# APPROXIMATION TO BAYES RISK IN COMPOUND DECISION PROBLEMS[1]

By Allan Oaten

*University of California, Santa Barbara*

We consider, simultaneously, $N$ statistical decision problems with identical generic structure: state space $\Omega$, action space $A$, sample space $\mathcal{X}$ and nonnegative loss function $L$ defined on $\Omega \times A \times \mathcal{X}$. With $x = (x_1, \cdots, x_N)$ distributed according to $\prod_{i=1}^{N} P_{\theta_i} = P_\theta$, a compound procedure is a vector, $\phi = (\phi_1, \cdots, \phi_N)$, such that $\phi_i(\mathbf{x}) \in A$ for each $i$ and $\mathbf{x}$. The risk of the procedure $\phi$ is $R(\theta, \phi) = N^{-1} \sum_{r=1}^{N} \int L(\theta_r, \phi_r(\mathbf{x}), x_r) P_\theta(d\mathbf{x})$ and the modified regret is $D(\theta, \phi) = R(\theta, \phi) - R(G)$ where $G$ is the empirical distribution of $\theta_1, \cdots, \theta_N$, and $R(G)$ is the Bayes risk against $G$ in the component problem.

We discuss quite wide classes of procedures, $\phi$, which consist of using $\mathbf{x}$ to estimate $G$, and then playing $\varepsilon$-Bayes against the estimate in each component problem. For one class we establish a type of uniform convergence of the conditional risk in the $m \times n$ problem (i.e. $\Omega$ has $m$ elements, $A$ has $n$), and use this to get $D(\theta, \phi) < \varepsilon + o(1)$ for another class in the $m \times n$ and $m \times \infty$ problems. Similar, but weaker, results are given in part II for the case when $\Omega$ is infinite.

**0. Summary.** A compound decision problem involves simultaneous consideration of $N$ statistical decision problems, each with the same generic structure. The risk is defined as the average of the risks of the component problems. The quantity $R(G)$, the risk of the best "simple symmetric" procedure based on knowing $G$, the empirical distribution of the states, is usually used as a standard against which compound procedures can be judged.

In papers most similar to the present one, Hannan and Robbins [6], Hannan and Van Ryzin [7] and Van Ryzin [14] have proved various types of convergence to $R(G)$ of the risks of compound procedures which consist of using the $N$ observations to estimate $G$ and then playing Bayes against the estimate in each component problem. Only component problems for which both state and action spaces are finite were considered.

Part I of this paper extends these results by considerably weakening the conditions previously imposed on the compound procedures, especially in the case of "ties", and by establishing some results for the case when the action space is infinite. We also establish, for the finite action space case, a type of uniform almost sure convergence of the conditional risk which appears in [6] only for the 2-state, 2-action case. In Part II we obtain some rather weaker results for the case when the state space is infinite.

**1. Introduction and notation.** In the component problem there is a state space $\Omega$ indexing a family of probability distributions $\{P_\omega, \omega \in \Omega\}$ over a $\sigma$-field $\mathcal{B}$ of

a sample space $\mathscr{X}$; a measurable action space $(A, \mathscr{A})$; and a loss function, $L \geq 0$, defined on $\Omega \times A \times \mathscr{X}$, which is $A$-measurable for each $\omega$ and $x$.

A (randomized) decision function, $\phi$, has domain $\mathscr{X} \times \mathscr{A}$, and is such that $\phi(x)(\cdot)$ is a probability measure on $\mathscr{A}$ for each fixed $x \in \mathscr{X}$. If the state is $\omega$, the conditional risk of $\phi$ given $x$ is

$$(1.1) \qquad L(\omega, \phi(x), x) = \int_A L(\omega, a, x)\phi(x)(da) .$$

The notation of (1.1) will be extended to $L(\omega, \lambda, x) = \int_A L(\omega, a, x)\lambda(da)$ for any signed measure $\lambda$ for which the right side exists; in this context, "$a$" will stand for the probability measure degenerate at $a$ so that, for example, $L(\omega, a - b, x) = L(\omega, a, x) - L(\omega, b, x)$. If $L(\omega, \phi(x), x)$ is $\mathscr{B}$-measurable, the unconditional risk is

$$(1.2) \qquad R(\omega, \phi) = \int L(\omega, \phi(x), x) \, dP_\omega(x) ,$$

which we abbreviate to $\int L(\omega, \phi) \, dP_\omega$.

In the compound decision problem, $\mathbf{x}_\infty = (x_1, x_2, \cdots)$ is distributed according to $P_\theta = \prod_{i=1}^\infty P_{\theta_i}$, $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots) \in \Omega^\infty$ is the unknown vector of states, and if there are $N$ problems only $\mathbf{x} = (x_1, \cdots, x_N)$ is observed. (We shall usually, as here, omit indicating dependence on $N$.) We remark that the ordering implicit in $(\theta_1, \theta_2, \cdots)$ and $(x_1, x_2, \cdots, x_N)$ may be quite arbitrary since the problems are considered simultaneously; the vector notation is used because it makes many of our statements simpler.

The choice of action for the $r$th component problem is allowed to depend on $\mathbf{x}$; this distinguishes the "set" compound problem from the "sequence" problem in which the action at the $r$th stage can depend on $(x_1, \cdots, x_r)$ only. Formally, a compound procedure of the type we consider is an array, $\boldsymbol{\phi} = \{(\phi_1^N, \phi_2^N, \cdots, \phi_N^N): N = 1, 2, \cdots\}$, such that, for each $N$ and $r \leq N$, $\phi_r^N$ is defined on $\mathscr{X}^N \times \mathscr{A}$, with $\phi_r^N(\mathbf{x})$ being, for each $\mathbf{x}$, the probability measure on $\mathscr{A}$ according to which an action is chosen for the $r$th problem. We will omit the superscript and simply write $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_N)$.

