CONVERGENCE OF THE LOSSES OF CERTAIN DECISION RULES FOR THE SEQUENTIAL COMPOUND DECISION PROBLEM

By Ester Samuel

The Hebrew University of Jerusalem

1. Summary. This paper is a continuation of [8], and considers the sequential compound decision problem for the case where the component decisions are of the simple versus simple hypothesis testing type, and thus can be stated in terms of testing whether $\theta = 0$ or $\theta = 1$. The loss for the compound decision is taken to be the average of the losses in the component decisions, and the risk for the compound decision is defined correspondingly. Let $R(\ )$ denote the Bayes envelope function of the component problem. In [8] two sequences of compound decision rules $\{T_n^*\}$ and $\{\hat{T}_n\}$ are exhibited, such that for n sufficiently large, the risk incurred by $\hat{T}_n$ never exceeds $R(\vartheta_n) + \epsilon$ where ϑ_n is the average of the true θ-values in the n first components, and this holds uniformly in all possible sequences of θ's: for T_n^* a corresponding statement is valid provided $R(\ )$ is differentiable for all $0 \leqq \eta \leqq 1$. Here we prove that for any sequence of θ-values, the difference between the loss incurred by $\hat{T}_n$ and $R(\vartheta_n)$ converges to zero in probability, and under the differentiability assumption a corresponding statement holding with probability one is proved for T_n^*. Numerical data is provided to indicate the rate of convergence.

2. Introduction. For convenience we shall briefly review some of the concepts and notation introduced in [8]. Since this paper leans heavily on [8], familiarity with the latter is desirable.

We are confronted with a sequence of independent random variables, $X_1, X_2, \cdots$, and "parameter values," $\theta_1, \theta_2, \cdots$, where each θ_i equals 0 or 1, and X_i has distribution function P_{θ_i}, where P_0 and P_1 are two completely specified distribution functions, given in terms of their densities $f(x, 0)$ and $f(x, 1)$, with respect to some measure μ. The sequence of θ's is unknown, and the statistician is required to decide, for each i, $i = 1, 2, \cdots$ whether θ_i equals 0 or 1. A sequential compound decision function is one where the decision about θ_i may depend upon the observed values $\mathbf{x}_i = (x_1, x_2, \cdots, x_i)$ of $\mathbf{X}_i = (X_1, X_2, \cdots, X_i)$.

The component problem is thus to decide whether the random variable X has distribution function P_0 or P_1, on the basis of an observation x on X. We assume that the loss structure is the following

Decision	True $\theta = 0$	True $\theta = 1$
$\theta = 0$	0	a
$\theta = 1$	b	0,

where $a > 0, b > 0$.

Received May 30, 1963; revised June 3, 1964.

Let t be a decision function for the component problem, i.e. t is a measurable function with $0 \leqq t(x) \leqq 1$, where $t(x)$ and $1 - t(x)$ denote the probabilities with which one decides $\theta = 1$ and $\theta = 0$ respectively, when $X = x$ is observed. The risk of t as a function of θ is thus given by

$$(1) \qquad \begin{aligned} R(t, \theta) &= bE_\theta(t(X)) && \text{for } \theta = 0 \\ &= aE_\theta(1 - t(X)) && \text{for } \theta = 1, \end{aligned}$$

where E_θ denotes the expectation with respect to P_θ, $\theta = 0, 1$. The function $(1 - \eta)R(t, 0) + \eta R(t, 1)$, for fixed η, is minimized with respect to all possible decision functions t by any measurable t_η of the form

$$(2) \qquad \begin{aligned} t_\eta(x) &= 1 && \text{if } (1 - \eta)bf(x, 0) < \eta af(x, 1) \\ &= 0 && \text{if } (1 - \eta)bf(x, 0) > \eta af(x, 1) \\ &= \text{arbitrary in } [0, 1] && \text{if } (1 - \eta)bf(x, 0) = \eta af(x, 1) \end{aligned}$$

and we denote by t_η^0 the particular non-randomized function where the arbitrary part in (2) is taken to be zero. (Notice that t_η is a "Bayes rule with respect to a priori probability η", when one considers θ as a random variable with $P(\theta = 1) = \eta = 1 - P(\theta = 0)$.) The minimum value

$$\min_t [(1 - \eta)R(t, 0) + \eta R(t, 1)] = (1 - \eta)R(t_\eta, 0) + \eta R(t_\eta, 1)$$

is denoted $R(\eta)$ and $R(\ )$ is called the Bayes envelope function.

Let Ω^∞ denote the set of all possible infinite sequences of 0's and 1's, and let Ω^n denote the set of all 2^n n-vectors of 0's and 1's. For any $\boldsymbol{\theta} \in \Omega^\infty$ let $\boldsymbol{\theta}_n \in \Omega^n$ denote its initial n-vector.

By a (n-step) *compound decision rule* we mean any n-vector $T_n = (t_1, t_2, \cdots, t_n)$ of measurable functions where $0 \leqq t_i = t_i(\mathbf{x}_i) \leqq 1$ and where t_i and $1 - t_i$ denote the probabilities with which one decides $\theta_i = 1$ and $\theta_i = 0$ respectively, when $\mathbf{X}_i = \mathbf{x}_i$ is observed. Corresponding to (1) we therefore have the risk on the ith decision

$$(3) \qquad \begin{aligned} R(t_i, \boldsymbol{\theta}_i) &= bE_{\boldsymbol{\theta}_i}(t_i(\mathbf{X}_i)) && \text{for } \theta_i = 0 \\ &= aE_{\boldsymbol{\theta}_i}(1 - t_i(\mathbf{X}_i)) && \text{for } \theta_i = 1, \end{aligned}$$

where $E_{\boldsymbol{\theta}_i}$ denotes the expectation under the product probability measure $P_{\theta_1} \times \cdots \times P_{\theta_i} = P_{\boldsymbol{\theta}_i}$. It should be noticed that (3) is a function not only of θ_i, but of all $\boldsymbol{\theta}_i$. The risk of T_n at the point $\boldsymbol{\theta}_n$ is thus defined as

$$(4) \qquad R(T_n, \boldsymbol{\theta}_n) = n^{-1} \sum_{i=1}^{n} R(t_i, \boldsymbol{\theta}_i).$$

Correspondingly we denote the (random) loss incurred in the ith decision by

$$(5) \qquad L(t_i(\mathbf{X}_i), \boldsymbol{\theta}_i).$$

(Notice that $L(t_i(\mathbf{X}_i), \boldsymbol{\theta}_i)$ depends not only on $\mathbf{X}_i$ but also on the independent random variable U_i introduced in order to carry out the randomization required by t_i, except in the case where $t_i(\mathbf{X}_i) = 0$ or 1 with probability one, i.e., the case where t_i is non-randomized, and no such random variable is required. In the latter case

$$R(t_i, \boldsymbol{\theta}_i) = E_{\boldsymbol{\theta}_i}(L(t_i(\mathbf{X}_i), \boldsymbol{\theta}_i),$$

whereas generally

$$R(t_i, \boldsymbol{\theta}_i) = E_{\boldsymbol{\theta}_i}E_{U_i}(L(t_i(\mathbf{X}_i), \boldsymbol{\theta}_i).$$

