THE COMPOUND DECISION PROBLEM IN THE OPPONENT CASE

BY
ESTER SAMUEL

ABSTRACT

The compound decision problem is considered under the assumption that Nature plays the role of an opponent. The empirical Bayes approach is a particular case of this approach. It is shown that in many instances sequential compound rules with good asymptotic properties for the fixed case have similar good properties in the opponent case.

1. Introduction and Summary. Consider a general statistical decision problem, with its usual components: The parameter space Ω with elements θ, the action space $\mathfrak{A}$ with elements A and the loss function $L(A,\theta) \geqq 0$ defined on $\mathfrak{A} \times \Omega$. The action chosen depends upon the value $x \in \mathscr{X}$ of an observable random variable X which is distributed according to the known distribution P_θ, when the true parameter is θ. A decision function ϕ is a function measurable in x, which for each fixed x is a distribution over $(\mathfrak{A}, \sigma_{\mathfrak{A}})$, (where σ_B denotes a suitable σ-field of subsets of B). The risk function $R(\phi,\theta)$ of ϕ is the expected loss incurred by use of ϕ, considered as a function of $\theta \in \Omega$.

In the Bayesian approach θ is considered as a realization of a random variable Λ with (a priori) distribution G over (Ω, σ_Ω). The Bayes risk of ϕ with respect to G is

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

(1) is minimized (for fixed G) by any Bayes rule ϕ_G for which

$$\inf_\phi R(\phi, G) = R(\phi_G, G) = R(G). \tag{2}$$

$R(G)$ is called the Bayes envelope functional. We assume that there always exists a minimizing ϕ_G.

The sequential compound problem has been of interest lately. (See [3], [9], [4], [11] and [12]). Such a problem arises when one is confronted with the same statistical decision problem, called the component problem, not only once, but

Received Feb. 17, 1965

sequentially n times. One therefore has a vector $\boldsymbol{\theta}_n = (\theta_1, \cdots, \theta_n)$ of (unknown) parameters, $\theta_i \in \Omega$, and a corresponding vector $X_n = (X_1, \cdots, X_n)$ of random variables, where the X_i's are assumed independent, and X_i has distribution P_{θ_i}. In this situation one can use a sequential compound decision rule $\boldsymbol{\phi}_n = (\phi_1, \cdots, \phi_n)$ where ϕ_i depends on X_i, and for each x_i is a distribution over $(\mathfrak{A}, \sigma_{\mathfrak{A}})$ by means of which the i-th action is chosen, $i = 1, \cdots, n$. (Generally ϕ_i may depend on previous actions, i.e., there may be dependence between the randomizations.)

We denote the random loss incurred on the i-th decision by $L(\phi_i, \boldsymbol{\theta}_i)$, and its average over the n decisions by $L(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n)$. We shall define this average to be the compound loss. The corresponding expected values are $R(\phi_i, \boldsymbol{\theta}_i)$ and $R(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n)$ respectively, the latter being the compound risk of $\boldsymbol{\phi}_n$.

Let $\boldsymbol{\phi} = (\phi_1, \phi_2, \cdots)$ be a strongly sequential compound rule (i.e. a sequene where ϕ_i is as described above but does not depend on n), and denote by $\boldsymbol{\phi}_n$ its initial n-vector.

Since we shall be interested in asymptotic properties of sequential compound rules, we shall for convenience limit our attention to strongly sequential rules. Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots)$, $\theta_i \in \Omega$, be a sequence of parameter values, with $\boldsymbol{\theta}_n$ as initial n-vector. Let G_n denote the empirical distribution of $\boldsymbol{\theta}_n$, i.e. the distribution over (Ω, σ_Ω) which assigns to θ the probability s/n if $\theta_i = \theta$ for s values of i, $i = 1, \cdots, n$. If G_n were known in advance, one could, by using for every i a rule ϕ_{G_n} for the component problem with the observation x_i to decide on θ_i, achieve $R(G_n)$ defined in (2), as the value of the compound risk at every $\boldsymbol{\theta}_n$ which has G_n as its empirical distribution. G_n is, however, usually unknown, so this simple rule cannot be used. One is therefore led to consider rules $\boldsymbol{\phi}$ with good asymptotic properties, in the following sense:

$\boldsymbol{\phi}$ is said to have property (A) if for every $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots)$ and every $\varepsilon > 0$ there exists a $N(\boldsymbol{\theta}, \varepsilon)$ such that for all $n > N(\boldsymbol{\theta}, \varepsilon)$

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

If $N(\boldsymbol{\theta}, \varepsilon)$ can be chosen to be independent of $\boldsymbol{\theta}$, $\boldsymbol{\phi}$ is said to have uniform property (A).

$\boldsymbol{\phi}$ is said to have property (B), almost everywhere if

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

$\boldsymbol{\phi}$ is said to have property (B) in probability if for $\varepsilon > 0$

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

For the problem of testing a simple hypothesis against a simple alternative rules having uniform property (A) are exhibited in [9]. In [11] they are shown to have property (B). For some estimation problems, rules with properties (A) or (B) are exhibited in [12].

In the present paper we change the theoretical setup, and consider what we call the opponent case. Rather than considering a fixed but unknown sequence $\boldsymbol{\theta}$ we consider Nature as an active opponent of the statistician. We assume that Nature has a *strategy* γ by means of which the parameters $(\theta_1, \theta_2 \cdots)$ are chosen. Let Λ_i denote the random variable whose observed value θ_i determines the distribution P_{θ_i} of X_i. We shall let $\Lambda_i = \Lambda_i(\gamma) = \Lambda_i(\boldsymbol{\phi}, \boldsymbol{X}_{i-1}, \boldsymbol{\Lambda}_{i-1}, \boldsymbol{Z}_{i-1})$. Here Z_j denotes the auxiliary random variable used to select the j-th action (needed when the j-th decision is randomized). Thus we let Λ_i depend upon all previous random variables, and we may even let it depend upon the functional form of $\boldsymbol{\phi}$. Clearly under these assumptions the X_i's are no longer independent. The conditional distribution of X_i, given $\Lambda_i = \theta_i$, is still P_{θ_i}. We shall call γ a strategy only if it together with $\boldsymbol{\phi}$ and the family P_θ, $\theta \in \Omega$, generates a probability measure on the measurable space of interest. This probability measure will be denoted by P_γ, expectation with respect to it by E_γ, and we assume that all conditional expectations considered in the sequel are meaningful. We shall maintain the notation introduced earlier, but replace θ_i by the random Λ_i, and we let K_n denote the (random) empirical distribution of $\boldsymbol{\Lambda}_n$. Corresponding to the earlier definitions we say that $\boldsymbol{\phi}$ has property (A_0) if for every strategy γ and every $\varepsilon > 0$ there exists $N(\gamma, \varepsilon)$ such that for all $n > N(\gamma, \varepsilon)$

$$(\mathrm{A}_0) \qquad E_\gamma[L(\boldsymbol{\phi}, \boldsymbol{\Lambda}_n) - R(K_n)] < \varepsilon$$

If $N(\gamma, \varepsilon)$ can be chosen to be independent of γ $\boldsymbol{\phi}$ is said to have uniform property (A_0).

$\boldsymbol{\phi}$ is said to have property (B_0) almost everywhere P_γ, if

$$(\mathrm{B}_0) \qquad \lim_{n \to \infty} \sup[L(\boldsymbol{\phi}_n, \boldsymbol{\Lambda}_n) - R(K_n)] = 0 \qquad \text{a.e. } P_\gamma.$$

$\boldsymbol{\phi}$ is said to have property (B_0) in probability if for every $\varepsilon > 0$

$$(\mathrm{B}_0) \qquad \lim_{n \to \infty} P_\gamma(L(\boldsymbol{\phi}_n, \boldsymbol{\Lambda}_n) - R(K_n) < \varepsilon) = 1\,.$$

Clearly (A_0) and (B_0) are stronger properties than (A) and (B), respectively, since (A) and (B) follow from (A_0) and (B_0) if we consider only such strategies γ which with probability one select a specified sequence $\boldsymbol{\theta}$. Notice that whenever (5) below holds, property (B_0) implies property (A_0).

