SEQUENTIAL COMPOUND ESTIMATORS

By Ester Samuel

The Hebrew University of Jerusalem

1. Summary. This paper deals with the sequential compound decision problem, when the component problem is an estimation problem. The aim is to find sequential compound rules which have Property (A) or (B), as described in Section 2. That is, one seeks rules the compound risk or loss of which do not exceed the value of the Bayes envelope functional at the "empirical distribution" of the parameters, by more than ϵ, for n sufficiently large. In Section 4 some general results for the estimation problem are obtained, which are applied in Sections 5 and 6 to several discrete and continuous distributions to obtain rules with Property (A) or (B). In Section 3 some results are obtained which hold for the sequential compound decision problem with any general component problem. These are applied for the estimation problem in the later sections.

2. Introduction. Consider a decision problem with its usual components: The space of states of nature Ω with elements θ; the action space $\mathfrak{A}$ with elements A; the real valued loss function $L(A, \theta) \geqq 0$ defined over $\mathfrak{A} \times \Omega$; and the observable random variable X with distribution P_θ when the state of nature is θ. We assume that for each θ P_θ is known. A (randomized) decision function ϕ is for each x (in the space $\mathfrak{X}$ of values of X) a distribution over $(\mathfrak{A}, \sigma_{\mathfrak{A}})$ (where σ_B denotes a properly defined σ-field of subsets of B) which is measurable in x. We denote by $R(\phi, \theta)$ the risk function of ϕ, i.e. the expected loss incurred by the use of ϕ, as a function of $\theta \varepsilon \Omega$. Let G be a (prior) distribution over (Ω, σ_Ω). $R(\phi, G)$ denotes the Bayes risk of ϕ, i.e.

$$R(\phi, G) = \int_\Omega R(\phi, \theta)\, dG(\theta). \tag{1}$$

Any ϕ minimizing (1) is called a Bayes rule with respect to G and will be denoted ϕ_G. The *Bayes envelope functional* is denoted $R(G)$ and for fixed G is the infimum value with respect to ϕ, of (1).

A compound decision problem arises when the same decision problem, called the component problem, occurs not only once, but n times. One thus has an (unknown) vector $\boldsymbol{\theta}_n = (\theta_1, \cdots, \theta_n)$, $\theta_i \varepsilon \Omega$, and a corresponding vector of random variables $\mathbf{X}_n = (X_1, \cdots, X_n)$ where the X_i's are independent, and X_i has distribution P_{θ_i}. If the problems occur sequentially, one may let the ith decision depend upon the observed value $\mathbf{x}_i = (x_1, \cdots, x_i)$ of $\mathbf{X}_i = (X_1, \cdots, X_i)$, rather than on x_i alone, $i = 1, \cdots, n$. This may seem irrational, since the X_i's are independent, and no relationship between the θ_i's is assumed. We shall see, however, that considering such rules may be worth while. We call $\boldsymbol{\phi} = (\phi_1, \phi_2, \cdots)$ a *strongly sequential compound decision rule if* ϕ_i is a measurable function of $\mathbf{x}_i$, which for each $\mathbf{x}_i$ is a distribution over $(\mathfrak{A}, \sigma_{\mathfrak{A}})$ by means of which

Received 27 April 1964; revised 22 December 1964.

the ith action is chosen. If the number n of components of the compound problem is unknown in advance, one may want to use such a rule $\boldsymbol{\phi}$. We shall denote the initial n-vector of $\boldsymbol{\phi}$ by $\boldsymbol{\phi}_n$. Let Ω^n denote the set of all $\boldsymbol{\theta}_n$, with $\theta_i \varepsilon \Omega$, and let Ω^∞ denote the set of all infinite sequences $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots)$ with $\theta_i \varepsilon \Omega$. For the compound rule $\boldsymbol{\phi}_n$ we shall let

$$R(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n) = n^{-1} \sum_{i=1}^n R(\phi_i, \boldsymbol{\theta}_i)$$

denote the compound risk of $\boldsymbol{\phi}_n$ as a function of $\boldsymbol{\theta}_n \varepsilon \Omega^n$. Here $R(\phi_i, \boldsymbol{\theta}_i)$ denotes the expected loss on the ith decision, and this will usually be a function of the initial i-vector $\boldsymbol{\theta}_i$ of $\boldsymbol{\theta}_n$, and not of θ_i alone. For every vector $\boldsymbol{\theta}_n$ we let $G_n = G_n(\boldsymbol{\theta}_n)$ denote its *empirical distribution.* G_n is the distribution over (Ω, σ_Ω) which assigns probability k/n to θ if $\theta_i = \theta$ for k among the n values of i, $i = 1, \cdots, n$. A strongly sequential rule $\boldsymbol{\phi}$ is said to have *Property* (A), if for every $\boldsymbol{\theta} \varepsilon \Omega^\infty$ and every $\epsilon > 0$ there exist $N(\boldsymbol{\theta}, \epsilon)$ such that for all $n > N(\boldsymbol{\theta}, \epsilon)$

$$\text{(A)} \qquad R(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n) - R(G_n) < \epsilon,$$

where $\boldsymbol{\theta}_n$ is the initial n-vector of $\boldsymbol{\theta}$, and G_n is its empirical distribution. $\boldsymbol{\phi}$ has *uniform Property* (A) if $N(\boldsymbol{\theta}, \epsilon) = N(\epsilon)$, i.e. if for every $\boldsymbol{\theta} \varepsilon \Omega^\infty$ (A) holds provided $n > N(\epsilon)$.