Thus the notation (5) is in fact incomplete.) Corresponding to (4) we have

$$(6) \qquad \lrcorner(T_n, \boldsymbol{\theta}_n) = n^{-1}\sum_{i=1}^{n} L(t_i(\mathbf{X}_i), \boldsymbol{\theta}_i).$$

We shall now describe the compound decision rules T_n^* and $\hat{T}_n$ discussed in [8]. Let $h(x)$ be an unbiased estimate of θ, and define, for $i = 1, 2, \cdots$

$$(7) \qquad \begin{aligned} p_i = p_i(\mathbf{x}_i) &= 0 && \text{if } i^{-1}\textstyle\sum_{j=1}^{i} h(x_j) \leqq 0 \\ &= i^{-1}\textstyle\sum_{j=1}^{i} h(x_j) && \text{if } 0 \leqq i^{-1}\textstyle\sum_{j=1}^{i} h(x_j) \leqq 1 \\ &= 1 && \text{if } 1 \leqq i^{-1}\textstyle\sum_{j=1}^{i} h(x_j) \end{aligned}$$

and set $p_0 = \frac{1}{2}$. Then $T_n^* = (t_1^*, \cdots, t_n^*)$ is defined by

$$(8) \qquad t_i^* = t_i^*(\mathbf{x}_i) = t^0_{p_{i-1}}(x_i),$$

where the right hand side of (8) is defined by (7) and the particular version of (2). Let $\vartheta_i = i^{-1}\sum_{j=1}^{i} \theta_j$. Thus, to decide on θ_i, T_n^* uses a rule which is Bayes with respect to the estimate p_{i-1} of ϑ_{i-1}.

Theorem 1 of [8] states that if the Bayes envelope function $R(\eta)$ has a derivative for $0 \leqq \eta \leqq 1$ then for every $\epsilon > 0$ there exists $N(\epsilon)$ such that for all $n \geqq N(\epsilon)$

$$(9) \qquad R(T_n^*, \boldsymbol{\theta}_n) - R(\vartheta_n) < \epsilon \quad \text{uniformly in} \quad \boldsymbol{\theta}_n \,\varepsilon\, \Omega^n.$$

Notice that (9) is a one-sided inequality, i.e., $R(\vartheta_n) + \epsilon$ is an upper bound for $R(T_n^*, \boldsymbol{\theta}_n)$. Since one is interested that the risk should be as small as possible, an upper bound is of particular interest. Intuitively one would however not expect that $R(T_n^*, \boldsymbol{\theta}_n)$, for n sufficiently large, could be less than $R(\vartheta_n) - \epsilon$. We shall in fact show that if $R(\eta)$ is differentiable for $0 \leqq \eta \leqq 1$, then

$$(10) \qquad \lim_{n\to\infty} |R(T_n^*, \boldsymbol{\theta}_n) - R(\vartheta_n)| = 0$$

uniformly in all $\boldsymbol{\theta} \,\varepsilon\, \Omega^\infty$.

We shall make use of the following vector notation introduced in [8]. For any $Y = (y_i, y_2)$ with $y_1 \geqq 0$, $y_2 \geqq 0$, and $y_1 + y_2 > 0$, let

$$(11) \qquad \mathcal{R}(Y) = \begin{bmatrix} R(t_\eta^0, & 0) \\ R(t_\eta^0, & 1) \end{bmatrix}$$

where $\eta = y_2/(y_1 + y_2)$, and where the right hand side of (11) is defined through (1) and (2). Expressions (16), (17) and (18) of [8] state

$$\mathcal{R}(cY) = \mathcal{R}(Y) \qquad \text{for } c > 0 \tag{12}$$

$$Y^*\mathcal{R}(Y) \geqq Y^*\mathcal{R}(Y^*) \tag{13}$$

$$Y[\mathcal{R}(Y^*) - \mathcal{R}(Y)] \leqq (Y - Y^*)[\mathcal{R}(Y^*) - \mathcal{R}(Y)] \tag{14}$$

Let

$$\varphi_i = (1 - \theta_i, \theta_i) \quad \text{and} \quad \psi_i = \sum_{j=1}^{i} \varphi_j$$

Lemma 2 of [8] states that for all $\boldsymbol{\theta}_n \in \Omega^n$

$$n^{-1}\sum_{i=1}^{n} R(t^0_{\vartheta_i}, \theta_i) \leqq R(\vartheta_n).$$

We shall now prove

LEMMA 1: *If $R(\eta)$ is differentiable for $0 \leqq \eta \leqq 1$, then*

$$\lim_{n\to\infty}\left[R(\vartheta_n) - n^{-1}\sum_{i=1}^{n} R(t^0_{\vartheta_i}, \theta)\right] = 0 \tag{15}$$

uniformly in all $\boldsymbol{\theta} \in \Omega^\infty$.

PROOF. By the definitions, (14) and (12) it follows that

$$\begin{aligned}
0 \leqq R(\vartheta_n) - n^{-1}\sum_{i=1}^{n} R(t^0_{\vartheta_i}, \theta_i) &= n^{-1}\psi_n\mathcal{R}(\psi_n) - n^{-1}\sum_{i=1}^{n}\varphi_i\mathcal{R}(\psi_i)\\
&= n^{-1}\left\{\psi_n\mathcal{R}(\psi_n) - \left(\psi_n\mathcal{R}(\psi_n) + \sum_{i=1}^{n-1}\psi_i[\mathcal{R}(\psi_i) - \mathcal{R}(\psi_{i+1})]\right)\right\}\\
&= n^{-1}\sum_{i=1}^{n-1}\psi_i[\mathcal{R}(\psi_{i+1}) - \mathcal{R}(\psi_i)]\\
&\leqq n^{-1}\sum_{i=1}^{n-1}(\psi_i - \psi_{i+1})[\mathcal{R}(\psi_{i+1}) - \mathcal{R}(\psi_i)]\\
&= n^{-1}\sum_{i=1}^{n-1}\varphi_i[\mathcal{R}(1 - \vartheta_i, \vartheta_i) - \mathcal{R}(1 - \vartheta_{i+1}, \vartheta_{i+1})]\\
&\leqq n^{-1}\sum_{i=1}^{n-1}\text{m.a.v.}[\mathcal{R}(1 - \vartheta_i, \vartheta_i) - \mathcal{R}(1 - \vartheta_{i+1}, \vartheta_{i+1})]
\end{aligned} \tag{16}$$

where m.a.v. $Y = \max(|y_1|, |y_2|)$. Notice that

$$|(1 - \vartheta_i) - (1 - \vartheta_{i+1})| = |\vartheta_i - \vartheta_{i+1}| \leqq 1/(i + 1). \tag{17}$$

Let $\epsilon > 0$ be given. Notice that since $R(\eta)$ is differentiable, Lemma 1 of [8] is applicable. From that lemma it follows that

$$\text{m.a.v. }[\mathcal{R}(1 - \vartheta_i, \vartheta_i) - \mathcal{R}(1 - \vartheta_{i+1}, \vartheta_{i+1})] < \epsilon/2$$

for $|\vartheta_i - \vartheta_{i+1}| < \delta$, thus, by (17), for $i > N(\delta)$. Thus

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n-1} \text{m.a.v.}[\mathcal{R}(1 - \vartheta_i, \vartheta_i) - \mathcal{R}(1 - \vartheta_{i+1}, \vartheta_{i+1})] \\ (18) \qquad & \leqq n^{-1}\sum_{i=1}^{N(\delta)} \max(a, b) + n^{-1}(n - 1 - N(\delta))\epsilon/2 \\ & < \epsilon/2 + \epsilon/2 = \epsilon \end{aligned}$$

provided $n > 2N(\delta)\max(a, b)/\epsilon$. Finally (15) follows from (16) and (18).

