The Annals of Mathematical Statistics
1970, Vol. 41, No. 2, 609–620

ESTIMATION FOR MONOTONE PARAMETER SEQUENCES: THE DISCRETE CASE[1]

BY HAROLD SACKROWITZ

Rutgers, The State University

1. Introduction and summary. Let $x_1, x_2, \cdots$ be a sequence of independent, observable, discrete, real random variables with respective density functions $f(x, \theta_1), f(x, \theta_2), \cdots$, where $\theta_i \in \Omega$ (a real interval), $i = 1, 2, \cdots$. At any stage s, having observed $x_1, \cdots, x_s$, we wish to estimate θ_s, relative to the loss function $L(d, \theta) \geqq 0$, under the restraint that $\theta_1 \leqq \cdots \leqq \theta_s$.

Most of the literature concerned with estimating ordered parameters deals with maximum likelihood estimators (for example, Robertson and Waltman [3] and references contained therein). Recently, Blumenthal and Cohen [1] considered questions of minimaxity and admissibility of certain estimators of ordered translation parameters of continuous distributions.

In Section 3 we obtain results concerning admissible and minimax estimators of θ_s for a large class of discrete distributions. Theorem 3.2 states that the minimax value M_k of the problem of estimating θ_k at stage k is the same for all $k = 1, 2, \cdots$. Furthermore, the Corollary to Theorem 3.1 states that if $t(x_1, \cdots, x_s)$ is admissible for estimating θ_s at stage s, then $t(x_{k+1}, \cdots, x_{k+s})$ is admissible for estimating θ_{k+s} at stage $k+s$. This is not true, for example, when x_k is normally distributed with mean θ_k and unit variance [4]. Hence in situations satisfying the conditions of Theorems 3.1 and 3.2, if there exists an admissible estimator $t(x_1)$ of θ_1 having constant risk, then $t(x_s)$ will be the unique admissible minimax estimator of θ_s, at stage s. It is this undesirable property (i.e. being based only on the last observation) which prompts us to look for other estimators which may be "better" in some sense.

In Section 4 we use the early observations ($x_1, \cdots, x_{s-1}$, at stage s) to construct a sequence of estimators which is asymptotically subminimax ($s \to \infty$) as well as having desirable properties for finite s. The results in Section 4 are obtained for general cumulative distribution functions as the methods are not peculiar to the discrete situation. However, it should be noted that the main results of Section 4 (Theorem 4.2 and Theorem 4.3) do not yield a worthwhile sequence of estimators unless we are in a situation similar to that of Section 3. That is, unless the existing admissible minimax estimators are undesirable as they are in the discrete case of Section 3.

2. Notation. We observe the discrete, real random variables $x_1, \cdots, x_s$, having respective density functions $f(x, \theta_1), \cdots, f(x, \theta_s)$, where $\theta_1 \leqq \cdots \leqq \theta_s$, $\theta_k \in \Omega$ (a real interval), $k = 1, \cdots, s$. We wish to estimate θ_s relative to the loss function

---

Received December 9, 1969; revised October 27, 1969.

[1] Research supported in part by the Office of Naval Research under Contract No. Nonr-4259(08) while the author was at Columbia University.

$0 \leqq L(d, \theta) < \infty$, where $L(d, \theta_s)$ represents the loss in taking action d when, in fact, the true parameter value is θ_s.

We define the following sets of estimators. Let $\mathscr{A}_k{}^s = \{t(\cdot, \cdots, \cdot): t(x_k, \cdots, x_s)$ is admissible at stage $s\}$, $k = 1, \cdots, s$. Furthermore, we denote by $R(t(x_k, \cdots, x_s); \theta_1, \cdots, \theta_s)$ the risk in estimating θ_s (at stage s) with $t(x_k, \cdots, x_s)$ when $\theta_1, \cdots, \theta_s$ are the true parameter values (i.e., x_i has density $f(x, \theta_i)$).

3. Admissibility and minimax. Throughout this section we will be working under the following assumptions.

A1. Ω is a real interval with lower limit $\gamma > -\infty$ (possibly $\gamma \notin \Omega$).

A2. $A = \{a_k : f(a_k, \theta) > 0$ for some $\theta \in \Omega\}$, that is, $a_1, a_2, \cdots, a_v$ are the atoms of $\{f(x, \theta): \theta \in \Omega\}$, where $v = \text{card}(A) \leqq \aleph_0$ (the cardinality of the set of integers).

A3. There exists an ordering, a_{i_j}, of the atoms such that

$$f^{-1}(a_{i_n}, \theta) \sum_{j=n+1}^{v} f(a_{i_j}, \theta) \to_{\theta \to \gamma} 0, \qquad n = 1, 2, \cdots, v-1.$$

A4. Given any $\theta \in \Omega$, either

(i) $f(a, \theta) > 0$ for all $a \in A$, or
(ii) $f(a^*, \theta) = 1$ for some $a^* \in A$.

A5. If $f(a^*, \theta^*) = 1$ for some $a^* \in A$, $\theta^* \in \Omega$, then θ^* is a boundary point of Ω and the model is such that the risk functions of all admissible estimators are continuous at θ^*.

We note that

$$f^{-1}(a_{i_1}, \theta) \sum_{j=2}^{v} f(a_{i_j}, \theta) = f^{-1}(a_{i_1}, \theta)[1 - f(a_{i_1}, \theta)] = f^{-1}(a_{i_1}, \theta) - 1,$$

and

$$0 \leqq f^{-1}(a_{i_n}, \theta) f(a_{i_{n+1}}, \theta) \leqq f^{-1}(a_{i_n}, \theta) \sum_{j=n+1}^{v} f(a_{i_j}, \theta).$$

Therefore A3 implies that $f(a_{i_1}, \theta) \to_{\theta \to \gamma} 1$, and $f^{-1}(a_{i_n}, \theta) f(a_{i_{n+1}}, \theta) \to_{\theta \to \gamma} 0$, $n = 1, 2, \cdots, v-1$.

We now give two examples of estimation problems satisfying assumptions A1 through A5.