One may object to the consideration of the opponent case in statistics, since it seems unrealistic to believe that Nature chooses the parameter values dependent upon the previous decisions of the statistician. From this point of view the opponent case seems more justified in a game-theoretic setup. It has been dealt with for nonstatistical games in [1] and [2] by Blackwell, who exihibits rules having a property corresponding to (B_0). We remark that a similar criticism is equally valid regarding the minimax attitude, which nevertheless has received much attention in statistics. Our aim in the present paper is to show that the rules considered earlier for the sequential compound situation actually have properties (A_0) or (B_0), respectively.

One particular example of the opponent case is the empirical Bayes approach. This approach is certainly motivated in statistics. (See e.g. [7], [8] and [10].) In the next Section we proceed to prove that every ϕ with property (A) is an asymptotically optimal empirical Bayes rule. In the Section thereafter we show that an optimal strategy for the opponent is the repeated maximin strategy of the component problem. In Section 4 some theoretical results are stated, which are used in Sections 5 and 6 when dealing with the problem of testing simple hypotheses, and the estimation problem, respectively.

2. **The empirical Bayes approach.** In this approach it is assumed that Λ_i are independent, identically distributed random variables, with unknown apriori distribution G over (Ω, σ_Ω). Let G^n denote the product distribution over the n-fold Cartesian product $(\Omega^n, \sigma_{\Omega^n})$, generated by G. Corresponding to (1), the Bayes risk on the n-th decision will be denoted $R(\phi_n, G^n)$, and is the expectation with respect to G^n, of $R(\phi_n, \Lambda_n)$. ϕ is called an *asymptotically optimal empirical Bayes rule* if for every (apriori) distribution G

$$\lim_{n\to\infty} R(\phi_n, G^n) = R(G). \tag{3}$$

(3) implies

$$\lim_{n\to\infty} R(\boldsymbol{\phi}_n, G^n) = R(G). \tag{4}$$

It follows from [13], Theorem 1 that always $R(\phi_n, G^n) \geqq R(G)$, and hence (4) implies (3), and asymptotic optimality could be defined either by (3) or (4). It is well known that there exist component problems for which (3) cannot be achieved. See [10] and [8]. We shall need the following lemma, which is quite well-known.

LEMMA 1. *The functional $R(G)$ is concave.*

Proof. Let $0 \leqq p \leqq 1$ and $q = 1 - p$. By (1) and (2)

$$\begin{aligned} R(pG_1 + qG_2) &= R(\phi_{pG_1+qG_2}, pG_1 + qG_2) \\ &= pR(\phi_{pG_1+qG_2}, G_1) + qR(\phi_{pG_1+qG_2}, G_2) \\ &\geqq pR(\phi_{G_1}, G_1) + qR(\phi_{G_2}, G_2) = pR(G_1) + qR(G_2). \end{aligned}$$

THEOREM 1. *If*

$$\sup_{A\in\mathfrak{A}}\sup_{\theta\in\Omega} L(A,\theta) = M < \infty \tag{5}$$

then every ϕ with property (A) *is an asymptotically optimal empirical Bayes rule.*