REMARKS:

(1) Quite similarly to Lemma 1, one can prove that if $R(\eta)$ is differentiable for $0 \leqq \eta \leqq 1$ and j is any fixed integer (positive or non-positive), then uniformly in $\boldsymbol{\theta} \in \Omega^\infty$

$$\lim_{n\to\infty} \left| R(\vartheta_n) - n^{-1} \sum_{i=1}^{n} R(t^0_{\vartheta_{i-j}}, \theta_i) \right| = 0 \tag{19}$$

(where one defines $\vartheta_{i-j} = \frac{1}{2}$ (or any other number) for $i < j$).

(2) Notice that if at the ith stage one uses a rule Bayes against ϑ_{i-1} and arbitrary at the first stage, then the compound loss is always greater than, or equal to $R(\vartheta_n)$. That is, for any $\boldsymbol{\theta}_n \in \Omega^n$

$$n^{-1}\sum_{i=1}^{n} R(t^0_{\vartheta_{i-1}}, \theta_i) \geqq R(\vartheta_n). \tag{20}$$

(20) follows since by (13) (where $\mathcal{R}(\psi_0)$ may be arbitrary)

$$\begin{aligned} n^{-1}\sum_{i=1}^{n} R(t^0_{\vartheta_{i-1}}, \theta_i) &= n^{-1}\sum_{i=1}^{n} \varphi_i\mathcal{R}(\psi_{i-1}) \\ &= n^{-1}(\psi_n\mathcal{R}(\psi_{n-1}) + \sum_{i=1}^{n-1} \psi_i[\mathcal{R}(\psi_{i-1}) - \mathcal{R}(\psi_i)]) \\ &\geqq n^{-1}\psi_n\mathcal{R}(\psi_{n-1}) \geqq n^{-1}\psi_n\mathcal{R}(\psi_n) = R(\vartheta_n). \end{aligned}$$

It should be remembered that the rule T_n^* uses at the ith stage, a rule Bayes against the estimate p_{i-1} of ϑ_{i-1}.

(3) Lemma 1 and (19) are *not* valid if $R(\eta)$ is not differentiable for all $0 \leqq \eta \leqq 1$. In fact there exist sequences $\boldsymbol{\theta} \in \Omega^\infty$ such that the limit in the left hand side of (19) exists but differs from 0. This can be seen in the following example. Let

$$\begin{aligned} f(x, 0) &= 1 \quad \text{for } 0 \leqq x \leqq 1, & f(x, 1) &= 2 \quad \text{for } 0 \leqq x \leqq \tfrac{1}{2} \\ &= 0 \quad \text{otherwise,} & &= 0 \quad \text{otherwise.} \end{aligned}$$

Then

$$
\begin{aligned}
t_\eta(x) &= 0 \quad \text{for all } x && \text{if } 0 \leqq \eta < \tfrac{1}{3} \\
t_{\frac{1}{3}}(x) &= \text{arbitrary in } [0, 1] && \text{for } 0 \leqq x \leqq \tfrac{1}{2} \\
&= 0 && \text{for } \tfrac{1}{2} < x \leqq 1 \\
t_\eta(x) &= 1 \quad \text{for } 0 \leqq x \leqq \tfrac{1}{2} \\
&= 0 \quad \text{for } \tfrac{1}{2} < x \leqq 1, && \text{if } \tfrac{1}{3} < \eta \leqq 1 \\
R(t_\eta, \theta) &= 0 \quad \text{for } \theta = 0 \\
&= 1 \quad \text{for } \theta = 1, && \text{if } 0 \leqq \eta < \tfrac{1}{3}
\end{aligned}
$$

and the above is also $R(t^0_{\frac{1}{3}}, \theta)$.

$$
\begin{aligned}
R(t_\eta, \theta) &= \tfrac{1}{2} \quad \text{for } \theta = 0 \\
&= 0 \quad \text{for } \theta = 1 \quad \text{if } \tfrac{1}{3} < \eta \leqq 1.
\end{aligned}
$$

Thus

$$
\begin{aligned}
R(\eta) &= \eta && \text{for } 0 \leqq \eta \leqq \tfrac{1}{3} \\
&= \tfrac{1}{2}(1 - \eta) && \text{for } \tfrac{1}{3} \leqq \eta \leqq 1
\end{aligned}
$$

Thus for $\boldsymbol{\theta} = 001001001\cdots$ we have

$$\lim_{n\to\infty} \left[R(\vartheta_n) - n^{-1} \sum_{i=1}^n R(t^0_{\vartheta_i}, \theta_i) \right] = \frac{1}{3}.$$

For $\boldsymbol{\theta} = 100100100\cdots$ we have

$$\lim_{n\to\infty} \left[R(\vartheta_n) - n^{-1} \sum_{i=1}^n R(t^0_{\vartheta_{i-1}}, \theta_i) \right] = -\frac{1}{3}.$$

It should be noticed that this result is due to the fact that we chose $t^0_{\frac{1}{3}}$ and not any other $t_{\frac{1}{3}}$. However, for any particular choice of t_η examples of the above type can be constructed. Actually (15) and (19) under the differentiability assumption and (20) generally are valid also when one omits the superscripts 0. (The above is an example of a statistical game corresponding to the example of the non-statistical game of matching pennies given in [8].)

To prove (10) we notice that the proof of Theorem 1 in [8] actually implies

$$\left| R(T^*_n, \boldsymbol{\theta}_n) - n^{-1} \sum_{i=1}^n R(t^0_{\vartheta_i}, \theta_i) \right| < \epsilon \tag{21}$$

for all $n > N(\epsilon)$, uniformly in $\boldsymbol{\theta}_n \, \varepsilon \, \Omega^n$. Thus (10) follows from (21) and Lemma 1.