EXAMPLE 3.1. (The Binomial distribution). Suppose $L(d, \theta) = (d - \theta)^2$ and $f(x, \theta) = \binom{m}{x} \theta^x (1-\theta)^{m-x}$, $x = 0, \cdots, m$. Then $\Omega = [0, 1]$, $\gamma = 0$, $A = \{0, \cdots, m\}$, $a_{i_j} = j - 1$, $v = m + 1$, and so A1 and A2 hold. A3 holds since

$$\begin{aligned} f^{-1}(n-1, \theta) \sum_{k=n+1}^{m+1} f(k-1, \theta) &= \left[\binom{m}{n-1} \theta^{n-1} (1-\theta)^{m-n+1}\right]^{-1} \\ &\quad \cdot \sum_{k=n+1}^{m+1} \binom{m}{k-1} \theta^{k-1} (1-\theta)^{m-k+1} \\ &= \binom{m}{n-1}^{-1} \sum_{k=n+1}^{m+1} \binom{m}{k-1} \theta^{k-n} (1-\theta)^{n-k} \to_{\theta \to 0} 0 \end{aligned}$$

as $k-n \geqq 1$ for $k = n+1, \cdots, m$. Furthermore, A4 and A5 are satisfied for $L(d, \theta) = (d-\theta)^2$ as

$$R(t(x_1, \cdots, x_s); \theta_1, \cdots, \theta_s) = \sum_{x_1=0}^{m} \cdots \sum_{x_s=0}^{m} (t(x_1, \cdots, x_s)-\theta_s)^2 \cdot \prod_{i=1}^{s} \binom{m}{x_i} \theta_i^{x_i}(1-\theta_i)^{m-x_i}$$

is continuous at any $(\theta_1, \cdots, \theta_s)$ since t admissible implies $(t(x_1, \cdots, x_s)-\theta_s)^2 \leqq 1$.

EXAMPLE 3.2. (The Poisson distribution). Suppose $f(x, \theta) = e^{-\theta}\theta^x/x!$, $\theta > 0$, $x = 0, 1, \cdots$. Then $\Omega = (0, \infty)$, $\gamma = 0$, $a_{i_j} = j-1$, $A = \{0, 1, \cdots\}$, $v = \aleph_0$ and A1 and A2 hold. To show A3 holds it suffices to show $f(n, \theta)^{-1} \sum_{k=n+1}^{\infty} f(k, \theta) \to 0$ as $\theta \to 0$ through values less than 1. This is so, as

$$[e^{-\theta}\theta^n/n!]^{-1} \sum_{k=n+1}^{\infty} e^{-\theta}\theta^k/k! = \sum_{k=n+1}^{\infty} \theta^{k-n} n!/k! \leqq n! \sum_{k=n+1}^{\infty} \theta^{k-n} = n!\theta/(1-\theta)$$

for $0 < \theta < 1$, and $\theta/(1-\theta) \to_{\theta \to 0} 0$. A4 and A5 hold for all loss functions as $f(k, \theta) > 0$ for all $\theta \in \Omega$ and $k = 0, 1, \cdots$.

Before stating Theorem 3.1 we mention the following.

$$R(t(x_k, \cdots, x_s); \theta_1, \cdots, \theta_s) = R(t(x_{k+r}, \cdots, x_{s+r}); \theta_1, \cdots, \theta_{s+r})$$

if $\theta_k = \theta_{k+r}$. The essence of this remark is that any estimator used at stage s (to estimate θ_s) can also be used at any succeeding stage $s+r$ (to estimate θ_{s+r}) and the risk functions at these stages will be essentially the same. We are now ready to state

THEOREM 3.1. *Under assumptions* A1–A5, $\mathscr{A}_k{}^s = \mathscr{A}_{k+r}^{s+r}$ *for* $k = 1, \cdots, s$, $s = 1, 2, \cdots, r = 0, 1, \cdots$.

PROOF. Without loss of generality we may assume that $\gamma = 0$ and $a_{i_j} = j-1$, $j = 1, 2, \cdots, v$ (i.e., the set A is composed of consecutive integers starting with 0).

Let $t \in \mathscr{A}_{k+r}^{s+r}$. If $t \notin \mathscr{A}_k{}^s$, then there exists a function $t^* \in \mathscr{A}_1{}^s$ which dominates $t(x_k, \cdots, x_s)$ as an estimator of θ_s. Since $t^*(x_{1+r}, \cdots, x_{s+r})$ can be used as an estimator of θ_{s+r}, it dominates $t(x_{k+r}, \cdots, x_{s+r})$ as an estimator of θ_{s+r}. This is a contradiction as $t \in \mathscr{A}_{k+r}^{s+r}$. Hence $\mathscr{A}_{k+r}^{s+r} \subseteq \mathscr{A}_k{}^s$.

To complete the proof it suffices to show that $\mathscr{A}_k{}^s \subseteq \mathscr{A}_{k+1}^{s+1}$. We shall exhibit the proof for $k = 1, s = 1$ only, as it can be readily extended to the case of arbitrary k and s.

Let $\phi(\cdot) \in \mathscr{A}_1{}^1$. If $\phi(\cdot) \notin \mathscr{A}_2{}^2$, then there exists an estimator $\delta(\cdot, \cdot) \in \mathscr{A}_1{}^2$ which dominates $\phi(x_2)$ as an estimator of θ_2. If δ is a non-randomized rule, then $\delta(x_1, x_2)$ is simply a real number for each pair (x_1, x_2). If δ is a randomized rule, then for every fixed pair (x_1, x_2), $\delta(x_1, x_2)$ defines a probability distribution $\Delta(a; x_1, x_2)$, $a \in (-\infty, \infty)$, on the real line. Define

$$L_\delta(x_1, x_2; \theta_2) = L(\delta(x_1, x_2), \theta_2) \quad \text{if} \quad \delta \quad \text{is non-randomized}$$

$$= \int L(a, \theta_s)\, d\Delta(a; x_1, x_2) \quad \text{if} \quad \delta \quad \text{is randomized.}$$

Since δ dominates ϕ, it follows that

$$R(\phi(x_2);\theta_1,\theta_2)\geqq R(\delta(x_1,x_2);\theta_1,\theta_2)$$

$$(1)\qquad = E_{\theta_1\theta_2}\{L_\delta(x_1,x_2;\theta_2)\} = E_{\theta_1}\{E_{\theta_2}[L_\delta(x_1,x_2;\theta_2)\,|\,x_1]\}$$

$$= E_{\theta_2}\{L_\delta(0,x_2;\theta_2)\}f(0,\theta_1)+\textstyle\sum_{i=1}^{\infty}E_{\theta_2}\{L_\delta(i,x_2;\theta_2)\}f(i,\theta_1),$$

for all $\theta_1, \theta_2 \in \Omega$, $\theta_1 \leqq \theta_2$.