The (nonsequential) compound decision problem, where all observations are at hand before the individual decisions must be made, was introduced by Robbins, [7]. The sequential compound decision problem, has, to the best of the present authors knowledge, earlier been considered in detail only for the case where the component problem is that of testing a simple hypothesis against a simple alternative. See [9]. The rules exhibited there have uniform Property (A) for the problem considered. Hannan, in an abstract [2], considers the multiple decision problem where there are m possible distributions. Johns has recently announced, in an abstract [3], asymptotically optimal sequential compound decision rules for several parametric and nonparametric component problems, involving a finite number of actions. Our aim in the present paper is to exhibit, for several parametric families, rules $\boldsymbol{\phi}$ which have Property (A) when the component problem is an estimation problem, and the loss function is the customary squared error.

It is easily seen that if G_n is known in advance, and one uses the simple rule which decides on θ_i by the use of x_i and some version ϕ_{G_n} of the Bayes rule for the component problem, $i = 1, \cdots, n$, one incurs a risk $R(G_n)$ at every point $\boldsymbol{\theta}_n \varepsilon \Omega^n$ which has G_n as its empirical distribution. Clearly in that case the risk is of interest only for such $\boldsymbol{\theta}_n$. When G_n is not known in advance, one cannot use this rule. It seems that (A) would then be a goal worth aiming at. It is shown in [13] that no simple rule, using x_i only, to decide on θ_i, achieves (A).

It may be of interest to consider the (random) loss

$$L(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n) = n^{-1} \sum_{i=1}^n L(\phi_i, \boldsymbol{\theta}_i)$$

rather than its expected value $R(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n)$. It should be noticed that $L(\phi_i, \boldsymbol{\theta}_i)$ is a function of the random variable $\mathbf{X}_i$, and when ϕ_i is randomized, it is a function also of the random variable required to carry out randomization. For every fixed $\boldsymbol{\theta} \varepsilon \Omega^{\infty}$ we shall let $P_{(\infty)}$ and $P_{(n)}$ denote the product measures generated by P_{θ_i}, $i = 1, 2, \cdots$, and $i = 1, \cdots, n$, respectively, and we let $E_{(\infty)}$ and $E_{(n)}$ denote expectations with respect to the corresponding distributions. For a strongly sequential rule $\boldsymbol{\phi}$ we let $P_{(\infty)}(\boldsymbol{\phi})$ denote the measure generated by $\boldsymbol{\theta}$ and the auxiliary random variables involved in $\boldsymbol{\phi}$. Corresponding to (A) we say that $\boldsymbol{\phi}$ has *Property* (B) *almost everywhere* $P_{(\infty)}(\boldsymbol{\phi})$ if for every $\boldsymbol{\theta} \varepsilon \Omega^{\infty}$

$$\text{(B)} \qquad \limsup_{n\to\infty} [L(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n) - R(G_n)] \leqq 0 \quad \text{a.e.} \quad P_{(\infty)}(\boldsymbol{\phi}).$$

$\boldsymbol{\phi}$ is said to have *Property* (B) *in probability* if for every $\epsilon > 0$

$$\text{(B)} \qquad \lim_{n\to\infty} P[L(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n) - R(G_n) < \epsilon] = 1.$$

It should be noticed that whenever

$$(2) \qquad \sup_{\theta\varepsilon\Omega} \sup_{A\varepsilon\alpha} L(A, \theta) = C < \infty$$

Property (B) implies Property (A). In [11] it is shown that the rules considered in [9] for the hypothesis testing problem, have Property (B). In the present paper we shall establish Property (B) for sequential compound estimators for some discrete families of underlying distributions.

In the next section we obtain some results which are of interest for the general sequential compound decision problem. In Section 4 we derive some results for the general estimation problem, which are then applied in Sections 5 and 6 for several discrete and continuous families of distributions.

3. General results. The following lemma is true for any decision problem. A particular version of it was used in [9], (see Lemma 2 there). It gives some insight how (A) may possibly be achieved.

LEMMA 1. *Let* $\boldsymbol{\theta}_n = (\theta_1, \cdots, \theta_n)$ *be any fixed vector in* Ω^n, *and let* G_i *denote the empirical distribution of its initial i-vector,* $i = 1, \cdots, n$. *Let* ϕ_{G_i}, $i = 1, \cdots, n$ *denote any versions of the Bayes rules with respect to* G_i. *Then*

$$(3) \qquad n^{-1} \textstyle\sum_{i=1}^{n} R(\phi_{G_i}, \theta_i) \leqq R(G_n).$$

(The lemma states that if at the ith stage one plays a simple rule which is Bayes against the empirical distribution of the i first choices of Nature, $i = 1, \cdots, n$, one incurs a compound risk which at the specified point $\boldsymbol{\theta}_n$ does not exceed the value of the Bayes envelope functional at G_n.)