Theorem 2 of [8] states that no matter whether or not $R(\eta)$ is differentiable, the decision rule $\hat{T}_n$ is such that for any $\epsilon > 0$ there exists $N(\epsilon)$ such that for all $n \geqq N(\epsilon)$

$$R(\hat{T}_n, \boldsymbol{\theta}_n) \leqq R(\vartheta_n) + \epsilon \quad \text{uniformly in } \boldsymbol{\theta}_n \, \varepsilon \, \Omega^n. \tag{22}$$

The decision rule $\hat{T}_n = (\hat{t}_1, \cdots, \hat{t}_n)$ is however harder to describe, and also the proof of (22) is more complicated than that of (9). In fact, $\hat{t}_i(\mathbf{x}_i) = t^0_{m_i}(x_i)$, i.e., $\hat{t}_i$ is also a Bayes rule with respect to some subscript m_i. m_i is however a function depending not on $\mathbf{x}_{i-1}$ alone, but also on a random variable $Z = (z_1, z_2)$ uniformly distributed on the unit square, and thus even for given $\mathbf{x}_{i-1}$, m_i is a random variable. In fact, consider a sequence $Z_1, Z_2, \cdots$ of independent random variables uniformly distributed on the unit square, i.e., $Z_i = (z^i_1, z^i_2)$. Let, for $i = 0, 1, \cdots$

$$m_{i+1} = m(Z_{i+1}, \mathbf{x}_i) = \frac{p_i(\mathbf{x}_i) + i^{-\frac{1}{4}} z_2^{i+1}}{1 + i^{-\frac{1}{4}}(z_1^{i+1} + z_2^{i+1})} \tag{23}$$

where we now assume that the estimate $h(x)$ used in the definition (7) of p_i is bounded, i.e., $|h(x)| \leqq H$. Then

$$\hat{t}_i(\mathbf{x}_i) = t^0_{m_i}(x_i). \tag{24}$$

Notice that in correspondence with (3), here

$$\begin{aligned} R(\hat{t}_i, \boldsymbol{\theta}_i) &= bE_{\boldsymbol{\theta}_i}E_{Z_i}(\hat{t}_i(\mathbf{X}_i, Z_i)) \qquad &\text{if } \theta_i = 0 \\ &= aE_{\boldsymbol{\theta}_i}E_{Z_i}(1 - \hat{t}_i(\mathbf{X}_i, Z_i)) \qquad &\text{if } \theta_i = 1 \end{aligned}$$

where E_{Z_i} denotes the expectation under the distribution of Z_i. (It should be remarked that in order to verify (22) one need not assume that $Z_1, Z_2, \cdots$ are independent random variables, since $R(\hat{T}_n, \boldsymbol{\theta}_n)$ is the average of the expectations of the losses at the individual decisions, and expectations of a sum equals the sum of the expectations, whether or not the random variables are independent. Since we shall in the next section discuss the losses incurred by $\hat{T}_n$ we shall assume that $\hat{t}_i$ is defined through (23) and (24) where the Z_i's are independent, and also independent of the X's.)

3. Convergence of $L(T^*_n, \boldsymbol{\theta}_n) - R(\vartheta_n)$ and $L(\hat{T}_n, \boldsymbol{\theta}_n) - R(\vartheta_n)$ to zero, for every $\boldsymbol{\theta} \varepsilon \Omega^\infty$. The proofs of Theorems 1 and 2 which follow lean heavily on a martingale theorem, which is a direct consequence of the stability theorem, and is stated as

Lemma 2: *Let $\{Y_n\}$ be a martingale relative to $\{\mathfrak{F}_n\}$. (See [4] p. 294.) Define $Y_0 = 0$ and let $\sigma_n^2 = \mathrm{Var}(Y_n - Y_{n-1})$, $n \geqq 1$. Let $\{b_n\}$ be a monotone sequence such that $\lim_{n\to\infty} b_n = \infty$. If*

$$\sum_{n=1}^{\infty} \sigma_n^2/b_n^2 < \infty, \tag{25}$$

then

$$b_n^{-1} Y_n \to 0 \quad \text{a.e.} \tag{26}$$

Proof. Let $V_n = Y_n - Y_{n-1}$, $n \geqq 1$. Then under (25) the stability theorem implies (see [6] p. 387 E)

$$b_n^{-1} \sum_{i=1}^{n} [V_i - E(V_i \mid V_1, \cdots, V_{i-1})] \to 0 \quad \text{a.e.} \tag{27}$$

But since $\{Y_n\}$ is a martingale relative to $\{\mathfrak{F}_n\}$ it follows that the σ-field generated by $V_1, \cdots, V_{i-1}$ is a sub-σ-field of $\mathfrak{F}_{i-1}$, and thus (by the smoothing property)

$$E(V_i \mid V_1, \cdots, V_{i-1}) = E(E(V_i \mid \mathfrak{F}_{i-1}) \mid V_1, \cdots, V_{i-1}) = 0 \quad \text{a.e.}$$

and (27) becomes equivalent to (26).

Consider the loss of T_n^*, $L(T_n^*, \boldsymbol{\theta}_n)$ defined for T_n^* through (6), and for any $\boldsymbol{\theta} \,\varepsilon\, \Omega^\infty$ let $P_{\boldsymbol{\theta}}$ denote the (infinite) product probability corresponding to $\boldsymbol{\theta}$. We have

THEOREM 1. *If $R(\eta)$ is differentiable for $0 \leqq \eta \leqq 1$, then for any $\boldsymbol{\theta} \,\varepsilon\, \Omega^\infty$ $P_{\boldsymbol{\theta}}[\lim_{n\to\infty}(L(T_n^*, \boldsymbol{\theta}_n) - R(\vartheta_n)) = 0] = 1$.*

PROOF. We shall use Lemma 2 and shall let $\mathfrak{F}_n$ be the product σ-field generated by $\mathbf{X}_n$, (i.e., the smallest σ-field with respect to which all X_i, $i = 1, \cdots, n$ are measurable). Set

$$Y_n^* = nL(T_n^*, \boldsymbol{\theta}_n) - \sum_{i=1}^n R(t_i^*, \boldsymbol{\theta}_i \mid \mathbf{X}_{i-1}) = \sum_{i=1}^n [L(t_i^*(\mathbf{X}_i), \boldsymbol{\theta}_i) - R(t_i^*, \boldsymbol{\theta}_i \mid \mathbf{X}_{i-1})]$$

where $R(t_i^*, \boldsymbol{\theta}_i \mid \mathbf{X}_{i-1})$ denotes the conditional risk of t_i^*, given the vector $\mathbf{X}_{i-1}$ of previous random variables, (and is defined in (19) of [8]). It follows that $\{Y_n^*\}$ is a martingale relative to $\{\mathfrak{F}_n\}$, since

$$\begin{aligned}
E_{\boldsymbol{\theta}}(Y_n^* \mid \mathfrak{F}_{n-1}) &= E_{\boldsymbol{\theta}}\left(\sum_{i=1}^n [L(t_i^*(\mathbf{X}_i), \boldsymbol{\theta}_i) - R(t_i^*, \boldsymbol{\theta}_i \mid \mathbf{X}_{i-1})] \mid \mathfrak{F}_{n-1}\right) \\
&= E_{\boldsymbol{\theta}}\left(\sum_{i=1}^{n-1} [L(t_i^*(\mathbf{X}_i), \boldsymbol{\theta}_i) - R(t_i^*, \boldsymbol{\theta}_i \mid \mathbf{X}_{i-1})] \mid \mathfrak{F}_{n-1}\right) \\
&\quad + R(t_n^*, \boldsymbol{\theta}_n \mid \mathbf{X}_{n-1}) - R(t_n^*, \boldsymbol{\theta}_n \mid \mathbf{X}_{n-1}) \\
&= \sum_{i=1}^{n-1} [L(t_i^*(\mathbf{X}_i), \boldsymbol{\theta}_i) - R(t_i^*, \boldsymbol{\theta}_i \mid \mathbf{X}_{i-1})] = Y_{n-1}^* \quad (\text{a.e. } P_{\boldsymbol{\theta}}).
\end{aligned} \tag{28}$$

Now with the above definition of Y_n^* it follows that

$$P[0 \leqq |Y_n^* - Y_{n-1}^*| \leqq M] = 1$$

where $M = \max(a, b)$, and thus $\sigma_n^2 \leqq M^2$. Condition (25) is thus satisfied with $b_n = n$ and (26) yields

$$\lim_{n\to\infty}\left[L(T_n^*, \boldsymbol{\theta}_n) - n^{-1}\sum_{i=1}^n R(t_i^*, \boldsymbol{\theta}_i \mid \mathbf{X}_{i-1})\right] = 0 \quad (\text{a.e. } P_{\boldsymbol{\theta}}). \tag{29}$$

Thus, if we establish

$$\lim_{n\to\infty}\left[n^{-1}\sum_{i=1}^n R(t_i^*, \boldsymbol{\theta}_i \mid \mathbf{X}_{i-1}) - R(\vartheta_n)\right] = 0 \quad (\text{a.e. } P_{\boldsymbol{\theta}}), \tag{30}$$

then the theorem follows upon considering the intersection of the events described in (29) and (30), which also must have probability one.