Fix $\theta_2 = \theta^* \in \operatorname{int}\Omega$ (interior of Ω). Then (1) implies

$$(2)\qquad R(\phi(x_2);\theta_1,\theta^*)f^{-1}(0,\theta_1)\geqq E_\theta{}^*\{L_\delta(0,x_2;\theta^*)\} + \textstyle\sum_{i=1}^{\infty}E_{\theta^*}\{L_\delta(i,x_2;\theta^*)\}f(i,\theta_1)f^{-1}(0,\theta_1),$$

for all $\theta_1 \leqq \theta^*$, $\theta_1 \in \operatorname{int}\Omega$. Since $\lim_{\theta_1\to 0}f(0,\theta_1)=1$, for θ_1 sufficiently close to 0, (2) implies

$$(3)\qquad R(\phi(x_2);\theta_1,\theta^*)(1+\varepsilon)\geqq \textstyle\sum_{i=1}^{\infty}E_{\theta^*}\{L_\delta(i,x_2;\theta^*)\}f(i,\theta_1)f^{-1}(0,\theta_1).$$

Suppose $R(\phi(x_2);\theta_1,\theta^*)<\infty$, then there exists an N such that

$$(4)\qquad \textstyle\sum_{i=N+1}^{\infty}E_{\theta^*}\{L_\delta(i,x_2;\theta^*)\}f(i,\theta_1)f^{-1}(0,\theta_1)\leqq\varepsilon.$$

Since $f(i,\theta_1)f^{-1}(0,\theta_1)\to 0$ as $\theta_1\to 0$ (for $i>0$), it follows from (2), (3), and (4) that

$$(5)\qquad R(\phi(x_2);\theta_1,\theta^*)\geqq E_{\theta^*}\{L_\delta(0,x_2;\theta^*)\}.$$

If $R(\phi(x_2);\theta_1,\theta^*)=\infty$, then (5) is obviously true. Hence we have shown that, for any $\theta_2\in\operatorname{int}\Omega$,

$$(6)\qquad R(\phi(x_2);\theta_1,\theta_2)\geqq E_{\theta_2}\{L_\delta(0,x_2;\theta_2)\} = R(\delta(0,x_2);\theta_1,\theta_2).$$

By assumptions A4 and A5, (6) also holds for boundary points of Ω. Equivalently,

$$(7)\qquad R(\phi(x_1);\theta_1)\geqq R(\delta'(x_1);\theta_1),\qquad \text{all}\quad \theta_1\in\Omega,$$

where $\delta'(\cdot)=\delta(0,\cdot)$.

Since $\phi(\cdot)\in\mathscr{A}_1{}^1$, it follows that $\delta'(\cdot)\in\mathscr{A}_1{}^1$ and $R(\phi(x_1);\theta_1)\equiv R(\delta'(x_1);\theta_1)$. That is,

$$(8)\qquad R(\phi(x_2);\theta_1,\theta_2)=E_{\theta_2}\{L_\delta(0,x_2;\theta_2)\},\qquad \text{for all}\quad \theta_2\in\Omega.$$

Using (8) in (1), we obtain

$$(9)\qquad R(\phi(x_2);\theta_1,\theta_2)[f(1,\theta_1)+\textstyle\sum_{i=2}^{\infty}f(i,\theta_1)] = R(\phi(x_2);\theta_1,\theta_2)[1-f(0,\theta_1)]$$

$$\geqq E_{\theta_2}\{L_\delta(1,x_2;\theta_2)\}f(1,\theta_1)+\textstyle\sum_{i=2}^{\infty}E_{\theta_2}\{L_\delta(i,x_2;\theta_2)\}f(i,\theta_1).$$

Fix $\theta_2=\theta^*\in\operatorname{int}\Omega$ and divide all terms in (9) by $f(1,\theta_1)$, where $\theta_1\leqq\theta^*$, $\theta_1\in\operatorname{int}\Omega$. Then we have

$$(10)\qquad R(\phi(x_2);\theta_1,\theta^*)[1+f^{-1}(1,\theta_1)\textstyle\sum_{i=1}^{\infty}f(i,\theta_1)]\geqq E_{\theta^*}\{L_\delta(1,x_2;\theta^*)\} + \textstyle\sum_{k=1}^{\infty}E_{\theta^*}\{L_\delta(i,x_2;\theta^*)\}f(i,\theta_1)f^{-1}(1,\theta_1).$$

Comparing (10) to (2) we see that we can use the arguments used to show (8) to show $R(\phi(x_2); \theta_1, \theta_2) = E_{\theta_2}\{L_\delta(1, x_2; \theta_2)\}$, for all $\theta_2 \in \Omega$. Continuing in this manner we find that, for $i = 0, 1, \cdots$ $R(\phi(x_2); \theta_1, \theta_2) = E_{\theta_2}\{L_\delta(i, x_2; \theta_2)\}$, for all $\theta_2 \in \Omega$. Therefore $R(\phi(x_2); \theta_1, \theta_2) = R(\delta(x_1, x_2); \theta_1 \theta_2)$, for all $\theta_1, \theta_2 \in \Omega$, $\theta_1 \leqq \theta_2$. Hence $\phi(\cdot) \in \mathscr{A}_2{}^2$ and the proof is complete.

COROLLARY. *Under the assumptions* A1–A5, $\mathscr{A}_j{}^r \subseteq \mathscr{A}_k{}^s$, $r \leqq s$, $j = k, \cdots, r$, $s = 1, 2, \cdots$.

PROOF. It is clear from the definition of $\mathscr{A}_k{}^s$ that $\mathscr{A}_i{}^s \subseteq \mathscr{A}_k{}^s$, $i = k, \cdots, s$. The result now follows from Theorem 3.1 as $\mathscr{A}_j{}^r = \mathscr{A}^s_{j+s-r} \subseteq \mathscr{A}_k{}^s$.

THEOREM 3.2. *Define M_k, $k = 1, 2, \cdots$ to be the minimax value for the problem of estimating θ_k at stage k. Then under assumptions* A1–A5, $M_1 = M_2 = \cdots$.

PROOF. Since any estimator used at stage k can also be used at stage $k+1$, it follows that $M_1 \geqq M_2 \geqq \cdots$. We will show $M_1 = M_2$ only as the proof can easily be extended to show $M_k = M_{k+1}$.

Suppose $M_1 > M_2$. Then there exists an $\varepsilon > 0$ and a function $t(x_1, x_2)$ such that $M_1 - \varepsilon \geqq R(t(x_1, x_2); \theta_1, \theta_2)$, $\theta_1 \leqq \theta_2$. This implies, by the same argument used in Theorem 1 to obtain (7), that $M_1 - \varepsilon \geqq R(t'(x_1); \theta_1)$, all $\theta_1 \in \Omega$, where $t'(\cdot) = t(0, \cdot)$. Since M_1 is the minimax value at stage 1, we have a contradiction and the proof is complete.

EXAMPLE 3.3. (Bernoulli trials). Let $L(d, \theta) = (d-\theta)^2$, $f(x, \theta) = \theta^x(1-\theta)^{1-x}$, $x = 0, 1$. At stage s we wish to estimate θ_s. Even if $x_1 = \cdots = x_s = 1$, the (unique) admissible minimax estimator says to estimate θ_s with $\frac{1}{4} + \frac{1}{2}x_s$.