PROOF. For any ϕ of the component problem, one has by definition

$$R(\phi, G_i) = i^{-1} \textstyle\sum_{j=1}^{i} R(\phi, \theta_j).$$

Hence

$$iR(\phi, G_i) = \textstyle\sum_{j=1}^{i} R(\phi, \theta_j),$$

and thus $iR(\phi, G_i) - (i-1)R(\phi, G_{i-1}) = R(\phi, \theta_i)$ for $i = 1, \cdots, n$. Therefore

$$\begin{aligned}(4) \qquad \textstyle\sum_{i=1}^{n} R(\phi_{G_i}, \theta_i) &= \textstyle\sum_{i=1}^{n} [iR(\phi_{G_i}, G_i) - (i-1)R(\phi_{G_i}, G_{i-1})] \\ &= \textstyle\sum_{i=1}^{n-1} i[R(\phi_{G_i}, G_i) - R(\phi_{G_{i+1}}, G_i)] + nR(\phi_{G_n}, G_n).\end{aligned}$$

The first term in the right hand side of (4) is nonpositive by the definition of a Bayes rule, and the second term is $nR(G_n)$. Thus (3) follows.

It is shown in [11] that the inequality (3) can be strict, and actually that the difference between the right and left hand sides of (3) may tend to a positive limit.

From Lemma 1 we have the following

THEOREM 1. *Let* $\boldsymbol{\phi}$ *be a strongly sequential rule which for every* $\boldsymbol{\theta} \in \Omega^\infty$ *satisfies*

$$\lim_{i\to\infty} [R(\phi_i, \boldsymbol{\theta}_i) - R(\phi_{G_i}, \theta_i)] = 0 \tag{5}$$

where ϕ_{G_i} *is some Bayes rule with respect to the empirical distribution* G_i *of the initial i-vector* $\boldsymbol{\theta}_i$ *of* $\boldsymbol{\theta}$. *Then* $\boldsymbol{\phi}$ *has Property* (A). *If* (5) *holds uniformly in* $\boldsymbol{\theta} \in \Omega^\infty$, *and* (2) *holds, then* $\boldsymbol{\phi}$ *has uniform Property* (A).

PROOF. The proof follows immediately upon taking limit of the Cesaro means.

We restate, without proof, a martingale theorem, which will be used to establish Property (B). For a proof see Lemma 2 of [11], or [5] p. 387 E.

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

$$\sum_{n=1}^{\infty} \sigma_n^2/b_n^2 < \infty \tag{6}$$

then $Y_n/b_n \to 0$ a.e.

For any $\phi_i = \phi_i(\mathbf{x}_i, \quad)$ we denote by

$$R(\phi_i, \boldsymbol{\theta}_i \mid \mathbf{X}_{i-1}) = R(\phi_i, \theta_i \mid \mathbf{X}_{i-1})$$

the conditional risk incurred on the ith decision when the previous random variables $\mathbf{X}_{i-1}$, as well as the previous randomization variables, when required, are given. The ith randomization variable is assumed to be independent of the previous randomization variables, when $\mathbf{X}_i$ is given. To establish Property (B) we use the following

LEMMA 3. *Let* $L(A, \theta)$ *satisfy* (2). *Then for any* $\boldsymbol{\phi}$ *and every* $\boldsymbol{\theta} \in \Omega^\infty$, *as* $n \to \infty$,

$$[L(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n) - n^{-1} \sum_{i=1}^{n} R(\phi_i, \theta_i \mid \mathbf{X}_{i-1})] \to 0 \quad \text{a.e.} \quad P_{(\infty)}(\boldsymbol{\phi}). \tag{7}$$

PROOF. Let $\mathfrak{F}_n$ be the σ-field generated by $\mathbf{X}_n$ and the first n randomization variables, when required. Let $Y_n = \sum_{i=1}^{n} [L(\phi_i, \boldsymbol{\theta}_i) - R(\phi_i, \theta_i \mid \mathbf{X}_{i-1})]$. It follows by the definition that $\{Y_n\}$ is a martingale relative to $\{\mathfrak{F}_n\}$. Since $\text{Var}(Y_n - Y_{n-1}) \leqq C^2$, (6) holds with $b_n = n$, and Lemma 2 implies (7).

4. The estimation problem. Here $\Omega = \mathcal{A}$ is some interval of the real line. It is well known (see e.g. [4] p. 4-3) that when $L(A, \theta)$ is continuous and convex in A for each θ the class of nonrandomized rules is essentially complete. We shall consider only nonrandomized estimators, and shall let $\phi_i(\mathbf{x}_i)$ denote the value in $\mathcal{A}$ which is selected on the ith decision with probability one. We consider only $L(A, \theta)$ having the following property: For every $\epsilon > 0$ there exist $\delta = \delta(\epsilon) > 0$ such that

(8) $|L(A, \theta) - L(A^*, \theta)| < \epsilon$ for all $\theta \varepsilon \Omega$, provided $|A - A^*| < \delta$.

(8) holds whenever $\Omega = \mathfrak{A}$ is compact and $L(A, \theta)$ is continuous.