To establish (30) we observe that given any $\delta > 0$ there exists $N(\delta)$ such that

$$P_{\boldsymbol{\theta}}(|p_{n-1}(\mathbf{X}_{n-1}) - \vartheta_n| < \delta \quad \text{for all } n > N(\delta)) > 1 - \delta, \tag{31}$$

this following from the strong law of large numbers and (17). Now, under the condition that $R(\eta)$ is differentiable for $0 \leqq \eta \leqq 1$, Lemma 1 of [8] is applicable, i.e., we can choose $\delta > 0$ such that

$$\max_{\theta=0,1}|R(t_\eta^0, \theta) - R(t_{\eta^*}^0, \theta)| < \epsilon/4 \quad \text{whenever } |\eta - \eta^*| < \delta.$$

Thus it follows from (31) that

$$P_{\boldsymbol{\theta}}(|R(t_n^*, \boldsymbol{\theta}_n \mid \mathbf{X}_{n-1}) - R(t_{\vartheta_n}^0, \theta_n)| < \epsilon/4 \quad \text{for all } n > N(\delta)) > 1 - \delta$$

i.e., there exists an $N(\delta, \epsilon)$ such that

$$(32) \qquad P_{\boldsymbol{\theta}}\left(\left|n^{-1}\sum_{i=1}^n R(t_i^*, \boldsymbol{\theta}_i \mid \mathbf{X}_{i-1}) - n^{-1}\sum_{i=1}^n R(t_{\vartheta_i}^0, \theta_i)\right| < \epsilon/2 \quad \text{for all} \quad n > N(\delta, \epsilon)\right) > 1 - \delta.$$

Now also Lemma 1 of the present paper is applicable, and thus from (15) and (32) it follows that there exists an $N_1(\delta, \epsilon)$ such that

$$(33) \qquad P_{\boldsymbol{\theta}}\left(\left|n^{-1}\sum_{i=1}^n R(t_i^*, \boldsymbol{\theta}_i \mid \mathbf{X}_{i-1}) - R(\vartheta_n)\right| < \epsilon \quad \text{for all} \quad n > N_1(\delta, \epsilon)\right) > 1 - \delta.$$

Since (33) is true for every $\epsilon > 0$ and $\delta > 0$, (30) follows and the proof is complete.

Remarks.

(1) Comparing Theorem 1 with Theorem 1 of [8], or correspondingly, with (10), it will be noticed that in our present theorem we lack the uniformity assertion which we had there, i.e., we have not been able to prove that given any $\epsilon > 0$ there exists $N(\epsilon)$ such that

$$P_{\boldsymbol{\theta}}(|L(T_n^*, \boldsymbol{\theta}_n) - R(\vartheta_n)| < \epsilon \quad \text{for all } n > N(\epsilon)) > 1 - \epsilon$$

uniformly in all $\boldsymbol{\theta} \,\varepsilon\, \Omega^\infty$. Notice that we have made no use of (10) in our proof of Theorem 1. In fact, a statement similar to (10), but lacking the uniformity in all $\boldsymbol{\theta} \,\varepsilon\, \Omega^\infty$, is a direct consequence of our present theorem.

(2) In [5] Hannan and Robbins consider the compound decision problem for the non-sequential case, i.e., the situation where all n random variables $X_1, X_2, \cdots, X_n$ may be observed before the decisions on θ_i, $i = 1, \cdots, n$ have to be made. The decision about θ_i may thus depend on $\mathbf{X}_n$. For a decision rule suggested in [5], Hannan and Robbins show (Theorem 3 in [5]) that

$$(34) \qquad P_{\boldsymbol{\theta}}(L(\boldsymbol{\theta}_n) - R(\vartheta_n) \leqq \epsilon \quad \text{for all } n > N) > 1 - \epsilon$$

uniformly in $\boldsymbol{\theta} \,\varepsilon\, \Omega^\infty$, where $L(\boldsymbol{\theta}_n)$ denotes the (random) loss incurred by their rule in the n first decisions. From (34) they conclude (Theorem 4 in [5]) a statement corresponding to (9), for their rule. Actually it can be deduced from

Theorem 5 in [5] that one may substitute $|L(\boldsymbol{\theta}_n) - R(\vartheta_n)|$ for $L(\boldsymbol{\theta}_n) - R(\vartheta_n)$ in (34), and the new statement will be correct. Thus, in the non-sequential case a statement corresponding to our Theorem 1, can be shown to hold *uniformly* in $\boldsymbol{\theta} \varepsilon \Omega^{\infty}$.

We shall now prove a theorem similar to Theorem 1, for the rule $\hat{T}_n$ but where the almost sure convergence is replaced by convergence in probability. Since the loss on the ith decision, when using $\hat{T}_n$, depends on Z_i as well as on $\mathbf{X}_i$, we shall denote it by $L(\hat{t}_i(\mathbf{X}_i, Z_i), \boldsymbol{\theta}_i)$ and write

$$L(\hat{T}_n, \boldsymbol{\theta}_n) = n^{-1}\sum_{i=1}^{n} L(\hat{t}_i(X_i, Z_i), \boldsymbol{\theta}_i).$$

We shall denote by $P_{\boldsymbol{\theta},Z}$ the product probability measure induced by $\boldsymbol{\theta}$ and the sequence of independent Z's. We then have

THEOREM 2. *For any* $\boldsymbol{\theta} \varepsilon \Omega^{\infty}$ *and every* $\epsilon > 0$ $\lim_{n\to\infty} P_{\boldsymbol{\theta},Z}[|L(\hat{T}_n, \boldsymbol{\theta}_n) - R(\vartheta_n)| < \epsilon] = 1$.