EXAMPLE 3.4. (Uniform distribution). Let $f(x, \theta) = [\theta]^{-1}$, $x = 1, 2, \cdots, [\theta]$, where $[\theta]$ is the greatest integer in θ and $\theta \in \Omega = [1, \infty)$. This distribution does not satisfy A3. Furthermore, the results of Theorems 3.1 and 3.2 will not hold for any reasonable loss function. To see this we note that observing $x_k = x_k{}^*$ is positive proof that $\theta_k \geqq x_k{}^*$. Hence the estimator $t'(x_1, \cdots, x_s) = \max(x_1, \cdots, x_{s-1}, t(x_s))$ would dominate the estimator $t(x_s)$ for estimating θ_s.

4. Asymptotic solutions. The results of Section 3 indicate that, for the situation under consideration, use of admissible minimax estimators will often lead to somewhat undesirable procedures. We have seen that, in the case of Bernoulli trials with squared error loss, the unique admissible minimax estimator of θ_s is based on x_s only.

In this section we will construct a sequence of estimators which are asymptotically subminimax, have reasonable and determinable (see discussion of Section 5) small sample properties, and are intuitively more appealing than the admissible minimax estimators tend to be.

We are motivated by the realization that knowing the values $x_1, \cdots, x_{s-1}$ at stage s cannot be more advantageous than knowing θ_{s-1} (i.e. $x_1, \cdots, x_{s-1}$ probably cannot do more than determine some sort of lower bound for θ_s) and if we knew θ_{s-1} and then observed x_s, it would be reasonable to use an admissible minimax estimator for $\theta_s \in \Omega \cap [\theta_{s-1}, \infty)$.

For simplicity we construct these sequences for squared error loss. However, similar results can be obtained for other loss functions $L(d, \theta)$ which are increasing functions of $|d-\theta|$. The idea is to use the fact that $(s-1)^{-1}\sum_{i=1}^{s-1} x_i$ becomes a fairly reliable lower bound for θ_s as s increases.

THEOREM 4.1. *Let x have distribution function $F(x, \theta)$, $\theta \in \Omega$ (a real interval). Let $t(x; \gamma)$ be an admissible minimax estimator of $\theta \in \Omega \cap [\gamma, \infty)$ relative to the loss function $L(d, \theta)$. If x_k has distribution $F(x, \theta_k)$, $k = 1, 2, \cdots$ and the x_k are independent, then there does* not *exist, for any $s = 1, 2, \cdots$, an estimator $\delta(x_1, \cdots, x_s)$ such that*

$$R(\delta(x_1, \cdots, x_s); \theta_1^*, \cdots, \theta_{s-1}^*, \theta_s) \leqq R(t(x_s; \theta_{s-1}); \theta_1^*, \cdots, \theta_{s-1}^*, \theta_s),$$

for any fixed $\theta_1^* \leqq \cdots \leqq \theta_{s-1}^*$ and all $\theta_s \geqq \theta_{s-1}^*$, $\theta_s, \theta_k^* \in \Omega$, $k = 1, \cdots, s$, with strict inequality for at least one point $(\theta_1^*, \cdots, \theta_s^*)$.

PROOF. If $\theta_1, \cdots, \theta_{s-1}, x_s$ are known, then $t(x_s; \theta_{s-1})$ is an admissible minimax estimator for $\theta_s \in \Omega$ and $\theta_1 \leqq \cdots \leqq \theta_s$. (Easily seen from the definition of $t(x; \gamma)$.)

Define $L_\delta(x_1, \cdots, x_s; \theta_s)$ as in the proof of Theorem 3.1. Then the risk function of any estimator δ is $E\{L_\delta(x_1, \cdots, x_s; \theta_s)\}$.

For every estimator $\delta(x_1, \cdots, x_s)$ and every fixed set $\theta_1, \cdots, \theta_{s-1}$, define the estimator $\Psi_\delta(\theta_1, \cdots, \theta_{s-1}; x_s)$ as follows: generate values $x_1, \cdots, x_{s-1}$ and use the estimator $\delta(x_1, \cdots, x_s)$ where x_k has distribution $F(x, \theta_k)$, $k = 1, \cdots, s-1$ (i.e. Ψ_δ is a randomized rule). Hence

$$\begin{aligned} R(\delta(x_1, \cdots, x_s); \theta_1, \cdots, \theta_s) &= E_{(\theta_1, \cdots, \theta_s)}\{L_\delta(x_1, \cdots, x_s; \theta_s)\} \\ &= E_{\theta_s}\{E_{(\theta_1, \cdots, \theta_{s-1})}\{L_\delta(x_1, \cdots, x_s; \theta_s)\}\} \\ &= R(\Psi_\delta(\theta_1, \cdots, \theta_{s-1}; x_s); \theta_1, \cdots, \theta_s). \end{aligned}$$

Since $t(x_s; \theta_{s-1})$ is admissible minimax among estimators based on $\theta_1, \cdots, \theta_{s-1}$, x_s, the theorem is proved.

Theorem 4.1 reinforces our intuitive suspicion that, even using $x_1, \cdots, x_{s-1}$ at stage s, we could not do better than if we knew θ_{s-1} and used $t(x_s; \theta_{s-1})$.

NOTE. In the remainder of this paper we will assume loss to be squared error, that is $L(d, \theta) = (d-\theta)^2$.

DEFINITION 1. Let x_k have distribution function $F(x, \theta_k)$, where $\theta_k \in \Omega$ (a real interval), $k = 1, 2, \cdots$ and $\theta_1 \leqq \theta_2 \leqq \cdots$. If there exists a sequence of functions $\{m_k(x_1, \cdots, x_k)\}$ such that for every $\varepsilon > 0$, there exists a constant $s(\varepsilon)$ such that for $s \geqq s(\varepsilon)$,

$$E_{(\theta_1 \cdots \theta_s)}[(m_s(x_1, \cdots, x_s) - \theta_s)^+]^2 \leqq \varepsilon,$$

for all $\theta_1 \leqq \cdots \leqq \theta_s$, $\theta_k \in \Omega$, $k = 1, \cdots, s$, then $F(x, \theta)$ is said to have the *lower bound property* and $\{m_k\}$ will be called a *lower bound sequence.*

Clearly lower bound sequences for a particular $F(x, \theta)$ are not unique.

DEFINITION 2. A family of distribution functions $\{F(x, \theta_k): \theta \in \Omega\}$ will be said to have *Property* 1 if there exist sequences of functions $\{u_n(x_1, \cdots, x_n)\}$ and $\{\mu_n(\theta_1, \cdots, \theta_n)\}$, where $\theta_1 \leqq \cdots \leqq \theta_n$ implies $\mu_n \leqq \theta_n$ and $\mu_n(\theta, \cdots, \theta) = \theta$, such that given $\varepsilon > 0$, there exists a constant $N(\varepsilon)$ such that $n \geqq N(\varepsilon)$ implies

$$E_{(\theta_1, \cdots, \theta_n)}[u_n(x_1, \cdots, x_n) - \mu_n(\theta_1, \cdots, \theta_n)]^2 \leqq \varepsilon,$$

where $x_1, x_2, \cdots$ are independent and x_k has distribution $F(x, \theta_k)$, $\theta_k \in \Omega$, $k = 1, 2, \cdots$.