THEOREM 2. *Let $L(A, \theta)$ satisfy (2) and (8). Let $\boldsymbol{\phi}$ be a strongly sequential estimator where $\phi_i(\mathbf{x}_i)$ can be written as $\phi_i(\mathbf{x}_i, x)$ with $x = x_i$. Suppose that for every $\epsilon > 0$ there exists a set $S_\epsilon \subset \mathfrak{X}$ such that*

$$(9) \qquad P_\theta(S_\epsilon) > 1 - \epsilon \qquad \text{for every} \quad \theta \varepsilon \Omega,$$

and for every $\delta > 0$

$$(10) \qquad \lim_{i\to\infty} P_{(\infty)}[\sup_{x \varepsilon S_\epsilon} |\phi_i(\mathbf{X}_i, x) - \phi_{G_i}(x)| < \delta] = 1$$

for every $\boldsymbol{\theta} \varepsilon \Omega^\infty$, where G_i denotes the empirical distribution of the initial i-vector of $\boldsymbol{\theta}$. Then $\boldsymbol{\phi}$ has Property (A).

PROOF. By Theorem 1 it suffices to show that for given $\epsilon > 0$ and for every $i > I(\epsilon, \boldsymbol{\theta})$

$$(11) \qquad |R(\phi_i, \boldsymbol{\theta}_i) - R(\phi_{G_i}, \theta_i)| < \epsilon.$$

Let $\delta > 0$ be such that (8) holds for $\epsilon/3$. Let $\epsilon^* = \epsilon/3C$ and let $I(\epsilon, \boldsymbol{\theta})$ be such that for all $i > I(\epsilon, \boldsymbol{\theta})$

$$(12) \qquad P_{(i)}[\sup_{x \varepsilon S_{\epsilon^*}} |\phi_i(\mathbf{X}_i, x) - \phi_{G_i}(x)| < \delta] > 1 - \epsilon/3C.$$

Denote the event in the square bracket of (12) by B_i, and denote by $\bar{B}_i$ its complement. Let

$$W_i = |L(\phi_i(\mathbf{X}_i), \theta_i) - L(\phi_{G_i}(X_i), \theta_i)|.$$

Then $0 \leqq W_i \leqq C$, and it follows by (8) and the definitions that

$$\begin{aligned} |R(\phi_i, \boldsymbol{\theta}_i) - R(\phi_{G_i}, \theta_i)| &\leqq E_{(i)}[W_i] = E_{(i)}[W_i \mid X_i \varepsilon S_{\epsilon^*}, B_i] P_{(i)}[X_i \varepsilon S_{\epsilon^*}, B_i] \\ &\quad + E_{(i)}[W_i \mid X_i \not\varepsilon S_{\epsilon^*}, B_i] P_{(i)}[X_i \not\varepsilon S_{\epsilon^*}, B_i] \\ &\quad + E_{(i)}[W_i \mid \bar{B}_i] P_{(i)}[\bar{B}_i] < \epsilon/3 + C\epsilon^* + C(\epsilon/3C) = \epsilon, \end{aligned}$$

which establishes (11).

Corresponding to the above Theorem we have

THEOREM 3. *Let $L(A, \theta)$ satisfy (2) and (8). Let $\boldsymbol{\phi}$ be a strongly sequential estimator where $\phi_i(\mathbf{x}_i)$ can be written as $\phi_i(\mathbf{x}_{i-1}, x)$ with $x = x_i$. Suppose (9) holds for Ω, and suppose that as $i \to \infty$*

$$(13) \qquad \sup_{x \varepsilon S_\epsilon} |\phi_i(\mathbf{X}_{i-1}, x) - \phi_{G_i}(x)| \to 0 \quad \text{a.e.} \quad P_{(\infty)}$$

for every $\boldsymbol{\theta} \varepsilon \Omega^\infty$, where G_i denotes the empirical distribution of the initial i-vector of $\boldsymbol{\theta}$. Then $\boldsymbol{\phi}$ has Property (B) a.e. $P_{(\infty)}$. (Here $P_{(\infty)} = P_{(\infty)}(\boldsymbol{\phi})$).

PROOF. In view of Lemmas 1 and 3 it suffices to show that

$$(14) \qquad \lim_{n\to\infty} n^{-1} \textstyle\sum_{i=1}^n [R(\phi_i, \theta_i \mid \mathbf{X}_{i-1}) \dot{-} R(\phi_{G_i}, \theta_i)] = 0 \quad \text{a.e.} \quad P_{(\infty)}.$$

(14) follows if we show that for every given $\epsilon > 0$ there exists $I(\epsilon, \boldsymbol{\theta})$ such that

(15) $P_{(\infty)}[|R(\phi_i, \theta_i \mid \mathbf{X}_{i-1}) - R(\phi_{G_i}, \theta_i)| < \epsilon \quad \text{for all} \quad i > I(\epsilon, \boldsymbol{\theta})] > 1 - \epsilon.$

Let $\delta > 0$ be such that (8) holds for $\epsilon/2$ and let $\epsilon^* = \epsilon/2C$. Let $I(\epsilon, \boldsymbol{\theta})$ be such that