If there are $N$ problems the conditional risk, given $\mathbf{x}$, of the procedure $\boldsymbol{\phi}$ is

$$(1.3) \qquad W(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{x}) = N^{-1} \sum_{r=1}^N L(\theta_r, \phi_r(\mathbf{x}), x_r)$$

and, if $W(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{x})$ is $\mathscr{B}^N$-measurable, the unconditional risk is

$$(1.4) \qquad R(\boldsymbol{\theta}, \boldsymbol{\phi}) = \int W(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{x}) \, dP_\theta(\mathbf{x}) = \int W(\boldsymbol{\theta}, \boldsymbol{\phi}) \, dP_\theta .$$

As is becoming standard (the terminology seems to be Samuel's [13]), we say a compound procedure $\boldsymbol{\phi}$ is simple if $\phi_r(\cdot)(C)$ is a function of $x_r$ for each $C \in \mathscr{A}$. If, in addition, the $\phi_r$ are identical, say $\phi_r = \phi$, we say $\boldsymbol{\phi}$ is simple symmetric with kernel $\phi$. (Since the component problems are unrelated and uninformative about each other, simple procedures might seem natural. If one component procedure, say $\phi$, was "better" than the rest, we might use it on every component problem; we would then be using the simple symmetric procedure with kernel $\phi$.)

We shall identify simple symmetric procedures with their kernels, and write $R(\boldsymbol{\theta}, \phi)$ and $W(\boldsymbol{\theta}, \phi, \mathbf{x})$ for the risk and conditional risk given $\mathbf{x}$ of the simple symmetric procedure whose kernel is $\phi$. For $\boldsymbol{\theta} \in \Omega^\infty$ and any simple symmetric procedure $\phi$, $R(\boldsymbol{\theta}, \phi) = N^{-1} \sum_{r=1}^{N} R(\theta_r, \phi)$. Thus if $R(F, \phi) = \int R(\omega, \phi) \, dF =$ the risk, in the component problem, of the procedure $\phi$ against the prior distribution $F$ on the states, and if $G$ denotes the empirical distribution of $\theta_1, \theta_2, \cdots, \theta_N$ (as it shall, henceforth), then $R(\boldsymbol{\theta}, \phi) = R(G, \phi)$ for all $\boldsymbol{\theta}$ and $\phi$, and we shall identify the two henceforth.

Let $\mathscr{G}$ be the set of distributions on $\Omega$ and, in the component problem, define the Bayes envelope by $R(F) = \inf R(F, \phi)$, for each $F \in \mathscr{G}$, where the infimum is taken over $\Phi$, the class of procedures for which $R(\omega, \phi)$ exists and is a measurable function of $\omega$ (this restriction is unnecessary when $\Omega$ is finite). In the compound problem, $R(G)$ is the "best" we could do with a simple symmetric procedure if we knew $G$. Defining the "modified regret" by

$$(1.5) \qquad\qquad D(\boldsymbol{\theta}, \phi) = R(\boldsymbol{\theta}, \phi) - R(G)$$

we have (cf. Gilliland [3], page 1890) that $D(\boldsymbol{\theta}, \phi) \geqq 0$ for all simple $\phi$, and if the component problem is non-trivial—i.e. if there is no strategy $\psi$ for which $R(\omega, \psi) = \inf_\phi R(\omega, \phi)$ for all $\omega$—then there is a $\delta > 0$ such that $\sup_\theta D(\boldsymbol{\theta}, \phi) \geqq \delta$ for all simple $\phi$.

The problem is to find non-simple procedures, $\phi$, for which $D(\boldsymbol{\theta}, \phi)$ converges in some sense to zero as $N \to \infty$. Robbins [11], introducing the compound decision problem, proposed estimating $G$ from $\mathbf{x}$, and then playing Bayes against this estimate in each component problem. This has been carried out in [6], [7] and [14], and various types of convergence of $D(\boldsymbol{\theta}, \phi)$ obtained.

Throughout this paper $\varepsilon$ denotes an arbitrary nonnegative number, possibly depending on $N$. In part I, we consider procedures which consist of playing $\varepsilon$-Bayes against an estimate of $G$. With fewer restrictions on the estimator than have previously been implicit, we consider two types of procedure: the "polytope" and the "equivariant". Each of these is considerably less restricted, especially in the case of ties (or, in our case, "$\varepsilon$-ties"), than procedures previously considered. For the first we shall prove a type of "uniform almost sure" convergence of $W(\boldsymbol{\theta}, \phi, \mathbf{x})$, analogous to Theorem 3 of [6], and use it to obtain $D(\boldsymbol{\theta}, \phi) < o(1) + \varepsilon$. This result helps establish the same for equivariant procedures, and this last result can be extended to infinite action spaces. In part II we modify these procedures and obtain slightly weaker results for the case when $\Omega$ is infinite.