The family of Normal distributions with mean θ and variance 1 is an example of a family having Property 1. This can be seen by setting

$$u_n = n^{-1} \textstyle\sum_{i=1}^n x_i \quad \text{and} \quad \mu_n = n^{-1} \sum_{i=1}^n \theta_i.$$

Clearly, the u_n and μ_n sequences need not be unique for any family having Property 1.

LEMMA 4.1. *If* $\{F(x, \theta): \theta \in \Omega\}$ *has Property* 1, *then* $F(x, \theta)$ *has the lower bound property. Furthermore, the sequence* $\{u_n(x_1, \cdots, x_n)\}$ *whose existence is guaranteed by Property* 1 *is a lower bound sequence.*

PROOF. Let $\{\mu_n(\theta_1, \cdots, \theta_n)\}$ be a sequence corresponding to the u_n sequence (by Property 1). Then as $\mu_n \leqq \theta_n$,

$$E_{(\theta_1, \cdots, \theta_s)}[(u_s - \theta_s)^+]^2 \leqq E_{(\theta_1, \cdots, \theta_s)}[(u_s - \mu_s)^+]^2 \leqq E_{(\theta_1, \cdots, \theta_s)}[u_s - \mu_s]^2.$$

By Property 1, there exists a constant $s(\varepsilon)$ (given an $\varepsilon > 0$) such that for $s \geqq s(\varepsilon)$, $E_{(\theta_1, \cdots, \theta_s)}[u_s - \mu_s]^2 \leqq \varepsilon$. This completes the proof.

DEFINITION 3. Let x be a random variable with distribution function belonging to the family $\{F(x, \theta): \theta \in \Omega\}$ and let $\{h(x; \gamma): \gamma \in \Omega\}$ be a collection of measurable functions of x. Then $h(x; \gamma)$ will be said to have *Property* 2 relative to $\{F(x, \theta): \theta \in \Omega\}$ if

(A) $h(x; \gamma) \leqq \max(\rho(x), \gamma) + \omega$, for all $\gamma \in \Omega$, where $\omega = \omega(\Omega)$ is a constant independent of x and γ, and $\rho(x) = \rho_\Omega(x)$ is a measurable function of x (independent of γ) such that $E_\theta(\rho(x) - \theta)^2 \leqq B = B(\Omega) < \infty$, for all $\theta \in \Omega$, and

(B) $h(x; \gamma)$ is equicontinuous in x as a function of γ.

EXAMPLE 4.1. Let $\{F(x, \theta): \theta \in \Omega\}$ be any family of distributions for which $E_\theta(x - \theta)^2 \leqq B < \infty$ for all $\theta \in \Omega$. Then the function $h(x; \gamma) = \max(x, \gamma)$ has Property 2 relative to $\{F(x, \theta): \theta \in \Omega\}$ (let $\rho(x) = x$, $\omega = 0$).

DEFINITION. The family of distributions $\{F(x, \theta): \theta \in \Omega\}$ will be said to have *Property* 3 if for any real γ there exists an admissible minimax estimator, $t(x; \gamma)$, for $\theta \in \Omega \cap [\gamma, \infty)$, relative to squared error loss, such that $t(x; \gamma)$ has Property 2 relative to $\{F(x, \theta): \theta \in \Omega\}$.

REMARK. If $t(x; \gamma)$ is an admissible estimator of $\theta \in \Omega \cap [\gamma, \infty)$, relative to squared error loss, then $t(x; \gamma) \geqq \gamma$ with probability one relative to $\Omega \cap [\gamma, \infty)$. If not, the estimator $\max(t(x, \gamma), \gamma)$ dominates $t(x; \gamma)$.

LEMMA 4.2. *Let the family $\{F(x, \theta): \theta \in \Omega\}$ have Property 3 and the lower bound property with lower bound sequence $\{m_k(x_1, \cdots, x_k)\}$. If there exists an estimator, $\delta(x_s)$, of $\theta_s \in \Omega$ which has a risk function bounded by v_0, then for any $\varepsilon > 0$, there exists a constant $s(\varepsilon)$ such that for $s \geqq s(\varepsilon)$*

$$R(t(x_s; m_{s-1}(x_1, \cdots, x_{s-1})); \theta_1, \cdots, \theta_s) \leqq v_0 + \varepsilon,$$

for all $\theta_1 \leqq \cdots \leqq \theta_s$ and t defined by Property 3.

PROOF. The estimator $\delta(x_s)$ of $\theta_s \in \Omega$ may, of course, also be used as an estimator of $\theta_s \in \Omega \cap [\gamma, \infty)$ for any γ. Since $t(x_s; \gamma)$ is defined to be admissible minimax for $\theta_s \in \Omega \cap [\gamma, \infty)$,

$$E_{\theta_s}[t(x_s; \gamma) - \theta_s]^2 \leqq v_0, \gamma \leqq \theta_s. \tag{11}$$

$$\begin{aligned} R(t(x_s; m_{s-1}); \theta_1, \cdots, \theta_s) &= E_{(\theta_1, \cdots, \theta_s)}[t(x_s; m_{s-1}) - \theta_s]^2 \\ &= E_{(\theta_1, \cdots, \theta_{s-1})}\{E_{\theta_s}\{[t(x_s; m_{s-1}) - \theta_s]^2 \mid m_{s-1}\}\} \\ &= E_{(\theta_1, \cdots, \theta_{s-1})}\{E_{\theta_s}[t(x_s; m_{s-1}) - \theta_s]^2\}. \end{aligned}$$

This allows us to consider $E_{\theta_s}[t(x_s; m_{s-1}) - \theta_s]^2$ for fixed values of m_{s-1} and then treat the risk of $t(x_s; m_{s-1})$ as the expected value of a function of m_{s-1}.

Given any $\alpha > 0$, there exists (by (B) of Property 2) a $\beta(\alpha)$ such that if $\theta_s < m_{s-1} \leqq \theta_s + \beta(\alpha)$, then $[t(x_s; m_{s-1}) - t(x_s; \theta_s)]^2 \leqq \alpha$.

Fix $\alpha = \min(\varepsilon/4, \varepsilon^2/64v_0)$.