PROOF. Let $L(\hat{t}_i(\mathbf{X}_i), \boldsymbol{\theta}_i) = E_{Z_i}(L(\hat{t}_i(\mathbf{X}_i, Z_i), \boldsymbol{\theta}_i) \mid \mathbf{X}_i)$. Notice that $L(\hat{t}_i(\mathbf{X}_i), \boldsymbol{\theta}_i)$ is again a random variable. However, conditionally on *any* fixed sequence $\mathbf{X} = X_1, X_2, \cdots$ it follows that $L(\hat{t}_i(\mathbf{X}_i, Z_i), \boldsymbol{\theta}_i)$ are independent random variables, (here we use the fact that the Z_i's are independent), and are uniformly bounded by 0 and $M = \max(a, b)$. Thus for any fixed sequence $\mathbf{X}$ it follows by the strong law of large numbers

$$P_{\boldsymbol{\theta},Z}\left[\lim_{n\to\infty}\left(n^{-1}\sum_{i=1}^{n} L(\hat{t}_i(\mathbf{X}_i, Z_i), \boldsymbol{\theta}) - n^{-1}\sum_{i=1}^{n} L(\hat{t}_i(\mathbf{X}_i), \boldsymbol{\theta}_i)\right) = 0 \mid \mathbf{X}\right] = 1$$

and thus also

$$P_{\boldsymbol{\theta},Z}\left[\lim_{n\to\infty}\left(n^{-1}\sum_{i=1}^{n} L(\hat{t}_i(\mathbf{X}_i, Z_i), \boldsymbol{\theta}_i) - n^{-1}\sum_{i=1}^{n} L(\hat{t}_i(\mathbf{X}_i), \boldsymbol{\theta}_i)\right) = 0\right] = 1. \tag{35}$$

Hence if we show that for every $\epsilon > 0$

$$\lim_{n\to\infty} P_{\boldsymbol{\theta}}\left[\left|n^{-1}\sum_{i=1}^{n} L(\hat{t}_i(\mathbf{X}_i), \boldsymbol{\theta}_i) - R(\vartheta_n)\right| < \epsilon\right] = 1, \tag{36}$$

the theorem will follow when considering the intersection of the events described in (35) and (36).

To establish (36) we shall again make use of Lemma 2. Again let $\mathfrak{F}_n$ be the σ-field generated by $\mathbf{X}_n$. Let

$$\hat{Y}_n = \sum_{i=1}^{n} [L(\hat{t}_i(\mathbf{X}_i), \boldsymbol{\theta}_i) - R(\hat{t}_i, \boldsymbol{\theta}_i \mid \mathbf{X}_{i-1})]$$

where $R(\hat{t}_i, \boldsymbol{\theta}_i \mid \mathbf{X}_{i-1})$ is the conditional risk of $\hat{t}_i$ when the preceding random variables $\mathbf{X}_{i-1}$ are given (and is defined in (33) of [8]). Since

$$E_{\boldsymbol{\theta}}(L(\hat{t}_n(\mathbf{X}_n), \boldsymbol{\theta}_n) \mid \mathfrak{F}_{n-1}) = R(\hat{t}_n, \boldsymbol{\theta}_n \mid \mathbf{X}_{n-1})$$

it follows exactly as in the proof of (28), with $*$ replaced by $\wedge$, that $\hat{Y}_n$ is a martingale, and finally as in the proof of Theorem 1, that

$$P_{\boldsymbol{\theta}}\left[\lim_{n\to\infty}\left(n^{-1}\sum_{i=1}^{n} L(\hat{t}_i(\mathbf{X}_i), \boldsymbol{\theta}_i) - n^{-1}\sum_{i=1}^{n} R(\hat{t}_i, \boldsymbol{\theta}_i \mid \mathbf{X}_{i-1})\right) = 0\right] = 1.$$

Thus in order to establish (36) it only remains to prove that for every $\epsilon > 0$

$$\lim_{n\to\infty} P_{\boldsymbol{\theta}}\left[\left|n^{-1}\sum_{i=1}^{n} R(\hat{t}_i, \boldsymbol{\theta}_i \mid \mathbf{X}_{i-1}) - R(\vartheta_n)\right| < \epsilon\right] = 1. \tag{37}$$

Now in the proof of Theorem 2 of [8] we show that for any $\epsilon > 0$ one can find a measurable set S_n (defined in (36) of [8]) such that $P_{\boldsymbol{\theta}_n}(S_n) > 1 - \epsilon$ and for all $\mathbf{x}_n \varepsilon S_n$ and n sufficiently large

$$n^{-1}\sum_{i=1}^{n} R(\hat{t}_i, \boldsymbol{\theta}_i \mid \mathbf{x}_{i-1}) - R(\vartheta_n) < \epsilon. \tag{38}$$

(See (37) of [8].) One may in fact, for every $\mathbf{x}_n \varepsilon S_n$, replace the left hand side of (38) by its absolute value, and the statement will still be valid. This follows from the proof of (37) in [8]. Thus

$$\left|n^{-1}\sum_{i=1}^{n} R(\hat{t}_i, \boldsymbol{\theta}_i \mid \mathbf{x}_{i-1}) - R(\vartheta_n)\right| < \epsilon \qquad \text{for} \quad \mathbf{x}_n \varepsilon S_n. \tag{39}$$

Since $P(S_n) > 1 - \epsilon$, (37) follows, and the proof of the theorem is complete.

REMARKS.

(1) From the above argument and the fact that also for $\mathbf{x}_n \not\varepsilon S_n$ the left hand side of (39) is bounded by M it follows that Theorem 2 of [8] could actually be replaced by:

For every $\epsilon > 0$ there exists $N(\epsilon)$ such that for all $n > N(\epsilon)$

$$|R(\hat{T}_n, \boldsymbol{\theta}_n) - R(\vartheta_n)| < \epsilon$$

uniformly in $\boldsymbol{\theta}_n \varepsilon \Omega^n$, i.e.,

$$\lim_{n\to\infty} |R(\hat{T}_n, \boldsymbol{\theta}_n) - R(\vartheta_n)| = 0 \tag{40}$$

uniformly in all $\boldsymbol{\theta} \varepsilon \Omega^\infty$. This corresponds to the change of Theorem 1 of [8] into (10).

(2) The author believes that a statement similar to Theorem 2, with weak convergence replaced by convergence almost everywhere is also true. It should be noticed that almost sure convergence would follow if one shows that a statement corresponding to (37) holds almost surely.

(3) It should be noticed that the approach adopted throughout this paper was non-Bayesian, i.e., we did not consider the parameter values θ_i as realizations of any stochastic process, and thus considered all $\boldsymbol{\theta} \varepsilon \Omega^\infty$. The usual Bayesian approach would be to assume that θ_i are independent, identically distributed random variables with $P(\theta_i = 1) = \eta^* = 1 - P(\theta_i = 0)$. In this case we need only consider the subset $\Omega^\infty(\eta^*)$ of Ω^∞, consisting of all $\boldsymbol{\theta}$ for which $\lim_{n\to\infty} \vartheta_n = \eta^*$ since the (random) $\boldsymbol{\theta}$ will belong to $\Omega^\infty(\eta^*)$ with probability one by the strong

law of large numbers. (Since η^* may not be known one should consider $\Omega^\infty(\eta)$ for all $0 \leqq \eta \leqq 1$.) Obviously Theorems 1 and 2 are valid for this situation, and could then be changed to read: Whenever $R(\eta)$ is differentiable for $0 \leqq \eta \leqq 1$

$$\lim_{n\to\infty} L(T_n^*, \boldsymbol{\theta}_n) = R(\eta^*) \quad \text{a.e.} \tag{41}$$

and correspondingly

$$\lim_{n\to\infty} L(\hat{T}_n, \boldsymbol{\theta}_n) = R(\eta^*) \quad \text{in probability,} \tag{42}$$

(where the a.e. and "in probability" statement of (41) and (42), respectively, are under the product distribution generated by η^* where P_θ is the conditional distribution of X_i given $\theta_i = \theta$). In fact (41) will hold provided only $R(\eta)$ is differentiable at the point $\eta = \eta^*$. Similar restatements of (10) and (40) for this situation are possible. It should, however, be noticed that Theorems 1 and 2 are stronger than (41) and (42).