Proof. Let γ_G denote the strategy by which the Λ_i are independent, identically distributed according to G. Then for every real t $E_{\gamma_G} K_n(t) = G(t)$, and it follows by Lemma 1 and by use of the usual approximation by step functions that

$$E_{\gamma_G} R(K_n) \leqq R(G). \tag{6}$$

Now whatever be $\boldsymbol{\theta}$, we have that conditionally on $\boldsymbol{\Lambda} = \boldsymbol{\theta}$ the X_i's are independent. Hence from property (A) it follows that conditionally on $\boldsymbol{\Lambda} = \boldsymbol{\theta}$

$$\limsup_{n\to\infty} [R(\boldsymbol{\phi}_n, \boldsymbol{\theta}_n) - R(G_n)] \leqq 0$$

and hence by the Fatou-Lebesgue theorem, since by (5) L and R are bounded functions,

$$\limsup_{n\to\infty} E_{\gamma_G}[R(\boldsymbol{\phi}_n, \boldsymbol{\Lambda}_n) - R(K_n)] \leqq 0. \tag{7}$$

(6) and (7) imply

$$\limsup_{n\to\infty} R(\boldsymbol{\phi}_n, G^n) \leqq R(G), \tag{8}$$

whereas from Theorem 1 of [13] we have

$$\inf_{\boldsymbol{\phi}_n} R(\boldsymbol{\phi}_n, G^n) = R(G).$$

Hence the theorem follows.

We remark that if $\boldsymbol{\phi}$ has property (A) and the Λ_i's are distributed independently of the X's, and Λ_i has (marginal) distribution G, but the Λ_i's need not be mutually independent, then (8) is still correct, but one can easily obtain examples where strict inequality holds in (8).

Theorem 1 states that it is at least as easy to find asymptotically optimal empirical Bayes rules as it is to find rules with property (A). In [6] p. 147 Robbins suggests that one should try to find rules with property (A) by acting as though one was treating the empirical Bayes problem.

3. **Maximin strategies for the opponent.** Suppose the component problem is such that there exists a G^* for which

$$\sup_G R(G) = R(G^*).$$

If the opponent has malicious intentions, in that he wants to select a strategy γ which maximizes

$$\limsup_{n\to\infty} E_\gamma L(\boldsymbol{\phi}_n, \boldsymbol{\Lambda}_n)$$

then if the statistician uses a $\boldsymbol{\phi}$ with property (A_0) an optimal strategy for the opponent is to select Λ_i independently, each with distribution G^*. (This is simply a sequence of independent repetitions of the maximin strategy for the component problem.) Clearly by (A_0) we have

$$\limsup_{n\to\infty} E_\gamma L(\boldsymbol{\phi}_n, \Lambda_n) \leqq R(G^*),$$

whereas from Theorem 1 it follows that with the above strategy

$$\lim_{n\to\infty} E_\gamma L(\boldsymbol{\phi}_n, \Lambda_n) = R(G^*).$$

Thus, when (A_0), holds there exists an optimal strategy for the opponent which does not take advantage of the fact that all the past observations are known to the opponent before he selects Λ_i.

4. **Some theoretical results.** In [12] Lemma 1, a pointwise inequality, valid for all the sequences $\boldsymbol{\theta}$ is given. We restate it here for the random sequence $\boldsymbol{\Lambda}$.

LEMMA 2. *With P_γ-probability one, for every n,*

$$\frac{1}{n}\sum_{i=1}^{n} R(\phi_{K_i}, \Lambda_i) \leqq R(K_n).$$

Let $\mathscr{F}_{i-1}$ denote the σ-field generated by $\Lambda_i, \boldsymbol{X}_{i-1}, \boldsymbol{Z}_{i-1}$, and let $R(\phi_i \mid \mathscr{F}_{i-1})$ denote the conditional expected loss incurred on the i-th decision, given $\mathscr{F}_{i-1}$. Corresponding to Lemma 3 of [12], we have

LEMMA 3. *If* (5) *holds*

$$[L(\boldsymbol{\phi}_n, \Lambda_n) - n^{-1}\sum_{i=1}^{n} R(\phi_i \mid \mathscr{F}_{i-1})] \xrightarrow[n\to\infty]{} 0 \quad \text{a.e.} \quad P_\gamma.$$