In both cases the convergence of $W(\boldsymbol{\theta}, \phi, \mathbf{x})$ to $R(G)$ is our first concern, and this is established by treating separately $|W(\boldsymbol{\theta}, \phi, \mathbf{x}) - R(G, \phi_{\mathbf{x}})|$ and $|R(G, \phi_{\mathbf{x}}) - R(G)|$, where $\phi_{\mathbf{x}}$ is a component procedure, to be defined later, which plays $\varepsilon$-Bayes against $\hat{G}(\mathbf{x})$, an estimate of $G$. We will be assuming $\hat{G}(\mathbf{x})$ is "close" to $G$ so the second of these terms is dealt with if we can show that whenever $F$ is "close" to $G$ and $\psi$ plays $\varepsilon$-Bayes against $F$ (i.e. $R(F, \psi) \leq R(F) + \varepsilon$) then $R(G, \psi) - R(G)$ is small. This is a consequence of the following lemma.

LEMMA 1. *Let $G$ and $F$ be any distributions on $\Omega$, and let $\phi \in \Phi$, the set of component procedures for which $R(\omega, \phi)$ exists and is $\omega$-measurable. Then*

$$R(G, \phi) - R(G) \leqq \sup_{\lambda, \psi \in \Phi} \int [R(\omega, \lambda) - R(\omega, \psi)]\, d(G - F)(\omega)$$
$$+ R(F, \phi) - R(F).$$

PROOF.

(1)
$$R(G, \phi) - R(F, \phi) = \int R(\omega, \phi)\, d(G - F)(\omega),$$

for any procedure $\phi$. If $\phi_n$ is any procedure for which $R(G, \phi_n) \leqq R(G) + n^{-1}$, then

(2)
$$R(F) - R(G) \leqq R(F, \phi_n) - R(G, \phi_n) + n^{-1}$$
$$= -\int R(\omega, \phi_n)\, d(G - F)(\omega) + n^{-1}.$$

Adding (1) and (2) and taking the supremum,

$$R(G, \phi) - R(G) - R(F, \phi) + R(F)$$
$$\leqq \sup_{\lambda, \psi \in \Phi} \int [R(\omega, \lambda) - R(\omega, \psi)]\, d(G - F)(\omega) + n^{-1}.$$

Since the left side is independent of $n$, the proof is complete.

We note that if $\Omega = \{1, 2, \cdots, m\}$, $\int L(\omega, \psi)\, dP_\omega \leqq M < \infty$ for every $\omega$ and $\psi$, and $\phi \in \Phi(F)$, (see Definition 2 below), then

(1.6)
$$R(G, \phi) - R(G) \leqq M \sum_{\omega=1}^{m} |G_\omega - F_\omega| + \varepsilon.$$

## I. FINITE STATE SPACES

**2. Definitions and preliminaries.** Throughout this part, $\Omega = \{1, 2, \cdots, m\}$. In this case we have $P_\omega \ll \mu = \sum_{\omega=1}^{m} P_\omega$, and we define $f(\omega, \cdot) = (dP_\omega/d\mu)(\cdot)$. We assume that $L(\omega, a, x)$ is $\mathscr{B}$-measurable for each fixed $\omega$ and $a$, and that

(2.1)
$$\int L(\omega, a)\, dP_\omega \leqq M < \infty \quad \text{for all} \quad \omega \quad \text{and} \quad a.$$

DEFINITION 1. $\hat{G}$ is a uniformly consistent estimator of $G$ if, for any $\eta > 0$ and $\gamma > 0$, there exists $N_1(\eta, \gamma)$ such that

$$\sup_{N > N_1(\eta, \gamma)} P_\theta [\sum_{\omega=1}^{m} |\hat{G}_\omega(\mathbf{x}) - G_\omega| > \eta] < \gamma \quad \text{for all} \quad \theta \in \Omega^\infty;$$

where, for each $F$, $F_\omega$ is the mass assigned by $F$ to the point $\omega$. With the supremum inside the square brackets, $\hat{G}$ is uniformly strongly consistent. (An estimator $\hat{G}$ is really a sequence of functions $\hat{G}_1, \hat{G}_2, \cdots$ with $\hat{G}_N : \mathscr{X}^N \to \mathscr{G}$ being $\mathscr{B}^N$-measurable for each $N$. We shall not need to emphasize this formality however.)

DEFINITION 2. For each $F \in \mathscr{G}$, let $\Phi(F)$ be the set of component procedures $\varepsilon$-Bayes against $F$, i.e. those $\phi$ for which $R(F, \phi) \leqq R(F) + \varepsilon$. For $\hat{G}$ an estimator, $\Phi(\hat{G})$ is the set of procedures of the following kind: on observing $\mathbf{x}$ we first estimate $G$ by $\hat{G}(\mathbf{x})$ and then choose (the choice can depend on $\mathbf{x}$) a component procedure $\phi_{\mathbf{x}} \in \Phi(\hat{G}(\mathbf{x}))$; we use this procedure in each component problem—i.e. we use the simple symmetric procedure whose kernel is $\phi_{\mathbf{x}}$. Formally,

$$\Phi(\hat{G}) = \{\boldsymbol{\phi} : \forall \mathbf{x}, \exists \phi_{\mathbf{x}} \in \Phi(\hat{G}(\mathbf{x})) \text{ such that } \forall r, \phi_r(\mathbf{x}) = \phi_{\mathbf{x}}(x_r)\}.$$

Henceforth, for each $\mathbf{x}$ and each $\boldsymbol{\phi} \in \Phi(\hat{G})$, $\phi_\mathbf{x}$ will denote the component procedure given by Definition 2.

We shall be concerned with two subsets of $\Phi(\hat{G})$: "polytope" procedures in Sections 3 and 4, and "equivariant" procedures in Sections 5 and 6.