(4) From theorems obtained by Blackwell in [2] and [3] he derives, for finite nonstatistical games, results related to those obtained here. In [2] and [3] it is not required that the sequence $\boldsymbol{\theta}$ be fixed in advance, and thus Nature may make his choice of $\boldsymbol{\theta}_i$ depend upon the losses in decisions $1, \cdots, i-1$. On the other hand it is assumed that the value ϑ_{i-1} and the past cumulative loss are known before the ith decision, $i = 1, 2, \cdots$, must be made.

4. Numerical examples indicating the rate of convergence of the risks and losses. From Theorem 1 of [8] and Theorem 1 of the present paper it follows quite easily that also the rule T_n^{**} with $t_i^{**}(\mathbf{x}_i) = t_{p_i}^0(x_i)$ has, whenever $R(\eta)$ is differentiable for $0 \leqq \eta \leqq 1$, a loss function $L(T_n^{**}, \boldsymbol{\theta}_n)$ satisfying

$$P_{\boldsymbol{\theta}}[\lim_{n\to\infty}(L(T_n^{**}, \boldsymbol{\theta}_n) - R(\vartheta_n)) = 0] = 1$$

for every $\boldsymbol{\theta} \,\varepsilon\, \Omega^\infty$. Notice that T_n^{**} differs from T_n^* in that at the ith step it uses an estimate of ϑ_i rather than of ϑ_{i-1} only.

The practical value of any of the rules discussed will be known only after one has some idea about the rate of the convergence of the risk functions and of the losses. Thus, extensive computations by means of Monte Carlo methods seem desirable. Because of limited resources, we have at present been able to furnish only quite limited information regarding this problem, but as seen from the tables below, the method seems to be quite useful already for moderate n.

In our example $f(x, 0)$ and $f(x, 1)$ are taken to be the normal density with variance one and mean 0 and 2 respectively. We consider the case $a = b = 1$. From (2) it follows that for this example

$$\begin{aligned} t_\eta^0(x) &= 1 \quad \text{if } x > 1 + \tfrac{1}{2}\log[(1-\eta)/\eta] \\ &= 0 \quad \text{otherwise} \end{aligned}$$

By symmetry it is easily seen that the minimax rule is given for $\eta = \frac{1}{2}$ and is

$$\begin{aligned} t_{\min}(x) &= 1 \quad \text{if } x > 1 \\ &= 0 \quad \text{otherwise} \end{aligned}$$

The values $\log[(1 - \eta)/\eta]$ have been tabulated for $0(0.001)1$ by Berkson in [1], pp. 568–569, and since our computations were performed on a desk calculator, we made extensive use of these tables. The normal deviates for which the calculations were made are taken from the three first columns of [7], pp. 1–4, and are referred to as sequences 1, 2, and 3 respectively. Whenever $\theta_i = 1$ the value 2 was added to the corresponding deviate of the tables, to make it a $N(2, 1)$ deviate. The unbiassed estimator $h(x)$ of θ was taken to be $h(x) = x/2$ and p_i was computed through (7).

Calculations were made for two kinds of sequences $\boldsymbol{\theta}_n$.

(I) The 'ones' among the θ_i's are equidistantly distributed. Intuitively this may seem to be the case where the rules discussed are best motivated.

Calculations were carried out for proportions $p = 0(0.1)0.5$. (Because of complete symmetry we did not consider the values $p = 0.6(0.1)1$.) The values of n considered are $n = 50(50)200$. Calculations were made for the three sequences referred to above.

(II) All the 'zeros' among the θ_i's precede all the ones. This is the case most extremely opposed to I.

TABLE I

Sequences of type I

Average of losses incurred through use of rules T_n^*, T_n^{**} and $t_{\min}$

		$n = 50$			$n = 100$			$n = 150$			$n = 200$		
p	$R(p)$	T_n^*	T_n^{**}	$t_{\min}$	T_n^*	T_n^{**}	$t_{\min}$	T_n^*	T_n^{**}	$t_{\min}$	T_n^*	T_n^{**}	$t_{\min}$
0	0	.013	.013	.180	.010	.010	.163	.007	.007	.153	.005	.005	.158
0.1	.0691	.087	.093	.173	.073	.073	.160	.067	.067	.153	.067	.067	.162
0.2	.1121	.153	.153	.187	.127	.127	.160	.129	.129	.158	.117	.118	.160
0.3	.1387	.207	.213	.193	.200	.207	.187	.185	.186	.178	.175	.178	.180
0.4	.1538	.220	.207	.200	.180	.173	.163	.180	.176	.162	.175	.172	.165
0.5	.1587	.207	.200	.193	.187	.183	.173	.191	.189	.178	.195	.193	.187

TABLE II

Sequences of type II

Average of losses incurred through use of rules T_n^*, T_n^{**} and $t_{\min}$

		$n = 100$			$n = 200$		
p	$R(p)$	T_n^*	T_n^{**}	$t_{\min}$	T_n^*	T_n^{**}	$t_{\min}$
0	0	.010	.010	.163	.005	.005	.158
0.1	.0691	.083	.073	.163	.078	.078	.160
0.2	.1121	.137	.137	.173	.133	.130	.163
0.3	.1387	.160	.157	.177	.147	.145	.160
0.4	.1538	.157	.147	.170	.158	.157	.172
0.5	.1587	.133	.130	.157	.172	.167	.172
0.6	.1538	.133	.130	.153	.182	.180	.173
0.7	.1387	.123	.123	.163	.150	.147	.175
0.8	.1121	.093	.093	.160	.107	.105	.167
0.9	.0691	.043	.043	.150	.058	.060	.170
1	0	.023	.023	.163	.012	.013	.172

Calculations were carried out for $p = 0(0.1)1$, and for $n = 100$ and $n = 200$. Also here calculations were made for the three sequences.

For the sequence described in I and II both the rules T_n^* and T_n^{**} were considered and for each it was listed whenever one came to a wrong conclusion about θ_i. As a check on the sequence, and for purpose of comparison, the same was done also for the minimax rule $t_{\min}$.