Proof. For $n = 1, 2, \cdots$ let

$$Y_n = \sum_{i=1}^{n} [L(\phi_i, \Lambda_i) - R(\phi_i \mid \mathscr{F}_{i-1})].$$

Then $\{Y_n\}$ is a martingale relative to $\{\mathscr{F}_n\}$, and since $\operatorname{Var}(Y_n - Y_{n-1}) \leqq M^2$ it follows (see [12] Lemma 2, or [5] p. 387E) that $Y_n/n \to 0$ a.e. P_γ.

From Lemmas 2 and 3 it follows that when (5) holds it suffices to show that

$$[R(\phi_n \mid \mathscr{F}_{n-1}) - R(\phi_{K_n}, \Lambda_n)] \xrightarrow[n\to\infty]{} 0 \tag{9}$$

a.e. P_γ, or in probability in order to prove that $\boldsymbol{\phi}$ has property (B_0) a.e. P_γ, or in probability, respectively.

5. **Testing a simple hypothesis against a simple alternative.** This problem is treated for the fixed sequential compound case in [9] and [11], where rules $\boldsymbol{t}^*$ and $\hat{\boldsymbol{t}}$ are exhibited having uniform property (A) and property (B). We shall

here consider these rules in the opponent situation. One could shorten the discussion somewhat by making more use of the results obtained in [9] and [11], but we believe it is for the benefit of the reader to give a complete proof of our statement concerning the rule $\mathbf{t}^*$, and only state the results concerning $\hat{\mathbf{t}}$. The interested reader should not find it too difficult to substitute the proofs. (See also [14].)

In this problem there are only two possible distributions, to be denoted by P_θ, $\theta = 0, 1$, and two actions $\mathfrak{A} = \{0,1\}$ where $A = j$ means "say $\theta = j$" $j = 0,1$. The loss function considered is $L(A,\theta) = a\theta(1-A) + b(1-\theta)A$, with $a > 0$, $b > 0$, and (5) holds with $M = \max(a,b)$. Without loss of generality we let $f(x,\theta)$ denote the density of P_θ, $\theta = 0, 1$, with respect to some dominating σ-finite measure μ. An apriori distribution G over $\Omega = \{0,1\}$ is entirely specified by $\eta = P(\Lambda = 1)$, and for any vector $\boldsymbol{\theta}_i$ of zeros and ones its empirical distribution G_i is completely specified by the proportion of ones $i^{-1}\sum_{j=1}^{i}\theta_j$. Here any (randomized) decision function $\phi(x)$ is completely specified by $t(x)$ which denotes the probability by which the actions $A = 1$ is taken, when $X = x$ is observed. We write t_η instead of ϕ_G, and $R(\eta)$ instead of $R(G)$. It follows easily (see e.g. (3) of [9]) that

$$(10)\qquad t_\eta(x) = \begin{cases} 1 & \text{if } \eta a f(x,1) - (1-\eta) b f(x,0) > 0 \\ 0 & \text{if } \qquad\qquad ,, \qquad\qquad < 0 \\ \text{arbitrary in } [0,1] & \text{if } \qquad\qquad ,, \qquad\qquad = 0. \end{cases}$$

Let $h(X)$ be an unbiassed estimator of θ with finite variance, and let $p_i = p_i(\mathbf{x}_i)$ equal $0, i^{-1}\sum_{j=1}^{i} h(x_j)$ or 1, as $i^{-1}\sum_{j=1}^{i} h(x_j)$ is less than 0, between 0 and 1 or greater than 1, respectively. Let $\lambda_i = i^{-1}\sum_{j=1}^{i}\Lambda_j$. By a martingale convergence theorem $|i^{-1}\sum_{j=1}^{i} h(X_j) - \lambda_i|$ converges to zero a.e. P_γ, and since $0 \leqq \lambda_i \leqq 1$ and $|\lambda_{i-1} - \lambda_i| \leqq i^{-1}$a.e. P_γ, it follows that also

$$(11)\qquad |p_{i-1}(\mathbf{X}_{i-1}) - \lambda_i| \to 0 \text{ as } i \to \infty \text{ a.e. } P_\gamma.$$

