdots$ is a sequence of* a priori *distributions on Ω such that the corresponding sequence $r(\xi_1)$, $r(\xi_2)$, $\cdots$ of Bayes risks converges to the minimax risk r^*, then $\lim_{i\to\infty} \sup_\omega R_{\xi_i}(\omega) = r^*$.*

The corollary says that we can get a decision function with maximum risk arbitrarily close to the minimax risk r^* if we take a Bayes decision function far enough out in the sequence δ_{ξ_1}, δ_{ξ_2}, $\cdots$. For example, if "close" means that ξ_k satisfies $\max_\omega R_{\xi_k}(\omega) - r^* \leqq \epsilon$ where ϵ is given, then it is sufficient to terminate the iteration with the first ξ_k that satisfies $\max_\omega R_{\xi_k}(\omega) - r(\xi_k) \leqq \epsilon$.

PROOF OF THEOREM 6. By Theorem 5, the minimax risk function R^* is unique. We assume that the sequence $\{R_{\xi_i}\}$ does not converge uniformly on Ω to R^* and arrive at a contradiction. Then there is an $\epsilon > 0$ and a subsequence $\{\xi_j\}$ of $\{\xi_i\}$ such that

$$\sup_\omega |R_{\xi_j}(\omega) - R^*(\omega)| \geqq \epsilon \tag{8}$$

or all j. By Helly's weak compactness theorem and the compactness of Ω, $\{\xi_j\}$ has a subsequence $\{\xi_k\}$ that converges weakly to an *a priori* distribution, say, ξ^*. By Theorem 2(iii), $\lim r(\xi_k) = r(\xi^*)$, and, by hypothesis, $\lim r(\xi_i) = r^*$. So $r(\xi^*) = r^*$ and ξ^* is least favorable. Then, by Theorem 4(ii), δ_{ξ^*} is minimax, and, by Theorem 5, $R_{\xi^*} = R^*$, the unique minimax risk function. But by Theorem 2(ii), $\lim R_{\xi_k} = R_{\xi^*} = R^*$ uniformly on Ω. This contradicts (8) and completes the proof.

The next theorem gives the convergence of the constructed sequence δ_{ξ_1}, δ_{ξ_2}, $\cdots$ of Bayes decision functions to the minimax decision function δ^* under a stronger version of Assumption 6. The stronger assumption is satisfied (1) if $\mathfrak{X}$ is countable and μ is counting measure or (2) if D is a convex set, $L(\cdot, \omega)$ is strictly convex on D for each ω in Ω, and $p(\cdot \mid \cdot) > 0$ for all x in $\mathfrak{X}$ and ω in Ω.

Because we judge a decision function solely by its risk function and seek a decision function having a risk function with as small a maximum as possible, the next theorem is of less interest than the corollary above.

THEOREM 7. *If Assumption 6 is satisfied for every x in $\mathfrak{X}$ and if ξ_1, ξ_2, $\cdots$ is a sequence of* a priori *distributions on Ω such that the corresponding sequence $r(\xi_1)$, $r(\xi_2)$, $\cdots$ of Bayes risks converges to the minimax risk r^*, then the sequence δ_{ξ_1}, δ_{ξ_2}, $\cdots$ of Bayes decision functions converges to the unique minimax decision function δ^*.*

PROOF. The hypothesis that Assumption 6 be satisfied for each x in $\mathfrak{X}$ implies that the results of Section 2 hold with "a.e. μ" replaced by "for each x in $\mathfrak{X}$." Use proof by contradiction. Assume that there is an x in $\mathfrak{X}$ such that $\{\delta_{\xi_i}(x)\}$ does not converge to $\delta^*(x)$, which is unique since δ^* is Bayes relative to any least favorable distribution. Then there is a subsequence $\{j\}$ of $\{i\}$ and an $\epsilon > 0$ such that

$$(9) \qquad |\delta_{\xi_j}(x) - \delta^*(x)| \geqq \epsilon > 0$$

for every j.