Case 1. If $m_{s-1} \leqq \theta_s$, then by (11)

$$E_{\theta_s}[t(x_s; m_{s-1}) - \theta_s]^2 \leqq v_0.$$

Case 2. If $\theta_s < m_{s-1} \leqq \theta_s + \beta(\alpha)$, then

$$\begin{aligned} E_{\theta_s}[t(x_s; m_{s-1}) - \theta_s]^2 &= E_{\theta_s}[t(x_s; m_{s-1}) - t(x_s; \theta_s) + t(x_s; \theta_s) - \theta_s]^2 \\ &\leqq E_{\theta_s}[t(x_s; m_{s-1}) - t(x_s; \theta_s)]^2 \\ &\qquad + 2v_0^{\frac{1}{2}} E_{\theta_s}^{\frac{1}{2}}[t(x_s; m_{s-1}) - t(x_s; \theta_s)]^2 + v_0, \end{aligned}$$

by the Schwarz inequality. But by our choice of α,

$$E_{\theta_s}[t(x_s; m_{s-1}) - \theta_s]^2 \leqq \tfrac{1}{4}\varepsilon + \tfrac{1}{4}\varepsilon + v_0 = v_0 + \tfrac{1}{2}\varepsilon.$$

Case 3. If $\theta_s + \beta(\alpha) < m_{s-1}$, then $t(x_s; m_{s-1}) \geqq \theta_s$. By (A) of Property 2. $t(x_s; m_{s-1}) \leqq \max(\rho(x_s), m_{s-1}) + \omega$. Therefore if $\theta_s + \beta(\alpha) < m_{s-1}$, then

$$[t(x_s; m_{s-1}) - \theta_s]^2 \leqq [\rho(x_s) + \omega - \theta_s]^2 + [m_{s-1} + \omega - \theta_s]^2.$$

Thus (for $B = B(\Omega)$ defined by (A) of Property 2)

$$E_{\theta_s}[t(x_s;\, m_{s-1})-\theta_s]^2 \leqq B+2\omega B^{\frac{1}{2}}+\omega^2+[m_{s-1}+\omega-\theta_s]^2.$$

Therefore, by the results of Cases 1, 2, and 3,

$$\begin{aligned} &E_{(\theta_1,\cdots,\theta_{s-1})}\{E_{\theta_s}[t(x_s;\, m_{s-1})-\theta_s]^2\} \\ (12)\qquad &\leqq v_0 P(m_{s-1} \leqq \theta_s+\beta(\alpha))+\varepsilon/2 \\ &\quad +(B+2\omega B^{\frac{1}{2}}+\omega^2)P(m_{s-1} > \theta_s+\beta(\alpha))+E[(m_{s-1}-\theta_s)^+]^2 \\ &\quad +2\omega E^{\frac{1}{2}}[(m_{s-1}-\theta_s)^+]^2+\omega^2 P(m_{s-1} > \theta_s+\beta(\alpha)). \end{aligned}$$

However, once α is chosen (and fixed) we can find, by the lower bound property, a constant $s(\alpha, \varepsilon)$ such that for $s \geqq s(\alpha, \varepsilon)$, each of the last four terms of (12) is less than $\varepsilon/8$. Since α is a function of ε, the lemma is proved.

We recall that by Lemma 4.1, a sequence $\{u_n(x_1, \cdots, x_n)\}$ defined by Property 1 is a lower bound sequence. Hence, in particular, Lemma 4.2 assures us that, if an admissible minimax estimator $\phi(x_s)$ with constant risk exists; after some finite stage $s(\varepsilon)$, the risk of $t(x_s; u_{s-1})$ will not exceed the risk of $\phi(x_s)$ by more than ε anywhere in the parameter space. For $t(x_s; u_{s-1})$ to be considered an improvement on $\phi(x_s)$, it remains to show that the risk of $\phi(x_s)$ exceeds that of $t(x_s; u_{s-1})$ over a large portion of the parameter space.

LEMMA 4.3. *Let the family $\{F(x, \theta): \theta \in \Omega\}$ have Properties* 1 *and* 3. *If there exists an estimator $\delta(x_s)$ of $\theta_s \in \Omega$ which has a risk function bounded by some $v_0 < \infty$, then for every $\varepsilon > 0$, there exists a constant $s(\varepsilon)$ such that, for $s \geqq s(\varepsilon)$,*

$$\sup_{\theta_1 \leqq \cdots \leqq \theta_n} |R(t(x_s;\, u_{s-1});\, \theta_1, \cdots, \theta_s)-R(t(x_s;\, \mu_{s-1});\, \theta_1, \cdots, \theta_s)| \leqq \varepsilon,$$

where u_{s-1} and μ_{s-1} are defined by Property 1 *and t by Property* 3.

PROOF. Since the argument used to obtain Equation (11) applies here as well, we have for $\theta \in \Omega$,

$$(13)\qquad E_\theta[t(x_s;\, \gamma)-\theta]^2 \leqq v_0,\ \gamma \leqq \theta.$$

$$\begin{aligned} R(t(x_s;\, u_{s-1});\, \theta_1, \cdots, \theta_s) &= E_{(\theta_1,\cdots,\theta_s)}[t(x_s;\, u_{s-1})-\theta_s]^2 \\ &= E_{(\theta_1,\cdots,\theta_{s-1})}\{E_{\theta_s}\{[t(x_s;\, u_{s-1})-\theta_s]^2 \mid u_{s-1}\}\} \\ &= E_{(\theta_1,\cdots,\theta_{s-1})}\{E_{\theta_s}[t(x_s;\, u_{s-1})-\theta_s]^2\}. \end{aligned}$$

Also, $R(t(x_s;\, \mu_{s-1});\, \theta_1, \cdots, \theta_s) = E_{(\theta_1,\cdots,\theta_{s-1})}\{E_{\theta_s}[t(x_s;\, \mu_{s-1})-\theta_s]^2\}$.

Case 1. If $u_{s-1} \leqq \mu_{s-1}-\beta$, then by (13)

$$E_{\theta_s}[t(x_s;\, u_{s-1})-\theta_s]^2 \leqq v_0, \qquad E_{\theta_s}[t(x_s;\, \mu_{s-1})-\theta_s]^2 \leqq v_0. \qquad \text{Thus}$$

$$(14)\qquad |E_{\theta_s}[t(x_s;\, u_{s-1})-\theta_s]^2-E_{\theta_s}[t(x_s;\, \mu_{s-1})-\theta_s]^2| \leqq 2v_0.$$

Case 2. If $\mu_{s-1}-\beta \leqq u_{s-1} \leqq \mu_{s-1}+\beta$, then

$$E_{\theta_s}[t(x_s;\, u_{s-1})-\theta_s]^2 = E_{\theta_s}[t(x_s;\, u_{s-1})-t(x_s;\, \mu_{s-1})+t(x_s;\, \mu_{s-1})-\theta_s]^2,$$

and using the Schwarz inequality we get

$$(15)\qquad \left|E_{\theta_s}[t(x_s; u_{s-1})-\theta_s]^2 - E_{\theta_s}[t(x_s; \mu_{s-1})-\theta_s]^2\right|$$
$$\leqq E_{\theta_s}[t(x_s; u_{s-1})-t(x_s; \mu_{s-1})]^2 + 2v_0^{\frac{1}{2}}E_{\theta_s}^{\frac{1}{2}}[t(x_s; u_{s-1})-t(x_s; \mu_{s-1})]^2.$$

By (B) of Property 2, there exists a $\beta(\alpha)$ such that, if $|u_{s-1}-\mu_{s-1}| \leqq \beta(\alpha)$, then $[t(x_s; u_{s-1})-t(x_s; \mu_{s-1})]^2 \leqq \alpha$. Hence, if we choose $\alpha = \min(\varepsilon/4, \varepsilon^2/64v_0)$, then the difference in equation (15) is less than $\varepsilon/2$.