(16) $P_{(\infty)}[\sup_{x \varepsilon S_{\epsilon^*}} |\phi_i(\mathbf{X}_{i-1}, x) - \phi_{G_i}(x)| < \delta \quad \text{for all} \quad i > I(\epsilon, \boldsymbol{\theta})] > 1 - \epsilon.$

Denote by B_ϵ the event in the square bracket of (16). We shall show that for every $\mathbf{X}$ such that B_ϵ occurs, and $i > I(\epsilon, \boldsymbol{\theta})$

(17) $$|R(\phi_i, \theta_i \mid \mathbf{X}_{i-1}) - R(\phi_{G_i}, \theta_i)| < \epsilon$$

with, conditional on B_ϵ, probability one. Thus also the countable union over $i > I(\epsilon, \boldsymbol{\theta})$ of (17) holds with conditional probability one, and (15) follows from (16). Let

$$W_i = |L(\phi_i(\mathbf{X}_{i-1}, X), \theta_i) - L(\phi_{G_i}(X), \theta_i)|$$

where X is distributed according to P_{θ_i} and is independent of $\mathbf{X}_{i-1}$. Then $0 \leqq W_i \leqq C$ and the left hand side of (17) does not exceed $E_{\theta_i}[W_i \mid \mathbf{X}_{i-1}]$, i.e. the random variable obtained from W_i when integrating with respect to X distributed according to P_{θ_i}. But if $\mathbf{X}_{i-1}$ is the initial vector of $\mathbf{X}$ satisfying B_ϵ, and $i > I(\epsilon, \boldsymbol{\theta})$ then by (16), (8) and (9)

$$E_{\theta_i}[W_i \mid \mathbf{X}_{i-1}] = E_{\theta_i}[W_i \mid \mathbf{X}_{i-1}, X \varepsilon S_{\epsilon^*}] P_{(i)}[X \varepsilon S_{\epsilon^*} \mid \mathbf{X}_{i-1}]$$
$$+ E_{\theta_i}[W_i \mid \mathbf{X}_{i-1}, X \not\varepsilon S_{\epsilon^*}] P_{(i)}[X \not\varepsilon S_{\epsilon^*} \mid \mathbf{X}_{i-1}] \leqq \epsilon/2 + C\epsilon^* = \epsilon.$$

Thus (17) follows, and the proof is complete.

We shall also have occasion to use a variant Theorem 3, which is proved similarly.

COROLLARY 1. *Let $L(A, \theta)$ satisfy* (2) *and* (8) *and let $\boldsymbol{\phi}$ be a strongly sequential estimator, where $\phi_i(\mathbf{x}_i)$ can be written as $\phi_i(\mathbf{x}_{i-1}, x)$ with x depending on x_i only. Suppose*

$$\sup_x |\phi_i(\mathbf{X}_{i-1}, x) - \psi_{G_i}(x)| \to 0 \quad \text{a.e.} \quad P_{(\infty)},$$

for every $\boldsymbol{\theta} \varepsilon \Omega^\infty$, where ψ_{G_i} is an estimator defined for each G_i, and satisfying, for every n and $\boldsymbol{\theta}_n \varepsilon \Omega^n$

$$n^{-1} \sum_{i=1}^n R(\psi_{G_i}, \theta_i) \leqq R^*(G_n).$$

Then

$$P_{(\infty)}[\limsup_{n\to\infty} (L(\boldsymbol{\phi}_n, \theta_n) - R^*(G_n)) \leqq 0] = 1.$$

5. Sequential compound estimators for some discrete distributions. In this and the next section we apply the results obtained in the previous section. We consider the conventional loss function $L(A, \theta) = k(A - \theta)^2$, $k > 0$, and for convenience we let $k = 1$. For this loss function (2) and (8) hold provided only $\Omega = \mathcal{A}$ is bounded. Here

(18) $$\phi_G(x) = E_G(\theta \mid x),$$

i.e., the conditional expectation of θ, under G, given $X = x$. Consider the family of nonnegative, integer valued random variables with distributions

$$(19) \qquad P[X = x] = p_\theta(x) = \theta^x h(\theta) g(x) \qquad \text{for} \quad x = 0, 1, \cdots,$$

where θ ranges in some subinterval of $(0, \infty)$. Special cases of (19) are the Poisson, geometric and negative binomial distributions. Consider the function

$$(20) \quad f(v, w) = cv/w \qquad \text{defined for} \quad 0 \leqq v < \infty, \quad 0 < a \leqq w < \infty,$$

where c is some constant. Since it is uniformly continuous in its range of definition, it follows that if $\{v_i^{(j)}\}$ and $\{w_i^{(j)}\}$, $j = 1, 2$, are sequences of random variables, which with probability one are in the range of definition of f, and if $|v_i^{(1)} - v_i^{(2)}| \to 0$ and $|w_i^{(1)} - w_i^{(2)}| \to 0$ with probability one or in probability, then

$$(21) \qquad |f(v_i^{(1)}, w_i^{(1)}) - f(v_i^{(2)}, w_i^{(2)})| \to 0$$

with probability one or in probability, respectively.