For the sequences of type I, the actual losses incurred for the three sequences are given in Table III. The second column in the table lists the corresponding values of $R(\vartheta_n) = R(p)$. (Notice that in our example $R(p) = R(1 - p)$ for each $0 \leqq p \leqq 1$.) Notice that $R(\frac{1}{2}) = .1587$ is the (constant) risk of the minimax rule, i.e., $R(t_{\min}, 0) = R(t_{\min}, 1) = R(\frac{1}{2})$. The average of the corresponding columns for the three sequences of Table III is given in Table I below. It gives some indication about the values of $R(T_n^*, \boldsymbol{\theta}_n)$ and $R(T_n^{**}, \boldsymbol{\theta}_n)$

TABLE III

Sequences of type I

Losses incurred through use of rules T_n^*, T_n^{**} and $t_{\min}$

		Sequence 1			Sequence 2			Sequence 3		
p	$R(p)$	T_n^*	T_n^{**}	$t_{\min}$	T_n^*	T_n^{**}	$t_{\min}$	T_n^*	T_n^{**}	$t_{\min}$
					$n = 50$					
0	0	.04	.04	.18	0	0	.20	0	0	.16
0.1	.0691	.10	.10	.18	.06	.08	.22	.10	.08	.12
0.2	.1121	.18	.18	.18	.18	.18	.22	.10	.10	.16
0.3	.1387	.20	.22	.22	.20	.20	.18	.22	.22	.18
0.4	.1538	.22	.18	.22	.18	.18	.14	.26	.26	.24
0.5	.1587	.22	.22	.20	.20	.20	.18	.20	.18	.20
					$n = 100$					
0	0	.03	.03	.16	0	0	.15	0	0	.18
0.1	.0691	.06	.06	.15	.06	.07	.17	.10	.09	.16
0.2	.1121	.12	.12	.15	.14	.14	.15	.12	.12	.18
0.3	.1387	.19	.20	.18	.22	.22	.19	.19	.20	.19
0.4	.1538	.18	.16	.17	.16	.16	.13	.20	.20	.19
0.5	.1587	.21	.21	.17	.15	.15	.14	.20	.19	.21
					$n = 150$					
0	0	.020	.020	.133	0	0	.147	0	0	.180
0.1	.0691	.060	.060	.147	.047	.053	.153	.093	.087	.160
0.2	.1121	.127	.127	.133	.140	.140	.160	.120	.120	.180
0.3	.1387	.160	.167	.153	.207	.207	.187	.187	.193	.193
0.4	.1538	.160	.147	.147	.160	.160	.133	.220	.220	.207
0.5	.1587	.220	.220	.187	.160	.160	.153	.193	.187	.193
					$n = 200$					
0	0	.015	.015	.125	0	0	.150	0	0	.200
0.1	.0691	.060	.060	.130	.055	.060	.165	.085	.080	.190
0.2	.1121	.100	.105	.120	.125	.125	.155	.125	.125	.205
0.3	.1387	.150	.155	.145	.200	.200	.185	.175	.180	.210
0.4	.1538	.170	.160	.155	.160	.160	.135	.195	.195	.205
0.5	.1587	.185	.185	.165	.190	.190	.185	.210	.205	.210

for $n = 50(50)200$ and θ_n of type I. Obviously one should average over more than three sequences in order to get a reliable estimate of $R(T_n^*, \theta_n)$ and $R(T_n^{**}, \theta_n)$. It should be noticed that both in Table I and in Table III there seems to be a positive correlation between the excess of the losses incurred by use of T_n^* and T_n^{**} over $R(p)$, and the corresponding excess of the losses incurred by use of $t_{\min}$ over $R(\frac{1}{2})$. It should also be remarked that the instances in which T_n^* and T_n^{**} lead to different decisions are altogether extremely rare, but they are more frequent in the beginning of the sequences than later on.

Table I seems to indicate that for the case considered the gain in using any of the rules T_n^* and T_n^{**} for p near zero or one is large, whereas the excess of the risk over $R(\frac{1}{2})$ for p near $\frac{1}{2}$ is moderate. This is true already for moderate n.

For the sequences of type II the actual losses incurred for the three sequences are given in Table IV, and averages of the corresponding columns for the three sequences of Table IV are given in Table II.

For Table II similar remarks are valid as those made about Table I. Table II seems to indicate that the rules T_n^* and T_n^{**} are worth consideration, also in the

TABLE IV

Sequences of type II

Losses incurred through use of rules T_n^*, T_n^{**} and $t_{\min}$

		Sequence 1			Sequence 2			Sequence 3		
p	$R(p)$	T_n^*	T_n^{**}	$t_{\min}$	T_n^*	T_n^{**}	$t_{\min}$	T_n^*	T_n^{**}	$t_{\min}$
					$n = 100$					
0	0	.03	.03	.16	0	0	.15	0	0	.18
0.1	.0691	.09	.08	.17	.06	.05	.14	.10	.09	.18
0.2	.1121	.13	.13	.18	.14	.14	.15	.14	.14	.19
0.3	.1387	.17	.17	.19	.16	.16	.16	.15	.14	.18
0.4	.1538	.13	.13	.17	.17	.16	.17	.17	.15	.17
0.5	.1587	.12	.12	.16	.12	.12	.16	.16	.15	.15
0.6	.1538	.12	.12	.16	.12	.11	.15	.16	.16	.15
0.7	.1387	.12	.12	.17	.09	.09	.15	.16	.16	.17
0.8	.1121	.06	.06	.17	.12	.12	.15	.10	.10	.16
0.9	.0691	.02	.03	.16	.02	.01	.14	.09	.09	.15
1	0	.02	.02	.17	.03	.02	.15	.02	.03	.17
					$n = 200$					
0	0	.015	.015	.125	0	0	.150	0	0	.200
0.1	.0691	.080	.080	.125	.090	.090	.165	.065	.065	.190
0.2	.1121	.135	.135	.135	.160	.155	.175	.105	.100	.180
0.3	.1387	.120	.120	.135	.180	.180	.160	.140	.135	.185
0.4	.1538	.170	.170	.155	.160	.155	.165	.145	.145	.195
0.5	.1587	.170	.170	.165	.175	.170	.170	.170	.160	.180
0.6	.1538	.165	.165	.175	.195	.190	.170	.185	.185	.185
0.7	.1387	.120	.120	.170	.170	.165	.180	.160	.155	.175
0.8	.1121	.080	.080	.165	.110	.105	.170	.130	.130	.165
0.9	.0691	.030	.030	.170	.080	.085	.170	.065	.065	.170
1	0	.010	.010	.170	.015	.010	.170	.010	.020	.175

case where there is no reason to believe that the zeros and ones are well "mixed" in the sequence, (i.e., not only in situations corresponding to the usual Bayesian approach).

Acknowledgment. This research was initiated while the author was at Columbia University and benefited from many stimulating discussions with Professor Herbert Robbins, which she gratefully acknowledges. Thanks are due also to Mrs. Ronit Eisenbach for her careful assistance with the calculations.

REFERENCES

[1] BERKSON, J. (1953). A statistically precise and relatively simple method of estimating the bio-assay with quantal response, based on the logistic function. *J. Amer. Statist. Assoc.* **48** 565–599.

[2] BLACKWELL, D. (1954). Controlled random walks. *Proc. Internat. Congress Math.* Amsterdam, **3** 336–338.

[3] BLACKWELL, D. (1956). An analogue of the minimax theorem for vector payoffs. *Pacific J. Math.* **6** 1–8.

[4] DOOB, J. L. (1953). *Stochastic Processes.* Wiley, New York.

[5] HANNAN, J. F. and ROBBINS, H. (1955). Asymptotic solutions of the compound decision problem for two completely specified distributions. *Ann. Math. Statist.* **26** 37–51.

[6] LOÈVE, M. (1960). *Probability Theory,* (2nd ed.). Van Nostrand, New York.

[7] THE RAND CORPORATION (1955). *A Million Random Digits With 100,000 Normal Deviates.* Free Press, Glencoe.

[8] SAMUEL, E. (1963). Asymptotic solutions of the sequential compound decision problem. *Ann. Math. Statist.* **34** 1079–1094.