By Helly's weak compactness theorem and the compactness of Ω, there is a subsequence $\{k\}$ of $\{j\}$ and an *a priori* distribution ξ^* such that $\xi_k \to \xi^*$. By the reasoning used in the proof of Theorem 6, ξ^* is least favorable and δ_{ξ^*} is minimax. Hence $\delta_{\xi^*} = \delta^*$. By Theorem 2(i)

$$\lim \delta_{\xi_k}(x) = \delta_{\xi^*}(x) = \delta^*(x).$$

This contradicts (9) and completes the proof.

We now give an iterative method of constructing a sequence of *a priori* distributions satisfying the hypothesis of Theorems 6 and 7 that the corresponding sequence of Bayes risks converges to the minimax risk r^*. We first show that we can construct a sequence $\xi_1, \xi_2, \cdots$ of *a priori* distributions such that corresponding sequence of Bayes risks is strictly increasing $r(\xi_1) < r(\xi_2) < \cdots$. Since the sequence of Bayes risks is bounded above by $r^* = \sup_\xi r(\xi)$, it converges to a number less than or equal to r^*. We then show that, under the conditions of Theorem 8, the sequence of Bayes risks does indeed converge to r^*.

To start the iteration we choose an arbitrary initial *a priori* distribution ξ_1. An initial distribution that is in some sense close to a least favorable distribution should speed the convergence of the iteration.

We now present the general iterative step. Let ξ_k denote the distribution obtained in the last iteration. Theorem 4(i) is used to determine if ξ_k is least favorable. Suppose ξ_k is not least favorable, for if it were least favorable, the iteration is terminated. To simplify notation, denote the corresponding Bayes decision function by δ_k, its risk function by R_k, and the Bayes risk by r_k. From ξ_k we construct a parametrized *a priori* distribution $\xi_k(\lambda, \cdot)$, $0 \leqq \lambda \leqq 1$, with corresponding Bayes decision function $\delta_k(\lambda, \cdot)$, risk function $R_k(\lambda, \cdot)$, and Bayes risk $r_k(\lambda)$. By Lemma 4 there is a λ_k, $0 < \lambda_k \leqq 1$, such that $r_k < r_k(\lambda_k)$. $\xi_k(\lambda_k, \cdot)$ is then relabeled as $\xi_{k+1}(\cdot)$, and we have $r_k < r_{k+1}$ as desired. Also, then $\delta_{k+1}(\cdot) = \delta_k(\lambda_k, \cdot)$ and $R_{k+1}(\cdot) = R_k(\lambda_k, \cdot)$.

We now construct $\xi_k(\lambda, \cdot)$. Let ζ_k by any *a priori* distribution on Ω that satisfies

$$(10) \qquad \int R_k(\omega)\, d\zeta_k(\omega) > r_k .$$

Such a ζ_k must exist when ξ_k is not least favorable. One such ζ_k puts probability one on ω' in Ω such that $R_k(\omega') > r_k$. If such an ω' did not exist, then $R_k(\omega) \leqq r_k$ for all ω in Ω, and, by Theorem 4(i), this implies ξ_k is least favorable, contradicting that ξ_k is not least favorable.

For the moment, regard an *a priori* distribution ξ_k as a strategy played against the statistician by antagonistic Nature, whose payoff is the Bayes risk r_k. Since $\int R_k(\omega)\, d\zeta_k(\omega) > r_k$, Nature should gain a higher Bayes risk by playing ζ_k in addition to ξ_k. That is, Nature should do better by playing a mixture of ζ_k and ξ_k. For $0 \leqq \lambda \leqq 1$, define the *a priori* distribution

$$\xi_k(\lambda, \cdot) = \lambda\zeta_k(\cdot) + (1 - \lambda)\xi_k(\cdot)$$

which is a mixture of ζ_k and ξ_k. As $\lambda \to 0^+$, $\xi_k(\lambda, \cdot) \to \xi_k(\cdot)$. As mentioned before, for the mixture $\xi_k(\lambda, \cdot)$, we denote the corresponding Bayes decision function by $\delta_k(\lambda, \cdot)$ and its risk function by $R_k(\lambda, \cdot)$. The corresponding Bayes risk is

$$\begin{aligned} r_k(\lambda) &= \int R_k(\lambda, \omega)\, d\xi_k(\lambda, \omega) \\ &= \lambda \int R_k(\lambda, \omega)\, d\zeta_k(\omega) + (1 - \lambda) \int R_k(\lambda, \omega)\, d\xi_k(\omega). \end{aligned}$$

We now show that there is a mixture with Bayes risk higher than r_k.