Whenever necessary we shall curtail the range of θ to be a closed interval $\Omega = \{\theta: \alpha \leqq \theta \leqq \beta\}$, and define $\inf_{\theta \varepsilon \Omega} p_\theta(x) = m(x)$. Assume $m(x) > 0$ for $x = 0, 1, \cdots$. Let $\boldsymbol{\theta} \,\varepsilon\, \Omega^\infty$ be fixed, and for $x = 0, 1, \cdots$, let $Y_j(x)$ be 1 or 0 according as $X_j = x$ or $X_j \neq x$, $j = 1, 2, \cdots$. Let

$$(22) \qquad p_{G_i}(x) = i^{-1} \textstyle\sum_{j=1}^i p_{\theta_j}(x), \qquad p_i^*(x) = i^{-1} \textstyle\sum_{j=1}^i Y_j(x)$$

and let $p_i(x)$ equal $m(x)$ or $p_i^*(x)$ according as $p_i^*(x)$ is less than $m(x)$, or is greater or equal to $m(x)$, respectively. It then follows by the strong law of large numbers that for $x = 0, 1, \cdots$, $|p_i(x) - p_{G_i}(x)| \to 0$ a.e. $P_{(\infty)}$. Since $|p_{G_{i-1}}(x) - p_{G_i}(x)| \leqq 1/i$ it follows that also $|p_{i-1}(x) - p_{G_i}(x)| \to 0$ a.e. $P_{(\infty)}$ and hence also

$$(23) \quad P_{(\infty)}[|p_{i-1}(x) - p_{G_i}(x)| \to 0 \quad \text{and} \quad |p_{i-1}(x+1) - p_{G_i}(x+1)| \to 0] = 1.$$

For (19) (18) becomes $\phi_{G_i}(x) = [g(x)/g(x+1)][p_{G_i}(x+1)/p_{G_i}(x)]$. Let

$$\phi_i(\mathbf{X}_{i-1}, x) = [g(x)/g(x+1)][p_{i-1}(x+1)/p_{i-1}(x)].$$

Then $\phi_{G_i}(x) = f(p_{G_i}(x+1), p_{G_i}(x))$ and $\phi_i(\mathbf{X}_{i-1}, x) = f(p_{i-1}(x+1), p_{i-1}(x))$ with $f(v, w)$ defined in (20) with $c = g(x)/g(x+1)$ and $a = m(x)$. Thus it follows from (23) and (21) that for every $x = 0, 1, \cdots$

$$(24) \qquad |\phi_i(X_{i-1}, x) - \phi_{G_i}(x)| \to 0 \text{ a.e. } P_{(\infty)}.$$

We want to apply Theorem 3. Notice that by the definition (19), $h(\theta)$ is decreasing. Let $\epsilon > 0$ be given, and let $N(\epsilon)$ be an integer such that

$$\textstyle\sum_{x=N(\epsilon)}^\infty \beta^x g(x) < \epsilon/h(\alpha).$$

(Such an integer exists, since $\sum_{x=0}^\infty \beta^x g(x) = 1/h(\beta)$.) Then for every $\theta \,\varepsilon\, \Omega$

$$\textstyle\sum_{x=N(\epsilon)}^\infty p_\theta(x) < \sum_{x=N(\epsilon)}^\infty h(\alpha)\beta^x g(x) < \epsilon.$$

Let $S_\epsilon = \{0, 1, \cdots, N(\epsilon) - 1\}$. Then S_ϵ is a *finite* set satisfying (9), and from

(24) it follows that (13) holds. Thus all assumptions of Theorem 3 hold and we have proved

THEOREM 4. *For estimating* $\theta \varepsilon \{\theta: \alpha \leqq \theta \leqq \beta\}$ *for a family* (19) *with loss function* $(A - \theta)^2$, *the strongly sequential rule* $\boldsymbol{\phi}$ *with*

$$\phi_i(\mathbf{x}_i) = [g(x_i)/g(x_i + 1)][p_{i-1}(x_i + 1)/p_{i-1}(x_i)] \tag{25}$$

has Property (B) a.e. $P_{(\infty)}$ (*and hence also has Property* (A)).

REMARKS. (1) $\phi_i(\mathbf{x}_i)$ defined in (25) may take values outside $\mathfrak{A}$, and hence may not be a permissible estimator for the problem considered. Since ϕ_{G_i} always takes values in $\mathfrak{A}$, and because of (24), we may, however, curtail $\phi_i(\mathbf{x}_i)$ of (25) to make it belong to $\mathfrak{A}$ with probability one, and the theorem remains unchanged.

(2) It may seem more natural to define a rule corresponding to (25), but with p_{i-1} replaced by p_i, which is at hand at the ith stage. It is easily seen that the rule obtained in this way satisfies the conditions of Theorem 2, and thus has Property (A). This rule is considered by Robbins in [8] for the empirical Bayes problem for the geometric and Poisson distributions. The author believes that this rule also has Property (B) but a proof is not quite straightforward.

In connection with the above, it is of interest to note that quite similarly to the proof of Lemma 1 one may show that for the general decision problem and every $\boldsymbol{\theta}_n \varepsilon \Omega^n$

$$n^{-1}\sum_{i=1}^{n} R(\phi_{G_{i-1}}, \theta_i) \geqq R(G_n).$$

(Compare (20) of [11]). Hence Theorem 4 may be somewhat surprising.