Throughout Sections 3, 4 and 5, $A = \{1, 2, \cdots, n\}$. Let $E^k$ be $k$-dimensional Euclidean space and let $z: \mathscr{H} \to E^{mn}$ be given by

$$(2.2) \qquad z(\omega, a, x) = L(\omega, a, x)f(\omega, x) , \qquad \omega = 1, 2, \cdots, m; \ a = 1, 2, \cdots, n .$$

We shall adopt, for $z$, the same convention as for $L$; i.e. $z(\omega, \lambda, x) = \int z(\omega, a, x) \, d\lambda(a)$ for any signed measure on $\mathscr{A}$ for which the right side exists.

**3. Polytope procedures.** In this section we define polytope procedures, and prove the main result about them.

DEFINITION 3. A set $H \in E^k$ is a half-space if, for some linear functional $l$ and some number $p$, either $H$ or its complement is $\{y: l(y) < p\}$. Let $\mathscr{H}_s$ be the set of all intersections of $s$ half-spaces, $\mathscr{H}_s{}^t$ the set of all unions of $t$ members of $\mathscr{H}_s$, and $\mathscr{K}_s{}^t = z^{-1}(\mathscr{H}_s{}^t)$.

For each $F \in \mathscr{G}$ we want to restrict attention to those members of $\Phi(F)$ which take on only finitely many values (at most $v$), and for which the corresponding induced partition of $\mathscr{H}$ is a collection of regions each of which is an element of $\mathscr{K}_s{}^t$ for some $t$ and $s$. ($s$, $t$ and $v$ are arbitrary finite numbers, fixed throughout what follows; we will frequently not indicate dependence on them.) Formally, $\psi \in \Phi(F)$ is an element of $\Phi_1(F)$ if $\{\psi(x): x \in \mathscr{H}\} \subset \{\nu_1, \nu_2, \cdots, \nu_v\}$, where $\nu_1, \cdots, \nu_v$ are distinct measures on $\mathscr{A}$, and, for each $j$, $\{x: \psi(x) = \nu_j\} = Q_j$ for some (possibly empty) $Q_j \in \mathscr{K}_s{}^t$. Hence $\psi(x)(a) = \sum_{j=1}^v Q_j(x)\nu_j(a)$ for each $x \in \mathscr{H}$ and $a \in A$, where, as we shall continue to do, we have identified sets $(Q_j)$ and their indicator functions $\{Q_j(x)\}$.

DEFINITION 4. (Polytope Procedures). Let $\hat{G}$ be an estimator. Then a procedure $\boldsymbol{\phi} \in \Phi(\hat{G})$ is an element of $\Phi_1(\hat{G})$ if, for each $\mathbf{x}$, $\phi_\mathbf{x} \in \Phi_1(\hat{G}(\mathbf{x}))$, where $\phi_\mathbf{x}$ is as in Definition 2.

Most results obtained so far (e.g. in [6] and [14]) have been obtained for special subsets of $\Phi_1(\hat{G})$—usually the class $\Phi_1(F)$ is replaced, in Definition 4, by those procedures $\psi$ for which $\psi(x)(a) = \sum_{B \subset A} Q_{FB}^0(x)\nu_B(a)$, where $x \in Q_{FB}^0$ if $B$ is the set of Bayes acts against $F$ when $x$ is observed. That $Q_{FB}^0 \in K_s{}^t$ for sufficiently large $t$ and $s$ follows immediately from Lemma 3 in Section 5. Usually $\nu_B(a)$ is restricted to the values 0 and 1. In addition the form of the estimator, $\hat{G}$, is usually restricted, especially when rates have been obtained; the most common form is Hannan's "average of unbiased estimators", as in Section 3 of [14].

THEOREM 1. *Let $\hat{G}$ be a uniformly strongly consistent estimator of $G$. Then given $\eta > 0$ and $\gamma > 0$ there exists $N(\eta, \gamma)$ such that*

$$P_\theta[\sup_{N > N(\eta, \gamma)} |W(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{x}) - R(G)| > \varepsilon + \eta] < \gamma$$

*for all $\boldsymbol{\theta} \in \Omega^\infty$ and $\boldsymbol{\phi} \in \Phi_1(\hat{G})$.*

The proof of this theorem will invoke the following lemma:

LEMMA 2. *Let* $(\mathscr{X}, \mathscr{B}, P)$ *be a probability space, and let* $P_N$ *be the empirical distribution of* $N$ *i.i.d. random variables* $\sim P$. *Let* $h: \mathscr{X} \to E^1$ *be* $P$-*integrable, and let* $g: \mathscr{X} \to E^k$ *be* $\mathscr{B}$-*measurable. Then, for any* $s$ *and* $t$,

$$P^\infty[\sup_{H \in H_s{}^t} |\int g^{-1}(H)h \, d(P_N - P)| \to 0 \quad as \quad N \to \infty] = 1 .$$

This (and much more) is proved by Ranga Rao ([10], especially Lemma 7.5) when $s = t = 1$, $\mathscr{X} = E^k$ and $g$ is the identity. Extending his results to intersections of half-spaces, to unions of such intersections, and finally to the inverse images of such sets under a measurable transformation, is not difficult. Details are given in [9] pages 12–13.