Case 3. If $\mu_{s-1}+\beta \leqq u_{s-1} \leqq \theta_s$, then as in Case 1, the relationship (14) holds. (This region need not exist.)

Case 4. If $\theta_s < u_{s-1}$, then as in Case 3 of Lemma 4.2,

$$E_{\theta_s}[t(x_s; u_{s-1})-\theta_s]^2 \leqq B+2\omega B^{\frac{1}{2}}+\omega^2+[u_{s-1}+\omega-\theta_s]^2.$$

Using the results of Cases 1, 2, 3, and 4, we find that

$$\left|R(t(x_s; u_{s-1}); \theta_1, \cdots, \theta_s)-R(t(x_s; \mu_{s-1}); \theta_1, \cdots, \theta_s)\right|$$
$$\leqq E_{(\theta_1, \cdots, \theta_{s-1})}\{\left|E_{\theta_s}[t(x_s; u_{s-1})-\theta_s]^2 - E_{\theta_s}[t(x_s; \mu_{s-1})-\theta_s]^2\right|\}$$
$$(16)\qquad \leqq 2v_0 P(u_{s-1} < \mu_{s-1}-\beta(\alpha))+\varepsilon/2+2v_0 P(\mu_{s-1}+\beta(\alpha) \leqq u_{s-1} \leqq \theta_s)$$
$$+(B+2\omega B^{\frac{1}{2}}+\omega^2)P(\mu_{s-1}+\beta(\alpha) < u_{s-1})+E[(u_{s-1}-\theta_s)^+]^2$$
$$+2\omega E^{\frac{1}{2}}[(u_{s-1}-\theta_s)^+]^2+\omega^2 P(\mu_{s-1}+\beta(\alpha) < u_{s-1}).$$

By Lemma 4.1, $\{u_s\}$ is a lower bound sequence. Therefore, for any fixed v_0, B, ω, and α (note α is a function of ε), we have that for s sufficiently large the sum of all terms in (16) is less than ε. Since v_0, B, and ω do not depend upon the particular parameter sequence, the proof is complete.

THEOREM 4.2. *Let the family* $\{F(x, \theta): \theta \in \Omega\}$ *have Properties* 1 *and* 3. *There does* not *exist a sequence of estimators* $\{\delta_k(x_1, \cdots, x_k)\}$, *where* $\delta_1(x_1)$ *has bounded risk, such that for any given* $\varepsilon > 0$, *a constant* $c(\varepsilon)$ *can be found such that* $s \geqq c(\varepsilon)$ *implies*

$$(17)\qquad R(t(x_s; u_{s-1}); \theta_1, \cdots, \theta_s)-R(\delta_s(x_1, \cdots, x_s); \theta_1, \cdots, \theta_s) \geqq \varepsilon,$$

for all $\theta_1 \leqq \cdots \leqq \theta_s$. ($u_{s-1}$ *and* t *are defined by Property* 1 *and Property* 3 *respectively.*)

PROOF. For any stage $s = 1, 2, \cdots$ consider the sequences for which $\theta_1 = \cdots = \theta_{s-1} \leqq \theta_s$. Then $\mu_{s-1} = \theta_{s-1}$. By Lemma 4.3, $s \geqq s(\varepsilon/2)$ implies

$$\left|R(t(x_s; u_{s-1}); \theta_1, \cdots, \theta_s)-R(t(x_s; \theta_{s-1}); \theta_1, \cdots, \theta_s)\right| < \varepsilon/2.$$

If there exists a sequence $\{\delta_k\}$ such that (17) holds, then it will hold, in particular, for any fixed s sufficiently large (greater than $s(\varepsilon/2)$), when $\mu_{s-1} = \theta_{s-1}$. Therefore, (17) implies that for s sufficiently large

$$R(t(x_s; \theta_{s-1}); \theta_1, \cdots, \theta_s)-R(\delta(x_1, \cdots, x_s); \theta_1, \cdots, \theta_s) \geqq \varepsilon/2,$$

for all $\theta_{s-1} \leqq \theta_s$ (θ_{s-1} fixed). This is a contradiction by Theorem 4.1.

Theorem 4.2 expresses a form of ε-admissibility which the sequence $\{t(x_s; u_{s-1})\}$ possesses. The following theorem combined with Theorem 4.2 indicates that the sequence of estimators $\{t(x_s; u_{s-1})\}$ is an improvement over any sequence of admissible minimax estimators each based on only one observation.

THEOREM 4.3. *Let the family* $\{F(x, \theta): \theta \in \Omega\}$ *have Properties* 1 *and* 3. *Let* $\phi(x_s)$ *be an estimator of* $\theta_s \in \Omega$ *with constant risk function* v. *For every* $\alpha > 0$, *define the set* $\Lambda_\alpha \subset \Omega \times \Omega$ *by*

$$\Lambda_\alpha = \{(\gamma, \theta): v - E_\theta[t(x; \gamma) - \theta]^2 \geqq \alpha, \gamma \leqq \theta, \gamma, \theta \in \Omega\},$$

where $t(\cdot\,;\cdot)$ *is defined by Property* 3. *Then, for every* $\varepsilon > 0$, *there exists a constant* $s(\varepsilon)$, *such that, for* $s \geqq s(\varepsilon)$,

(A) $R(t(x_s; u_{s-1});\ \theta_1, \cdots, \theta_s) \leqq v + \varepsilon$ *for all* $\theta_1 \leqq \cdots \leqq \theta_s$, $\theta_k \in \Omega$, $k = 1, \cdots, s$, *and*

(B) $R(t(x_s; u_{s-1});\ \theta_1, \cdots, \theta_s) \leqq v - (\alpha - \varepsilon)$, *for all* $\theta_1 \leqq \cdots \leqq \theta_s$, $\theta_k \in \Omega$, $k = 1, \cdots, s$, $(\mu_{s-1}, \theta_s) \in \Lambda_\alpha$, *where* μ_{s-1} *and* $u_{s-1} = u_{s-1}(x_1, \cdots, x_{s-1})$ *are defined by Property* 1.