LEMMA 4. *There is a* λ_k, $0 < \lambda_k \leqq 1$, *such that* $r_k < r_k(\lambda_k)$.

PROOF. This lemma follows from

$$\lim_{\lambda \to 0^+} [(r_k(\lambda) - r_k)/\lambda] > 0$$

which is proved in the same manner as the corresponding inequality (6) in the proof of Theorem 4(i).

The following well-known result aids in a numerical search for a λ_k that satisfies $r_k(\lambda_k) > r_k$. Also, it aids in a numerical search for a λ_k that satisfies condition (2) of Theorem 8.

LEMMA 5. *Let* $\xi(\lambda, \cdot) = \lambda\zeta(\cdot) + (1 - \lambda)\xi(\cdot)$, $0 \leqq \lambda \leqq 1$, *be a mixture of any two* a priori *distributions* ζ *and* ξ. *Then the corresponding Bayes risk* $r(\lambda)$ *is a continuous concave function for* $0 \leqq \lambda \leqq 1$.

PROOF. It is well known that, under assumptions weaker than the ones here, $r(\lambda)$ is concave for $0 \leqq \lambda \leqq 1$, which implies that $r(\lambda)$ is continuous for $0 < \lambda < 1$. Continuity at the end-points 0 and 1 follows from Theorem 2(iii).

We now prove a theorem giving conditions that guarantee that the sequence of Bayes risks of the iteratively constructed *a priori* distributions converges to the minimax risk. If one of the ξ_k is found least favorable by Theorem 4(i), then the iteration is terminated and is trivially convergent. So we need to consider only the case with no ξ_k least favorable for $k = 1, 2, \cdots$.

THEOREM 8. *At the* kth *stage of iteration* ($k = 1, 2, \cdots$), *if* ξ_k *is not least favorable and*

(1) *the chosen distribution* ζ_k *satisfies*

$$\int R_k(\omega)\, d\zeta_k(\omega) - r_k \geqq \alpha(\max_\omega R_k(\omega) - r_k)$$

where α *is any fixed number satisfying* $0 < \alpha < 1$ *and*

(2) *the chosen* λ_k *satisfies*

$$(r_k(\lambda_k) - r_k)/(\max_{0 \leqq \lambda \leqq 1} r_k(\lambda) - r_k) \geqq \beta$$

where β *is any fixed number satisfying* $0 < \beta < 1$, *then* $\lim r_k = r^*$.

Condition (1) requires only that the chosen distribution ζ_k put its probability on those ω in Ω that come close to maximizing R_k. The α and β can be chosen arbitrarily, but are fixed throughout the iteration. By Lemma 5, $\max_{0 \leqq \lambda \leqq 1} r_k(\lambda)$ exists, and, by Lemma 4, the denominator of condition (2) is positive. Condition (2) requires that, for each k, the increase in the Bayes risk $(r_k(\lambda_k) - r_k)$ achieved with the chosen λ_k be at least a fraction β of the maximum possible increase

$(\max_{0 \leqq \lambda \leqq 1} r_k(\lambda) - r_k)$. Lemma 6 provides a means of checking whether condition (2) is satisfied.

PROOF. *A priori* distributions ζ_k satisfying condition (1) exist; for example, one such ζ_k puts probability one on an ω that maximizes R_k. Since $\max_\omega R_k(\omega) > r_k$, condition (1) implies that ζ_k satisfies (10) and can be used to iterate on ξ_k. By Lemma 4, the constructed sequence $\xi_1, \xi_2, \cdots$ has a strictly increasing sequence of Bayes risks $r_1 < r_2 < \cdots$ bounded above by $r^* = \sup_\xi r(\xi)$. Thus the sequence of Bayes risks has a limit that is less than or equal to r^*. We assume that $\lim r_k = r_0 < r^*$ and derive a contradiction to condition (2).