PROOF OF THEOREM 1. Let $\phi \in \Phi_1(\hat{G})$ and, given $\mathbf{x}$, let $\phi_\mathbf{x}$ be the component procedure guaranteed by Definitions 2 and 4. Then by Lemma 1

$$(3.1) \quad |W(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{x}) - R(G)| \leqq |W(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{x}) - R(G, \phi_\mathbf{x})| + |R(G, \phi_\mathbf{x}) - R(G)|$$
$$\leqq |W(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{x}) - R(G, \phi_\mathbf{x})| + M \sum_{\omega=1}^m |G_\omega - \hat{G}_\omega(\mathbf{x})| + \varepsilon .$$

The first term is $|W(\boldsymbol{\theta}, \phi_\mathbf{x}, \mathbf{x}) - R(G, \phi_\mathbf{x})|$, so rather like $|W(\boldsymbol{\theta}, \psi, \mathbf{x}) - R(G, \psi)|$ which we could handle using a variant of the strong law of large numbers. However, in our case, the simple symmetric procedure, $\phi_\mathbf{x}$ or $\psi$, is a function of the observations $\mathbf{x}$ so we need, speaking roughly, to show convergence for the "worst" $\psi$, i.e. we need to show convergence of $\sup_\psi |W(\boldsymbol{\theta}, \psi, \mathbf{x}) - R(G, \psi)|$, the sup being over $\psi \in \Phi_1(F)$ for all $F$.

If $\psi \in \Phi_1(F)$ then $\psi = \sum_{j=1}^v Q_j \nu_j$, where $Q_j \in \mathscr{X}_s{}^t$ for each $j$, and $\nu_1, \cdots, \nu_v$ are distinct measures on $\mathscr{A}$. Then

$$W(\boldsymbol{\theta}, \psi, \mathbf{x}) = N^{-1} \sum_{r=1}^N \sum_{j=1}^v Q_j(x_r) \sum_{a=1}^n \nu_j(a)L(\theta_r, a, x_r)$$
$$= \sum_{\omega=1}^m N^{-1} \sum_{a=1}^n \sum_{j=1}^v \nu_j(a) \sum_{\{r: \theta_r=\omega\}} Q_j(x_r)L(\theta_r, a, x_r) .$$

Now $\sum_{\{r: \theta_r=\omega\}} Q_j(x_r)L(\theta_r, a, x_r)$ is the sum of $N_\omega = N_\omega(\boldsymbol{\theta}) = \sum_{r=1}^N [\theta_r = \omega]$ i.i.d. random variables; accordingly, if $P_{N_\omega}$ denotes the empirical distribution of the $N_\omega$ random variables $\{x_r : r = \omega\}$ we may write

$$W(\boldsymbol{\theta}, \psi, \mathbf{x}) = \sum_{\omega=1}^m N^{-1} \sum_{a=1}^n \sum_{j=1}^v \nu_j(a)N_\omega \int Q_j L(\omega, a) \, dP_{N_\omega} .$$

Subtracting from each side its expectation,

$$W(\boldsymbol{\theta}, \psi, \mathbf{x}) - R(G, \psi) = \sum_{\omega=1}^m \frac{N_\omega}{N} \sum_{a=1}^n \sum_{j=1}^v \nu_j(a) \int Q_j L(\omega, a) \, d(P_{N_\omega} - P_\omega) .$$

Thus, with $S(N, \boldsymbol{\theta}, \omega, \mathbf{x}) = \sum_{a=1}^n v \sup_{Q \in \mathscr{X}_s{}^t} |\int QL(\omega, a) \, d(P_{N_\omega} - P_\omega)|$, we have

$$(3.2) \qquad \sup_\psi |W(\boldsymbol{\theta}, \psi, \mathbf{x}) - R(G, \psi)| \leqq \sum_{\omega=1}^m \frac{N_\omega}{N} S(N, \boldsymbol{\theta}, \omega, \mathbf{x}) .$$

By applying Lemma 2, with $h = L(\omega, a)$ and $g(x) = (z(1, 1, x), \cdots, z(m, n, x))$, we have, for all $\boldsymbol{\theta}, \omega, \eta_1 > 0$ and $\gamma_1 > 0$,

$$P_\theta[\sup_J S(N, \boldsymbol{\theta}, \omega, \mathbf{x}) > \eta_1] < \gamma_1 ,$$

where $J = J(k_1, \omega, \boldsymbol{\theta}) = \{N : N_\omega(\boldsymbol{\theta}) > k_1\}$ and $k_1 = k_1(\eta_1, \gamma_1)$ is given by Lemma 2.

With $H = H(k_1, \omega, \boldsymbol{\theta}) = \{N : N > N' \text{ and } N_\omega(\boldsymbol{\theta}) \leqq k_1\}$, $\eta_1 = \eta/2m$ and $\gamma_1 = \gamma/2m$, we have

$$P_\theta\left[\sup_{N>N'} \sum_{\omega=1}^m \frac{N_\omega(\boldsymbol{\theta})}{N} S(N, \boldsymbol{\theta}, \omega, \mathbf{x}) > \frac{\eta}{2}\right]$$

$$\leqq \sum_{\omega=1}^m P_\theta\left[\sup_{N>N'} \frac{N_\omega(\boldsymbol{\theta})}{N} S(N, \boldsymbol{\theta}, \omega, \mathbf{x}) > \frac{\eta}{2m}\right]$$

(3.3)
$$\leqq \sum_{\omega=1}^m \left\{ P_\theta\left[\sup_J S(N, \boldsymbol{\theta}, \omega, \mathbf{x}) > \frac{\eta}{2m}\right]\right.$$

$$\left. + P_\theta\left[\frac{k}{N'} \max_H S(N, \boldsymbol{\theta}, \omega, \mathbf{x}) > \frac{\eta}{2m}\right]\right\}$$

$$< \gamma/2 + g(N')$$

where $g(N') \to 0$ as $N' \to \infty$ since $\max_H S(N, \boldsymbol{\theta}, \omega, \mathbf{x})$ is finite valued because $H$ yields at most $k_1$ different values of $S$.