For the *binomial distribution*

$$P(X = x) = p_{\theta,r}(x) = \binom{r}{x}\theta^x(1 - \theta)^{r-x},$$

$x = 0, \cdots, r$, no rule has Property (B). Still, if $r > 1$ something can be achieved. Let $\Omega = \{\theta: 0 < \alpha \leqq \theta \leqq \beta < 1\}$. Let $p_{G_i,r}(x)$, $p^*_{i,r}(x)$ and $p_{i,r}(x)$ be defined corresponding to (22). Then (18) implies

$$\phi_{G_i,r}(x) = [(x + 1)/(r + 1)][p_{G_i,r+1}(x + 1)/p_{G_i,r}(x)].$$

Consider X_j as the number of successes in r independent Bernoulli trials, with probability θ_j of success, and assume that the number of successes in the $(r - 1)$ first trials, $X_j^{(r-1)}$, is known. Denote its observed value by $x_j^{(r-1)}$. Define $p^*_{i,r-1}(x)$ and $p_{i,r-1}(x)$ by means of $X_j^{(r-1)}, j = 1, 2, \cdots$, corresponding to (22). Then

COROLLARY 2. *For estimating* $\theta \varepsilon \{\theta: 0 < \alpha \leqq \theta \leqq \beta < 1\}$ *for the family of binomial distributions with parameter* r *and loss function* $(A - \theta)^2$, *the strongly sequential rule* $\boldsymbol{\phi}$ *with*

$$\phi_i(\mathbf{x}_i) = [(x_i^{(r-1)} + 1)/r][p_{i-1,r}(x_i^{(r-1)} + 1)/p_{i-1,r-1}(x_i^{(r-1)})] \tag{26}$$

satisfies $P_{(\infty)}[\limsup_{n\to\infty} [L(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n) - R_{r-1}(G_n)] \leqq 0] = 1$, *where* R_{r-1} *denotes the Bayes envelope functional for the binomial distribution with parameter* $r - 1$.

PROOF. This follows from Corollary 1 since

$$\max_{x=0,\cdots,r-1} |\phi_i(\mathbf{X}_{i-1}, x) - \phi_{G_i,r-1}(x)| \to 0 \text{ a.e. } P_{(\infty)},$$

where $\phi_i(\mathbf{X}_{i-1}, x)$ is defined corresponding to (26).

6. Sequential compound estimators for some continuous distributions. In this section we assume that P_θ satisfies (9) and has density $f_\theta(x)$ with respect to Lebesgue measure. Further assume that for every $\epsilon > 0$, $f_\theta(x)$, $\theta \in \Omega$, are uniformly equicontinuous in S_ϵ i.e. for every $\eta > 0$ there exists $\delta > 0$ such that $|f_\theta(x) - f_\theta(x+h)| < \eta$ for every $\theta \in \Omega$, if $x \in S_\epsilon$ and $|h| < \delta$. Theorem 5, which follows, is based on Lemma 4, which also is of some independent interest. They are stated without proofs, as their proofs are quite similar to the proofs of Theorem 1A and 3A of Parzen [6]. Related detailed proofs can be found in [12], p. 49–53.

LEMMA 4. *Suppose $f_j(x)$, $j = 1, 2, \cdots$, are uniformly bounded and uniformly equicontinuous functions in the domain S_ϵ. Let $f_{(n)}(x) = n^{-1}\sum_{j=1}^n f_j(x)$. Suppose further that for $j = 1, 2, \cdots$ $\int_{-\infty}^{\infty} |f_j(x)|\, dx \leqq c < \infty$. Let $K(y)$ be a Borel function satisfying*

$$\sup_{-\infty<y<\infty} |K(y)| < \infty \tag{27}$$

$$\int_{-\infty}^{\infty} |K(y)|\, dy < \infty \tag{28}$$

$$\lim_{y\to\infty} |yK(y)| = 0. \tag{29}$$

Let $h(n)$ be a sequence of positive constants such that

$$\lim_{n\to\infty} h(n) = 0. \tag{30}$$

Define

$$f_n^*(x) = [1/nh(n)]\sum_{j=1}^n \int_{-\infty}^{\infty} K(y/h(n)) f_j(x-y)\, dy.$$

Then

$$\lim_{n\to\infty} \sup_{x\in S_\epsilon} |f_n^*(x) - f_{(n)}(x) \int_{-\infty}^{\infty} K(y)\, dy| = 0.$$

Let $\boldsymbol{\theta} \in \Omega^\infty$ be given, and let $f_{G_n}(x) = n^{-1}\sum_{j=1}^n f_{\theta_j}(x)$. In Theorem 5 we apply Lemma 4 with $f_j = f_{\theta_j}$. Let

$$f_n(\mathbf{X}_n, x) = [1/nh(n)]\sum_{j=1}^n K[(x - X_j)/h(n)]. \tag{31}$$

Then $E_{(\infty)} f_n(\mathbf{X}_n, x) = f_n^*(x)$. We have

THEOREM 5. *Suppose $f_\theta(x)$, $\theta \in \Omega$, are uniformly bounded and uniformly equicontinuous functions in the domain S_ϵ, and let $f_n(\mathbf{X}_n, x)$ be defined by (31) where $K(y)$ is an even function satisfying (27), (28), (29) and $\int_{-\infty}^{\infty} K(y)\, dy = 1$, and has a Fourier transform $\int_{-\infty}^{\infty} e^{-iuy} K(y)\, dy$ which is absolutely integrable. Let $h(n)$ satisfy (30) and $\lim_{n\to\infty} nh^2(n) = \infty$. Then for every $\epsilon > 0$*

$$\lim_{n\to\infty} P_{(\infty)}[\sup_{x\in S_\epsilon} |f_n(\mathbf{X}_n, x) - f_{G_n}(x)| < \epsilon] = 1.$$