By Helly's weak compactness theorem and the compactness of Ω, there is a subsequence $\{j\}$ of $\{k\}$ and *a priori* distributions ζ_0 and ξ_0 such that $\zeta_j \to \zeta_0$ and $\xi_j \to \xi_0$. By Theorem 2(iii),

$$r(\xi_0) = \lim r_j = r_0 < r^*.$$

Thus ξ_0 is not least favorable and the iteration can be used on it. We use ζ_0 to iterate on ξ_0; so we must show that ζ_0 satisfies (10). By Theorem 2(ii), $\lim R_j = R_{\xi_0}$ uniformly on Ω where R_{ξ_0} is continuous by Lemma 1. Lemma 3 and condition (1) imply

$$\begin{aligned} \int R_{\xi_0}(\omega)\, d\zeta_0(\omega) - r_0 &= \lim \left(\int R_j(\omega)\, d\zeta_j(\omega) - r_j\right) \\ &\geqq \alpha(\liminf \max_\omega R_j(\omega) - r_0) \geqq \alpha(r^* - r_0) > 0 \end{aligned}$$

since $\max_\omega R_j(\omega) \geqq r^*, j = 1, 2, \cdots$. Hence ζ_0 satisfies (10).

By Lemma 4, there is a mixture λ_0 of ζ_0 and ξ_0 such that $r_0(\lambda_0) > r_0$. Let $\epsilon = (r_0(\lambda_0) - r_0) > 0$. Since $\xi_j(\lambda_0, \cdot) = \lambda_0 \zeta_j(\cdot) + (1 - \lambda_0)\xi_j(\cdot) \to \xi_0(\lambda_0, \cdot)$, Theorem 2(iii) implies that $\lim r_j(\lambda_0) = r_0(\lambda_0)$. Thus there is a J such that $|r_0(\lambda_0) - r_j(\lambda_0)| < \frac{1}{2}\epsilon$ for all $j > J$. This implies that $(r_j(\lambda_0) - r_0) > \frac{1}{2}\epsilon$ for all $j > J$. $\lim r_j = r_0$ implies there is a J' such that $(r_0 - r_j) < \beta\frac{1}{2}\epsilon$ for all $j > J'$. Then for all $j > \max(J, J')$,

$$(r_j(\lambda_j) - r_j)/(\max_{0 \leqq \lambda \leqq 1} r_j(\lambda) - r_j) \leqq (r_0 - r_j)/(r_j(\lambda_0) - r_0) < \beta.$$

This contradicts condition (2) and completes the proof.

The following example of the iterative method was chosen for its computational simplicity. The result is well known. [1], Section 11.2, contains the results used in the example to obtain Bayes estimators.

EXAMPLE. Let X be binomial with one observation and parameter ω, that is, $p(x \mid \omega) = \omega^x(1 - \omega)^{1-x}$, $x = 0, 1$. We use the iterative method to calculate the minimax estimator for ω with squared error loss $L(d, \omega) = (d - \omega)^2$. Here $\Omega = D = [0, 1]$. Assumption 6 is not satisfied here for *a priori* distributions that put probability 1 on 0 or on 1. More generally, for binomial problems allowing more than one observation, Assumption 6 is not satisfied for *a priori* distributions that put probability π on 0 and $1 - \pi$ on 1. However, it is proved later that, under Assumption 7 which is satisfied here, the iterative method works.