Now let $N(\eta, \gamma)$ be so large that $g(N(\eta, \gamma)) < \gamma/4$ and

$$P_\theta\left[\sup_{N>N(\eta,\gamma)} \sum_{\omega=1}^m |\hat{G}_\omega(x) - G_\omega| > \frac{\eta}{2Mn}\right] < \gamma/4 \quad \text{for all} \quad \boldsymbol{\theta}.$$

Then (3.1), (3.2) and (3.3) yield Theorem 1.

## 4. Convergence of $D(\boldsymbol{\theta}, \boldsymbol{\phi})$ and remarks.

REMARK 1. From (3.1) and (3.3) we have that if $\hat{G}$ is uniformly consistent (not necessarily strongly) and $\eta > 0$ then

(4.1)    $$\sup_\phi \sup_\theta P_\theta[|W(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{x}) - R(G)| > \eta + \varepsilon] \to 0 \quad \text{as} \quad N \to \infty,$$

where the sup is over $\boldsymbol{\phi} \in \Phi_1(\hat{G})$.

REMARK 2. An alternative way of looking at the component problem is: Let $P$ be any distribution on $\mathscr{X}^\Omega$ whose $\omega$th marginal is $P_\omega$. Let $Y \in \mathscr{X}^\Omega$ be distributed according to $P$ and suppose that, if the state is $\omega$, only $Y(\omega)$, the $\omega$th coordinate of $Y$, is observed; the loss resulting from action $a$ is $L(\omega, a, Y(\omega))$.

For the compound problem, $\mathbf{Y}_\infty = (Y_1, Y_2, \cdots)$ is distributed according to $P^\infty$, $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots) \in \Omega^\infty$ and if there are $N$ problems only $(Y_1(\theta_1), \cdots, Y_N(\theta_N)) = Y(\boldsymbol{\theta})$ is observed. The conditional risk, given $Y(\boldsymbol{\theta})$, of the procedure $\boldsymbol{\phi}$, is $W(\boldsymbol{\theta}, \boldsymbol{\phi}, Y(\boldsymbol{\theta})) = N^{-1} \sum_{r=1}^N L(\theta_r, \phi_r(Y(\boldsymbol{\theta})), Y_r(\theta_r))$.

This is clearly the same problem as we have been dealing with but, along with its notational clumsiness, this formulation allows us to state and prove results using the fixed measure $P^\infty$ rather than the state-dependent $P_\theta$. The usefulness of this will be demonstrated in the next remark and in the beginning of the proof of Theorem 2. For the moment we note that the "probability" part of the statement of Theorem 1 can be written

$$P^\infty[\sup_{N>N(\eta,\gamma)} |W(\boldsymbol{\theta}, \boldsymbol{\phi}, Y(\boldsymbol{\theta})) - R(G)| > \varepsilon + \eta] < \gamma.$$

REMARK 3. COROLLARY TO THEOREM 1. *If $\hat{G}$ is uniformly consistent, then* $\sup_{\phi} \sup_{\theta} D(\theta, \phi) < \varepsilon + o(1)$ *as* $N \to \infty$, *where the sup is over* $\phi \in \Phi_1(\hat{G})$.

PROOF. In the notation of Remark 2, (4.1) becomes

$$(4.2) \qquad \sup_{\phi} \sup_{\theta} P^{\infty}[|W(\theta, \phi, \mathbf{Y}(\theta)) - R(G)| > \eta + \varepsilon] \to 0 \quad \text{as} \quad N \to \infty .$$

Let $\phi \in \Phi_1(\hat{G})$ and $C_N = \{\mathbf{Y} : W(\theta, \phi, \mathbf{Y}(\theta)) > R(G) + \varepsilon + \delta/2\}$, so that

$$(4.3) \qquad R(\theta, \phi) = \int W(\theta, \phi, \mathbf{Y}(\theta)) \, dP^{\infty}$$
$$\leqq R(G) + \varepsilon + \delta/2 + \int_{C_N} W(\theta, \phi) \, dP^{\infty} .$$

Since $W(\theta, \phi, \mathbf{Y}(\theta)) \leqq N^{-1} \sum_{r=1}^{N} \max_{\omega, a} L(\omega, a, Y_r(\omega)) = N^{-1} \sum_{r=1}^{N} V(Y_r)$, say, we have $\int_{C_N} W(\theta, \phi) \, dP^{\infty} \leqq N^{-1} \sum_{r=1}^{N} \int_{C_N} V(Y_r) \, dP^{\infty} < \delta/2$ for $P(C_N)$ sufficiently small, because the $V(Y_r)$ are identically distributed and integrable. The corollary now follows from (4.2) and (4.3).

REMARK 4. None of these results is affected if $\varepsilon$ depends on $N$. In particular, if $\varepsilon = o(1)$ as $N \to \infty$, the conclusion of the corollary becomes $\sup_{\phi} \sup_{\theta} D(\theta, \phi) < o(1)$ as $N \to \infty$.