The rule $\mathbf{t}^*$ is defined by

$$t_i^*(\mathbf{x}_i) = \begin{cases} 1 & \text{if } p_{i-1}(\mathbf{x}_{i-1}) a f(x_i,1) - (1 - p_{i-1}(\mathbf{x}_{i-1})) b f(x_i,0) > 0 \\ 0 & \text{otherwise.} \end{cases}$$

A motivation for $\mathbf{t}^*$ is given in [9], and is also obtained by comparing t_i^* with (10). We shall prove

THEOREM 2. *If $R(\eta)$ is differentiable for $0 \leqq \eta \leqq 1$ then $\mathbf{t}^*$ has uniform property* (A_0), *and has property* (B_0) a.e. P_γ.

Proof. Let $\varepsilon > 0$ be given. If $R(\eta)$ is differentiable then by Lemma 1 of [9] there exists a $\delta > 0$ such that for any versions of (10)

$$(12)\qquad \max_{\theta=0,1} |R(t_\eta,\theta) - R(t_{\eta^*},\theta)| < \varepsilon/3 \quad \text{for} \quad |\eta - \eta^*| \leqq \delta.$$

Let ω denote a generic point in our measure space and let

$$S_{i-1} = \{\omega : |p_{i-1} - \lambda_i| < \delta\}.$$

Let N_1 be a fixed integer greater than $2/\delta + 1$. For $i > N_1$ we have by Chebychev's inequality

$$(13) \quad P_\gamma[|p_{i-1} - \lambda_i| \geqq \delta] \leqq P_\gamma[|p_{i-1} - \lambda_{i-1}| \geqq \delta - i^{-1}] \leqq V/[(i-1)\delta^2 - 2\delta]$$

where $V = \max\{E_0 h(X)^2, E_1[h(X) - 1]^2\}$. With our definitions we have

$$(14) \qquad R(t_i^* \mid \mathscr{F}_{i-1}) = R(t^0_{p_{i-1}}, \Lambda_i)$$

where t^0_η is defined through (10) with the arbitrary part taken to be 0. Thus for $\omega \in S_{i-1}$ we have from (12) and (14)

$$(15) \qquad |R(t_i^* \mid \mathscr{F}_{i-1}) - R(t_{\lambda_i}, \Lambda_i)| < \varepsilon/3$$

and hence for $i > N_1$, it follows from (13) and (15) that

$$(16) \quad E_\gamma L(t_i^*, \Lambda_i) = E_\gamma R(t_i^* \mid \mathscr{F}_{i-1}) \leqq E_\gamma R(t_{\lambda_i}, \Lambda_i) + \varepsilon/3 + MV/\delta[(i-1)\delta - 2].$$

Now for $n \geqq 3MN_1/\varepsilon$ $n^{-1}\sum_{i=1}^{N_1} L(t_i^*, \Lambda_i) \leqq \varepsilon/3$ a.e. P_γ and for $n > N_2$ sufficiently large $n^{-1}\sum_{i=N_1+1}^{n}[(i-1)\delta - 2]^{-1} < \varepsilon\delta/3MV$. Thus for $n > \max(N_2, 3MN_1/\varepsilon)$ we have from (16)

$$E_\gamma L(\mathbf{t}_n^*, \Lambda_n) = E_\gamma n^{-1} \sum_{i=1}^{n} L(t_i^*, \Lambda_n) \leqq E_\gamma n^{-1} \sum_{i=1}^{n} R(t_{\lambda_i}, \Lambda_i) + \varepsilon,$$

and uniform property (A_0) for $\mathbf{t}^*$ follows from Lemma 2.