Because of the symmetry of the problem, we work only with *a priori* distributions that are symmetric about $\omega = \frac{1}{2}$. Let the distribution that puts probability 1 on $\omega = \frac{1}{2}$ be ξ_1. Then $\delta_1(x) = \frac{1}{2}$ for $x = 0$ and 1, $R_1(\omega) = (\omega - \frac{1}{2})^2$, and $r_1 = 0$. R_1 has maxima at 0 and 1. Let the distribution that puts probability $\frac{1}{2}$ on 0 and on 1 be ζ_1. Then $\xi_1(\lambda, \cdot)$ puts probabilities $\frac{1}{2}\lambda$, $1 - \lambda$, $\frac{1}{2}\lambda$ on 0, $\frac{1}{2}$, 1, respectively, $0 \leqq \lambda \leqq 1$. The corresponding Bayes estimator is

$$\begin{aligned} \delta_1(\lambda, x) &= \tfrac{1}{2}(1 - \lambda) \qquad \text{for } x = 0 \\ &= \tfrac{1}{2}(1 + \lambda) \qquad \text{for } x = 1. \end{aligned}$$

Its risk function is

$$R_1(\lambda, \omega) = [\tfrac{1}{2}(1 + \lambda) - \omega]^2\omega + [\tfrac{1}{2}(1 - \lambda) - \omega]^2(1 - \omega)$$

and Bayes risk is

$$r_1(\lambda) = \int R_1(\lambda, \omega)\, d\xi_1(\lambda, \omega) = \tfrac{1}{4}\lambda(1 - \lambda).$$

This is maximized by $\lambda_1 = \frac{1}{2}$. Then $\xi_2 = \xi_1(\lambda_1, \cdot)$ puts probabilities $\frac{1}{4}$, $\frac{1}{2}$, $\frac{1}{4}$ on on 0, $\frac{1}{2}$, 1, respectively. Also,

$$\begin{aligned} \delta_2(x) = \delta_1(\lambda_1, x) &= \tfrac{1}{4} \qquad \text{for } x = 0 \\ &= \tfrac{3}{4} \qquad \text{for } x = 1, \\ R_2(\omega) = R_1(\lambda_1, \omega) &= \tfrac{1}{16}, \qquad \text{and } r_2 = \tfrac{1}{16}. \end{aligned}$$

By the Corollary of Theorem 4, δ_2 is the minimax estimator. The iterative method converged in one step here because of a fortuitous choice of ξ_1 and because the chosen mixture λ_1 maximized $r_1(\lambda)$.

In the example, an explicit expression for $r_k(\lambda)$ was obtained and a λ satisfying condition (2) of Theorem 8 was found easily. However, for many problems, $r_k(\lambda)$ does not have a manageable explicit expression and a λ satisfying condition (2) must be found by trial and error computation. The Bayes risk must be calculated for different values of λ until a λ is found that satisfies condition (2). The concavity of $r_k(\lambda)$ is used in the obvious way to aid in the search for such a λ. If B is an upper bound for $\max_{0 \leqq \lambda \leqq 1} r_k(\lambda)$ and if $r_k(\lambda')$ is calculated and satisfies

$$(r_k(\lambda') - r_k)/(B - r_k) \geqq \beta,$$

then λ' clearly satisfies condition (2). Lemma 6 provides such an upper bound.

Lemma 6. *Let $r(\lambda)$ be a continuous, concave function for $0 \leqq \lambda \leqq 1$. If $r(b) \geqq r(a)$ and $r(b) \geqq r(c)$ where $0 \leqq a < b < c \leqq 1$, then*

$$\begin{aligned} (11) \quad \max_{0 \leqq \lambda \leqq 1} r(\lambda) \leqq \max \{&[r(b) + [(r(b) - r(a))/(b - a)](c - b)], \\ &[r(b) + [(r(b) - r(c))/(b - c)](a - b)]\}. \end{aligned}$$

In applications, a, b, and c correspond to three mixtures for which the Bayes risks $r(a)$, $r(b)$, and $r(c)$ have been calculated.