REMARK 5. The class $\Phi_1(\hat{G})$ may include procedures, $\phi$, for which $W(\theta, \phi, \mathbf{x})$ is not $\mathscr{B}^N$-measurable (though, for each $\mathbf{x}$, the component procedures $\phi_{\mathbf{x}}$ must be elements of $\Phi$). Such procedures are included in Theorem 1 and its corollary in the sense that, whether or not $W(\theta, \phi, \mathbf{x})$ is $\mathscr{B}^N$-measurable, there is a measurable function $U(\theta, \mathbf{x}) \geqq W(\theta, \phi, \mathbf{x})$ for all $\theta$ and $\mathbf{x}$, with $U$ having the properties asserted for $W$. This follows from the fact that both $S(N, \theta, \omega, \mathbf{x})$ and $\sum_{\omega=1}^{m} |\hat{G}_{\omega}(\mathbf{x}) - G_{\omega}|$ are $\mathscr{B}^N$-measurable, and from (3.1) and (3.2).

REMARK 6. The restrictions on the class $\Phi_1(\hat{G})$ are two, both the results of restrictions on the classes $\Phi_1(F)$: that for $\phi \in \Phi_1(F)$, (1) $\{\phi(x) : x \in \mathscr{K}\}$ is finite and (2) $\{x : \phi(x) = \nu_j\} \in \mathscr{K}_s^t$ for each $j$. The crux of the proof of Theorem 1 is the convergence of $S(N, \theta, \omega, \mathbf{x})$ which follows from Lemma 2 because of the structure of the family of functions $\bigcup_{F \in \mathscr{E}} \{\phi(\cdot)(a) : a \in A, \phi \in \Phi_1(F)\}$. However if, for each $F \in \mathscr{E}$, $\Phi^*(F)$ is a subclass of $\Phi(F)$ and, for each $\omega$ and $a$, $P_{\theta}[\sup |gL(\omega, a) d(P_{N_{\omega}} - P_{\omega})| \to 0$ as $N \to \infty] = 1$ where the sup is taken over $g(\cdot) \in \bigcup_{F \in \mathscr{E}} \{\phi(\cdot)(a) : \phi \in \Phi^*(F)\}$, then Theorem 1 would hold with $\Phi_1(\hat{G})$ replaced by $\Phi^*(\hat{G})$. The families $\bigcup_{F \in \mathscr{E}} \{\phi(\cdot)(a) : \phi \in \Phi_1(F)\}$, $a \in A$, are by no means the only ones with this property; however they are of particular interest as a natural generalization of the standard situation, mentioned after Definition 4, in which, for each $F$, $\Phi(F)$ is restricted to those procedures, $\phi$, for which $\phi(x)(a) = \sum_{B \subset A} Q_{FB}^0(x) \nu_B(a)$.

In fact, although $\nu_B$ (the measure according to which an act is chosen when $B$ is the set of Bayes acts) is usually permitted only to depend on $B$ (most commonly, with $A = \{1, 2, \cdots, n\}$, $\nu_B$ is degenerate at the "minimum" member of $B$; see e.g. [6], [7] and [14]), Theorem 1 also applies to the case where $\nu_B$ is also permitted to depend on $F$; and the conclusions of Theorem 1 hold if $\nu_B$ is a measur-

able function of $x$ since, from Lemma 2, $\sup_{Q \in \mathscr{X}_s{}^t} |\int Q\nu_B(a)L(\omega, a) \, d(P_{N_\omega} - P_\omega)|$ converges to 0 with probability 1 as $N \to \infty$, for each $\omega$, $a$ and $B \subset A$. However Lemma 2 is not applicable if $\nu_B$ depends on both $F$ and $x$, for we would then need to take the sup over all $F \in \mathscr{S}$ as well; for details see page 19 of [9].

**5. Uniformly $\varepsilon$-Bayes and equivariant procedures.** In this section we define "equivariant" procedures and prove the main result concerning them. Since our notational requirements are somewhat different in this section, it is perhaps as well to note that, in addition to previous formulae,

$$R(\omega, \phi) = \int z(\omega, \phi, x) \, d\mu(x)$$

and

$$R(\theta, \phi) = \int W(\theta, \phi, \mathbf{x}) \prod_{r=1}^{N} f(\theta_r, x_r) \, d\mu^N(\mathbf{x}) \, .$$

DEFINITION 5. A procedure $\phi \in \Phi$ is uniformly $\varepsilon$-Bayes against $F \in \mathscr{S}$ if $\phi(x)(B(F, x, \varepsilon)) = 1$ for all $x$, where

$$(5.1) \qquad B(F, x, \varepsilon) = \{a : \sum_{\omega=1}^{m} z(\omega, a - b, x)F_\omega \leqq \varepsilon/m \text{ for all } b \in A\} \, .$$

If $\phi$ is uniformly $\varepsilon$-Bayes against $F$, then

$$R(F, \phi) = \sum_{\omega=1}^{m} F_\omega \int L(\omega, \phi(x), x) \, dP_\omega(x) = \sum_{\omega=1}^{m} F_\omega \int z(\omega, \phi) \, d\mu$$

$$= \int \sum_{\omega=1}^{m} z(\omega, \phi)F_\omega \, d\mu \leqq R(F) + \varepsilon$$

since $\sum_{\omega=1}^{m} z(\omega, \phi)F_\omega \leqq \min_{a \in A} \sum_{\omega=1}^{m} z(\omega, a)F_\omega + \varepsilon/m$ and $\mu(\mathscr{X}) = m$.

Hence a uniformly $\varepsilon$-Bayes procedure is $\varepsilon$-Bayes in the usual sense.