We shall show that (9) holds a.e. P_γ for $\mathbf{t}^*$. This follows immediately from (11), (12) and (14), since here $R(\phi_{\Lambda_i}, \Lambda_i) = R(t_{\lambda_i}, \Lambda_i)$. Thus $\mathbf{t}^*$ has property (B_0) a.e. P_γ, and the proof is complete.

We remark that uniform property (A) and property (B) for $\mathbf{t}^*$ was proved under the same condition on $R(\eta)$, as in Theorem 2 above.

The rule $\hat{\mathbf{t}}$ is randomized. Let $p_i = p_i(\mathbf{x}_i)$ be defined as before, but with $h(x)$ a bounded estimator. For $i = 1, 2, \cdots$ let $Z_i = (z_{i1}, z_{i2})$ be independent random variables, uniformly distributed on the unit square, Z_i independent of $\mathbf{X}_i, \Lambda_i$. For $i = 2, 3, \cdots$ let

$$m_{i-1} = m(Z_i, \mathbf{x}_{i-1}) = \frac{p_{i-1}(\mathbf{x}_{i-1}) + i^{-1/4} z_{i2}}{1 + i^{-1/4}(z_{i1} + z_{i2})}$$

and define m_0 arbitrarily. $\hat{\mathbf{t}}$ is defined by means of $\hat{t}_i(\mathbf{x}_i) = t^0_{m_{i-1}}(x_i)$, $i = 1, 2, \cdots$, where the right hand side is defined through the nonrandomized version of (10) mentioned earlier. We state without proof the following

THEOREM 3. *For any two distributions P_0, P_1, $\hat{t}$ has uniform property* (A_0) *and has property* (B_0) *in probability.*

In connection with Theorem 2 and 3 we remark that, as has been pointed out to the author by Professor H. Robbins, the third paragraph on p. 1084 of [9] is incorrect, in that there the opponent case sould have been considered.

6. **The estimation problem.** Rather than discussing the opponent-case properties of all the compound estimators considered in [12] in full generality, we shall consider a particular example.

Let the component problem be that of estimating the parameter θ of a Poisson distribution, where $\theta \in \Omega = \{\theta : 0 < \alpha \leqq \theta \leqq \beta < \infty\}$ (where α, β are otherwise arbitrary). Let $\mathfrak{A} = \Omega$, $L(A, \theta) = (A - \theta)^2$, so (5) holds with $M = (\beta - \alpha)^2$. We shall consider nonrandomized estimators only, and let $\phi(x)$ denote the value in $\mathfrak{A}$ chosen with probability one. Let $Y_j(x)$ be 1 or 0 according as $X_j = x$ or $X_j \neq x$, and for $x = 0, 1, \cdots$ let

$$(17) \qquad q_i(x) = i^{-1} \sum_{j=1}^{i} Y_j(x), \quad q_{K_i}(x) = i^{-1} \sum_{j=1}^{i} P(X_j = x \mid \Lambda_i).$$

The sequential compound rule ϕ considered in [12] (see also [7]) is defined by

$$(18) \qquad \phi_i(x_i) = (x_i + 1)\, q_{i-1}(x_i + 1) / q_{i-1}(x_i), \ \alpha, \ \text{or} \ \beta$$

according as the first term on the right hand side of (18) is between α and β, less than α, or greater than β, respectively. We shall prove

THEOREM 4. *The rule* $\boldsymbol{\phi}$ *defined by* (18) *has property* (B_0) a.e. P_γ.