Proof. Since $r(\lambda)$ is continuous on the compact set [0, 1], there exists a

λ^* that maximizes $r(\lambda)$. We first show there is such a λ^* satisfying $a \leqq \lambda^* \leqq c$. Suppose $c < \lambda^* \leqq 1$. (The proof for $0 \leqq \lambda^* < a$ is the same.) Clearly $r(\lambda^*) \geqq r(b)$. Assume $r(\lambda^*) > r(b)$, then a weighted average of them gives

$$[(c - b)/(\lambda^* - b)]r(\lambda^*) + [(\lambda^* - c)/(\lambda^* - b)]r(b) > r(b) \geqq r(c).$$

This contradicts the concavity of $r(\lambda)$, so $r(\lambda^*) = r(b)$. Thus we can choose $\lambda^* = b$.

Let λ^* satisfying $a \leqq \lambda^* \leqq c$ maximize $r(\lambda)$. Suppose $b \leqq \lambda^* \leqq c$. (The proof for $a \leqq \lambda^* \leqq b$ is the same.) Assume that (11) is not satisfied. Then

$$r(\lambda^*) > \{r(b) + [(r(b) - r(a))/(b - a)](c - b)\}. \tag{12}$$

Here $c - b \geqq \lambda^* - b$, and, by hypothesis, $(r(b) - r(a))/(b - a) \geqq 0$. So

$$[(r(b) - r(a))/(b - a)](c - b) \geqq [(r(b) - r(a))/(b - a)](\lambda^* - b).$$

Substitute this in (12) to get

$$\begin{aligned} r(\lambda^*) &> r(b) + [(r(b) - r(a))/(b - a)](\lambda^* - b) \\ &= [(\lambda^* - a)/(b - a)]r(b) - [(\lambda^* - b)/(b - a)]r(a) \end{aligned}$$

or, rewritten,

$$[(b - a)/(\lambda^* - a)]r(\lambda^*) + [(\lambda^* - b)/(\lambda^* - a)]r(a) > r(b).$$

This contradicts the concavity of $r(\lambda)$ and completes the proof.

In the example above, Assumption 6 is not satisfied by all of the *a priori* distributions. Given below is an additional assumption, satisfied in the example, that guarantees that the iterative method works there. The assumption guarantees that the *a priori* distributions that do not satisfy Assumption 6 do not appear in the iteratively constructed sequence and so cause no difficulties. This works for binomial, multinomial, and Poisson problems.

ASSUMPTION 7. For any *a priori* distribution ζ in the class Z of *a priori* distributions not satisfying Assumption 6, $r(\zeta) = 0$, whereas $r^* = \sup_\xi r(\xi) > 0$.

The results of Sections 2 and 3 hold if all of the *a priori* distributions appearing in the results are restricted to the class $\Xi - Z$ of *a priori* distributions satisfying Assumption 6. Thus it is necessary to insure that the *a priori* distributions $\{\xi_k\}$ and $\{\xi_k(\lambda, \cdot)\}$ appearing in the iteration are in the class $\Xi - Z$ and that the limit of any weakly convergent subsequence of $\{\xi_k\}$ is in $\Xi - Z$. We assume that ξ_1 is chosen so that $r_1 > 0$. This and the concavity of $r_1(\lambda)$, which does not depend on Assumption 6, imply that $r_1(\lambda) > 0$, which implies $\xi_1(\lambda, \cdot)$ is in $\Xi - Z$ for $0 \leqq \lambda < 1$. In particular, $r_2 = r_1(\lambda_1) > 0$ and $\xi_2(\cdot) = \xi_1(\lambda_1, \cdot)$ is in $\Xi - Z$. Hence by induction, the *a priori* distributions $\{\xi_k\}$ and $\{\xi_k(\lambda, \cdot)\}$ are all in $\Xi - Z$. It remains only to show:

LEMMA 7. *Under Assumption* 7, *if* $\{\xi_i\}$ *is a sequence of* a priori *distributions in* $\Xi - Z$ *such that* $\lim r(\xi_i) > 0$, *then the limit of any weakly convergent subsequence is in* $\Xi - Z$.