PROOF. By Lemma 4.1, $\{u_k\}$ is a lower bound sequence. Since $R(\phi(x_s);\ \theta_1, \cdots, \theta_s) \equiv v$ for all $\theta_1 \leqq \cdots \leqq \theta_s$, $\theta_k \in \Omega$, $k = 1, \cdots, s$; inequality (A) follows from Lemma 4.2. Inequality (B) follows immediately from Lemma 4.3.

EXAMPLE 4.2. (Bernoulli trials). Let $L(d, \theta) = (d - \theta)^2$ and $f(x, \theta) = \theta^x(1-\theta)^{1-x}$, $x = 0, 1$, $\theta \in \Omega = [0, 1]$. The unique admissible minimax estimator of θ_s at stage s is $\frac{1}{4} + \frac{1}{2}x_s$.

The Bernoulli family is easily seen to have Property 1 by setting u_n and μ_n of the definition equal to $n^{-1}\sum_{i=1}^n x_i$ and $n^{-1}\sum_{i=1}^n \theta_i$ respectively.

Consider the functions $\{t(x; \gamma): \gamma \in \Omega\}$ where

$$\begin{aligned} t(x; \gamma) &= \tfrac{1}{4} + \tfrac{1}{2}x && \text{if } 0 \leqq \gamma \leqq \tfrac{1}{4}; \\ &= \gamma + x(1 - \gamma^{\frac{1}{2}}) && \text{if } \tfrac{1}{4} < \gamma \leqq 1. \end{aligned}$$

Then the family $\{t(x; \gamma): \gamma \in \Omega\}$ is easily seen to have Property 2.

To show that $t(x; \gamma)$ is admissible minimax for $\theta \in [\gamma, 1]$ we note the following.

(I) If $0 \leqq \gamma \leqq \frac{1}{4}$, $t(x; \gamma)$ is unque Bayes for the prior distribution which puts probability 8/15 on $\theta = \frac{1}{4}$ and probability 7/15 on $\theta = 1$. Also, the risk function of $t(x; \gamma)$ attains its maximum at the points $\theta = \frac{1}{4}$ and $\theta = 1$.

(II) If $\frac{1}{4} < \gamma \leqq 1$, $t(x; \gamma)$ is unique Bayes for the prior distribution which puts probability $(1+\gamma)^{-1}(2-\gamma^{\frac{1}{2}})^{-1}$ on $\theta = \gamma$ and probability $1-(1+\gamma)^{-1}(2-\gamma^{\frac{1}{2}})^{-1}$ on $\theta = 1$. Also, the risk function of $t(x; \gamma)$ attains its maximum at the points $\theta = \gamma$ and $\theta = 1$.

Hence (by lemma, page 4–19 of [2]) $t(x; \gamma)$ is unique admissible minimax for $\theta_s \in [\gamma, 1]$ relative to squared error loss. Therefore, the Bernoulli family has Property 3.

It follows from Theorem 4.3 that the sequence of estimators $\{t(x_s; \bar{x}_{s-1})\}$, where $\bar{x}_{s-1} = (s-1)^{-1}\sum_{i=1}^{s-1} x_i$, is asymptotically subminimax. Furthermore,

$$\begin{aligned} &E_{(\theta_1, \cdots, \theta_s)}[t(x_s; \bar{x}_{s-1}) - \theta_s]^2 \\ &\to_{s\to\infty} \tfrac{1}{16} && \text{if} \quad 0 \leqq \bar{\theta}_{s-1} \leqq \tfrac{1}{4}, \\ &\to_{s\to\infty} \bar{\theta}_{s-1}^2 - \theta_s(1 + \bar{\theta}_{s-1} - \theta_s)(2\bar{\theta}_{s-1}^{\frac{1}{2}} - 1) && \text{if} \quad \tfrac{1}{4} < \bar{\theta}_{s-1} \leqq 1 \end{aligned}$$

where $\bar{\theta}_{s-1} = (s-1)^{-1}\sum_{i=1}^{s-1} \theta_i$, whereas

$$E_{(\theta_1 \cdots, \theta_s)}[\tfrac{1}{4} + \tfrac{1}{2}x_s - \theta_s]^2 = \tfrac{1}{16}, \qquad \text{all} \quad \theta_1 \leqq \cdots \leqq \theta_s.$$

We note that $\bar{\theta}_{s-1}^2 - \theta_s(1 + \bar{\theta}_{s-1} - \theta_s)(2\bar{\theta}_{s-1}^{\frac{1}{2}} - 1) < \frac{1}{16}$ for all $\frac{1}{4} < \bar{\theta}_{s-1} \leqq \theta_s \leqq 1$ and sometimes significantly less.

5. Discussion. In Theorem 4.3 we see that the portion of the parameter space on which the risk of $t(x_s; u_{s-1})$ is lower than that of $\phi(x_s)$ and the magnitude of this difference, depend on the nature of the sets Λ_α. In most estimation problems (for example, the case of Bernoulli trials which was mentioned in Section 3), the sets Λ_α (for reasonable choices of α) are sizable compared to $\Omega \times \Omega$.

Another favorable property of the sequence $\{t(x_s; u_{s-1})\}$, is that the $s(\varepsilon)$ of Theorem 4.3 does not depend on the parameter sequence. Therefore $s(\varepsilon)$ can be calculated for any particular family $\{F(x, \theta): \theta \in \Omega\}$. This allows us to set some sort of "tolerance level" ε, by using $\phi(x_s)$ for $s < s(\varepsilon)$ and then $t(x_s; u_{s-1})$ for $s \geqq s(\varepsilon)$. This is also important in the following situation: If Ω is bounded above, then the sequence of parameters will approach a finite limit (as they are increasing). This implies that $E[u_s(s_1, \cdots, x_s) - \theta_s]^2 \to 0$, and that u_s is asymptotically, an exact estimator of θ_s. Unfortunately, unless something is known about the convergence of the parameter sequence, it is not possible to know much about the risk of u_s at any finite stage s.

REFERENCES

[1] Blumenthal, S. and Cohen, A. (1968). Estimation of two ordered translation parameters. *Ann. Math. Statist.* **39** 517–530.

[2] Lehmann, E. L. (1962). *Notes on the Theory of Estimation.* Univ. of California, Berkeley.

[3] Robertson, T. and Waltman, P. (1968). On estimating monotone parameters. *Ann. Math. Statist.* **39** 1030–1039.

[4] Sackrowitz, H. (1968). Decision theoretic and empirical Bayes approaches to inference for monotone parameter sequences. Ph.D. thesis, Columbia Univ.