Proof. We shall show that (9) holds a.e. P_γ. It is easily checked that (cf. e.g. [7], (18) or [12], Section 5.)

$$(19) \qquad \phi_{K_i}(x_i) = (x_i + 1) q_{K_i}(x_i + 1) / q_{K_i}(x_i)$$

and $\alpha \leqq \phi_{K_i}(x_i) \leqq \beta$ for $x_i = 0, 1, \cdots$ a.e. P_γ. We shall first show that for every $x = 0, 1, \cdots$

$$(20) \qquad \phi_{K_i}(x) - \phi_i(X_{i-1}, x) \xrightarrow[i \to \infty]{} 0 \ \text{a.e.} \ P_\gamma.$$

(20) follows immediately from (18) and (19), since similarly to (11) one has by the definitions (17) and the martingale convergence theorem, that

$$|q_{i-1}(x) - q_{K_i}(x)| \to 0 \quad \text{a.e.} \quad P_\gamma,$$

and for fixed x $q_{K_i}(x)$ is a.e. bounded away from zero. That (20) implies (9) is seen as follows: For every $\varepsilon > 0$ there exists a finite N_ε such that $P_\theta(X \leqq N_\varepsilon) > 1 - \varepsilon / 2M$, for all $\theta \in \Omega$. Thus (20) holds uniformly in $0 \leqq x \leqq N_\varepsilon$, i.e. for every $\delta > 0$ there exists $I(\varepsilon, \delta)$ such that

$$P_\gamma(\left|\phi_{K_i}(x) - \phi_i(X_{i-1}, x)\right| < \varepsilon/6 \quad \beta \text{ for all } i > I(\varepsilon, \text{ and all } 0 \leqq x \leqq N_\varepsilon) > 1 - \delta. \tag{21}$$

But straightforward approximations show that (21) implies

$$P_\gamma(\left|R(\phi_i \mid \mathscr{F}_{i-1}) - R(\phi_{K_i}, \Lambda_i)\right| < \varepsilon \text{ for all } i > I(\varepsilon, \delta)) > 1 - \delta,$$

which is just an explicit form of (9).

References

1. D. Blackwell, *Controlled random walks*, Proc. International Congress Math. **3**, Amsterdam (1954), 336–338.
2. D. Blackwell, *An analogue of the minimax theorem for vector payoffs*, Pacific J. Math. **6**, (1956), 1–8.
3. J. F. Hannan, *The dynamic statistical decision problem when the component problem involves a finite number, m, of distributions*. Abstract, Ann. Math. Statist, **27** (1956), 212.
4. M. V. Johns Jr., *The parametric and nonparametric compound decision problem in the sequence case*. Abstract, Ann. Math. Statist. **34** (1963), 1620–1621.
5. M. Loève, *Probability Theory*. 2nd ed. Van Nostrand Comp., (1960).
6. H. Robbins, *Asymptotic subminimax solutions of compound statistical decision problems*, Proc. Second Berkeley Symp. Math. Statist and Prob. Univ. of California Press, (1951), 131–148.
7. H. Robbins, *An empirical Bayes approach to statistics*, Proc. Third Berkeley Symp. Math. Statist. and Prob. **1**, Univ. of California Press. (1955), 157–163.
8. H. Robbins, *The empirical Bayes approach to statistical decision problems*, Ann. Math. Statist. **35** (1964), 1–20.
9. E. Samuel, *Asymptotic solutions of the sequential compound decision problem*, Ann. Math. Statist. **34** (1963), 1079–1094.
10. E. Samuel, *An empirical Bayes approach to the testing of certain parametric hypotheses*, Ann. Math. Statist. **34** (1963), 1370–1385.
11. E. Samuel, *Convergence of the losses of certain decision rules for the sequential compound decision problem*, Ann. Math. Statist. **35** (1964), 1606–1621.
12. E. Samuel, *Sequential compound estimators*, Ann. Math. Statist. **36** (1965), 879–889.
13. E. Samuel, *On simple rules for the compound decision problem*, J. Roy. Statist. Soc. (B) **27** (1965), 238–244.
14. E. Samuel, *On the sequential compound decision problems in the fixed case and the opponent case*. Unpublished (1964).

THE HEBREW UNIVERSITY
OF JERUSALEM