Examples of functions $K(y)$ satisfying all the imposed conditions are given in Table 1 (except the first line) of [6]. Lemma 4 and Theorem 5 are of particular interest when the functions are uniformly equicontinuous over the whole real line.

Corresponding to (19) we shall consider densities of the form

$$(32) \qquad f_\theta(x) = \theta^x h(\theta) g(x) \qquad \text{for } a < x < \infty$$
$$= 0 \qquad \text{otherwise,}$$

where a is some constant (possibly $-\infty$) and h and g are continuous functions. When curtailing the parameter space, (32) satisfy (9) and are uniformly equicontinuous in S_ϵ for every $\epsilon > 0$. Though (32) seems a quite particular family of densities, it is shown in Section 6 of [10] that many families of distributions, including the exponential, normal and gamma families, can by simple transformations be given the form (32). Let $L(A, \theta) = (A - \theta)^2$. Then from (18)

$$\phi_{G_n}(x) = [g(x)/g(x+1)][f_{G_n}(x+1)/f_{G_n}(x)] \qquad \text{for } x > a$$

(and need not be defined for $x \leqq a$). Suppose the distributions are such that (9) holds, and

$$(33) \qquad \sup_{x \varepsilon S_\epsilon} [g(x)/g(x+1)] = D_\epsilon < \infty$$

and

$$(34) \qquad \inf_{x \varepsilon S_\epsilon} \inf_{\theta \varepsilon \Omega} f_\theta(x) > 0.$$

Suppose $\Omega = \mathfrak{A}$ is compact and for $x > a$ let

$$(35) \qquad \phi_n(\mathbf{X}_n, x) = [g(x)/g(x+1)][f_n(\mathbf{X}_n, x+1)/f_n(\mathbf{X}_n, x)]$$

if the right hand side of (35) is in Ω, and let $\phi_n(\mathbf{X}_n, {}_x)$ equal the nearest value in Ω, otherwise. Then by Theorem 5 and arguments similar to those of Section 5 it follows that (10) holds for every $\delta > 0$.

From Theorem 2 we therefore have

THEOREM 6. *For estimating $\theta \varepsilon \{\theta: \alpha \leqq \theta \leqq \beta\}$ for a family with density* (32), *satisfying* (9), (33) *and* (34) *with loss function* $(A - \theta)^2$, *the strongly sequential rule* ϕ *with*

$$\phi_n(\mathbf{x}_n) = [g(x_n)/g(x_n+1)][f_n(\mathbf{x}_n, x_n+1)/f_n(\mathbf{x}_n, x_n)]$$

when the right hand side is between α and β, and $\phi_n(\mathbf{x}_n)$ is equal α or β according as the right hand side is closer to α or β, otherwise, has Property (A).

Notice that application of the theory usually requires curtailing the natural parameter space of the family P_θ of distributions, to obtain Ω compact, with uniformly equicontinuous $f_\theta(x)$. A more detailed application for the case where P_θ is $N(\theta, \sigma^2)$ is given in [12].

REFERENCES

[1] DOOB, J. L. (1953). *Stochastic Processes*, Wiley, New York.
[2] HANNAN, J. F. (1956). The dynamic statistical decision problem when the component problem involves a finite number, m, of distributions. (Abstract) *Ann. Math. Statist.* **27** 212.

[3] Johns, M. V. Jr. (1963). The parametric and nonparametric compound decision problems in the sequence case. (Abstract) *Ann. Math. Statists.* **34** 1620–1621.

[4] Lehmann, E. L. (1950). *Notes on the Theory of Estimation*. Univ. of California, Berkeley.

[5] Loève, M. (1960). *Probability Theory*, 2nd ed. Van Nostrand, Princeton.

[6] Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33** 1065–1076.

[7] Robbins, H. (1951). Asymptotic subminimax solutions of compound statistical decision problems. *Proc. Second Berkeley Symp. Math. Statist. Prob.* 131–148. Univ. of California Press.

[8] Robbins, H. (1955). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Prob.* 157–163. Univ. of California Press.

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

[10] Samuel, E. (1963). An empirical Bayes approach to the testing of certain parametric hypotheses. *Ann. Math. Statist.* **34** 1370–1385.

[11] Samuel, E. (1964). Convergence of the losses of certain decision rules for the sequential compound decision problem. *Ann. Math. Statist.* **35** 1606–1621.

[12] Samuel, E. (1964). On the sequential compound decision problem in the fixed case and the opponent case. Unpublished.

[13] Samuel, E. (1965). On simple rules for the compound decision problem. *J. Roy. Statist. Soc. Ser. B* **27**.