PROOF. Let $\{\xi_j\}$ be a weakly convergent subsequence of $\{\xi_i\}$ with limit ξ_0. Assume that ξ_0 is in Z. By Assumption 7, the Bayes risk satisfies $\int R_{\xi_0}(\omega)\, d\xi_0(\omega) = 0$. Since R_{ξ_0} is continuous, $\xi_j \to \xi_0$ implies $\lim \int R_{\xi_0}(\omega)\, d\xi_j(\omega) = \int R_{\xi_0}(\omega)\, d\xi_0(\omega) = 0$. Then

$$0 \leqq \limsup_j \int R_{\xi_j}(\omega)\, d\xi_j(\omega) \leqq \limsup_j \int R_{\xi_0}(\omega)\, d\xi_j(\omega) = 0.$$

This contradicts the hypothesis $\lim r(\xi_i) > 0$.

The following example shows that Lemma 7 need not hold if Assumption 7 is not satisfied. It is on binomial estimation with absolute error loss.

EXAMPLE. Let X be binomial with one observation and parameter ω; that is, $p(x \mid \omega) = \omega^x(1-\omega)^{1-x}$, $x = 0, 1$. Use absolute error loss $L(d, \omega) = |d - \omega|$. Here $\Omega = D = [0, 1]$. General results for this loss function are given in [1], Section 11.2. The sequence $\xi_i(0) = \xi_i(1) = (1 + (-\frac{1}{2})^i)/4$ and $\xi_i(\frac{1}{2}) = (1 - (-\frac{1}{2})^i)/2$, $i = 1, 2, \cdots$, of *a priori* distributions on the points $0, \frac{1}{2}, 1$ has the limit distribution $\zeta(0) = \zeta(1) = \frac{1}{4}$ and $\zeta(\frac{1}{2}) = \frac{1}{2}$. Each ξ_i satisfies Assumption 6, but ζ does not. Also, ζ does not satisfy Assumption 7 since $r(\zeta) = \frac{1}{4} > 0$. Moreover, ζ is least favorable and

$$\begin{aligned} \delta(x) &= \tfrac{1}{4}, & x &= 0 \\ &= \tfrac{3}{4}, & x &= 1 \end{aligned}$$

is Bayes (ζ), though not unique, and minimax. The unique Bayes(ξ_i) estimator is

$$\begin{aligned} \delta_i(0) = 1 - \delta_i(1) &= 0 \quad && \text{for } i \text{ even} \\ &= \tfrac{1}{2} && \text{for } i \text{ odd.} \end{aligned}$$

The corresponding risk function and Bayes risk are

$$\begin{aligned} R_i(\omega) &= 2\omega(1-\omega) \quad && \text{for } i \text{ even} \\ &= |\omega - \tfrac{1}{2}| && \text{for } i \text{ odd,} \\ r(\xi_i) &= (1 - (\tfrac{1}{2})^i)/4. \end{aligned}$$

Thus $\xi_i \to \zeta$ in Z and $\lim r(\xi_i) = r(\zeta) = \frac{1}{4} > 0$, but $\lim \delta_i$ and $\lim R_i$ do not exist. Thus Lemma 7 and results of Section 2 need not hold when Assumption 7 is not satisfied.

The iterative method of Section 3 has been extended to obtain minimax truncated sequential procedures for certain sequential decision problems with a finite sequence of independent identically distributed chance variables and constant cost per observation.

Acknowledgment. The author expresses his sincere appreciation for the guidance of Professor Colin R. Blyth, who helped formulate and initiate this work, and of Professor Robert A. Wijsman, who carefully directed the major portion of it. The author takes great pleasure in acknowledging their considerable contribution to this work. The author thanks the referee for kindly suggesting the

present version of condition (1) of Theorem 8 which is more satisfying than the author's.

REFERENCES

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

[2] Ghosh, M. N. (1964). Uniform approximation of minimax point estimates. *Ann. Math. Statist.* **35** 1031–1047.

[3] Lehmann, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

[4] Royden, H. L. (1963). *Real Analysis.* Macmillan, New York.

[5] Wald, A. (1945). Statistical decision functions which minimize the maximum risk. *Ann. Math.* **16** 265–280.

[6] Wald, A. (1950). *Statistical Decision Functions.* Wiley, New York.