LEMMA 3. *Let $Q_{FB}^\varepsilon = \{x : B(F, x, \varepsilon) = B\}$. Then for each $F \in \mathscr{S}$, $B \subset A$ and $\varepsilon \geqq 0$, $Q_{FB}^\varepsilon \in \mathscr{X}_s{}^t$ for sufficiently large $t$ and $s$.*

PROOF. Let $T_{Fba} = \{x : \sum_{\omega=1}^{m} z(\omega, b - a, x)F_\omega \leqq \varepsilon/m\}$. Then $T_{Fba} \in \mathscr{X}_1$. Also, $Q_{FB}^\varepsilon = \bigcap_{b \in B} \bigcap_{a=1}^{n} T_{Fba} \cap \bigcap_{d \in A \sim B} \bigcup_{e=1}^{n} T_{Fde}^c$. But $\bigcup_{e=1}^{n} T_{Fde}^c = T_{Fd1}^c + T_{Fd1} T_{Fd2}^c + \cdots + T_{Fd1} \cdots T_{Fd(n-1)} T_{Fdn}^c$, a disjoint union of members of $\mathscr{X}_n$. The lemma now follows since, for any $s$ and $t$, $E_1 \in \mathscr{X}_s$ and $E_2 \in K_t$ implies $E_1 \cap E_2 \in \mathscr{X}_{s+t}$.

Since $\varepsilon$ is fixed in our discussion, we shall abbreviate $B(F, x, \varepsilon)$ and $Q_{FB}^\varepsilon$ to $B(F, x)$ and $Q_{FB}$.

DEFINITION 6. For each $F \in \mathscr{S}$, let $\Phi_u(F)$ be the set of component procedures uniformly $\varepsilon$-Bayes against $F$; and, for $\hat{G}$ an estimator, let $\Phi_u(\hat{G})$ be the class $\{\boldsymbol{\phi} : \phi_\mathbf{x} \in \Phi_u(\hat{G}(\mathbf{x})) \text{ for each } \mathbf{x}\}$, where $\phi_\mathbf{x}$ is as in Definition 2.

Let $g(1, 2, \cdots, N) = (g^{-1}1, g^{-1}2, \cdots, g^{-1}N)$ be an arbitrary permutation of $(1, 2, \cdots, N)$ and, for any vector $\mathbf{r} = (r_1, \cdots, r_N)$, let $g\mathbf{r} = (r_{g^{-1}1}, \cdots, r_{g^{-1}N})$. Let $\mathscr{P}$ be the set of permutations on $(1, 2, \cdots, N)$, and let $E$ denote expectation under the uniform distribution on $\mathscr{P}$.

DEFINITION 7. A procedure $\boldsymbol{\phi}$ is equivariant if, for each $N$, $\mathbf{x}$, and $g$, $\boldsymbol{\phi}(g\mathbf{x}) = g\boldsymbol{\phi}(\mathbf{x})$; i.e. $\phi_r(g\mathbf{x}) = \phi_{g^{-1}r}(\mathbf{x})$ for each $r$.

Since the problems are considered simultaneously, their labeling is artificial so equivariant procedures seem reasonable. Further, if $\boldsymbol{\phi}$ is any procedure, let

$\phi_r(\mathbf{x}) = E\psi_{gr}(g\mathbf{x})$ for each $\mathbf{x}$ and $r$. Then $\boldsymbol{\phi}$ is equivariant and it is easy to show that $EW(g\boldsymbol{\theta}, \boldsymbol{\phi}, g\mathbf{x}) = W(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{x})$ for each $\boldsymbol{\theta}$ and $\mathbf{x}$, so that $ER(g\boldsymbol{\theta}, \boldsymbol{\phi}) = R(\boldsymbol{\theta}, \boldsymbol{\phi})$. Thus if $R^*(G) = \inf R(\boldsymbol{\theta}, \boldsymbol{\phi})$ where $G$ is given by $\theta_1, \cdots, \theta_N$ and the infimum is taken over all equivariant $\boldsymbol{\phi}$, then $ER(g\boldsymbol{\theta}, \boldsymbol{\phi}) \geqq R^*(G)$ for all procedures, $\boldsymbol{\phi}$. (This might suggest that $R^*(G)$ would make a better "standard" than $R(G)$; however, the difference between these quantities has been shown to be small under various conditions in [6], [8] and, particularly, in [4], where it is shown that $R(G) - R^*(G) < O(N^{-\frac{1}{2}})$ uniformly in $\boldsymbol{\theta}$, under conditions considerably broader than we have adopted here. We shall use this result in the proof of Theorem 2.)

DEFINITION 8.  $\Phi_2(\hat{G})$ is the set of equivariant members of $\Phi_u(\hat{G})$.

THEOREM 2.  *Let $\hat{G}$ be a uniformly consistent estimator. Then* $\sup_{\Phi_2(\hat{G})} \sup_\theta D(\boldsymbol{\theta}, \boldsymbol{\phi}) < o(1) + \varepsilon$ *as $N \to \infty$.*

PROOF.  Let $\hat{G}'(\mathbf{x}) = E\hat{G}(g\mathbf{x})$.  We show that $\Phi_2(\hat{G}) \subset \Phi_2(\hat{G}')$ and that $\hat{G}'$ is uniformly consistent, so that it suffices to prove the theorem for $\Phi_2(\hat{G}')$ instead of $\Phi_2(\hat{G})$.

Let $\boldsymbol{\phi} \in \Phi_2(\hat{G})$ and, given $r$ and $g$, let $g^{-1}k = r$. Then $\phi_k(g\mathbf{x})(B(\hat{G}(g\mathbf{x}), x_{g^{-1}k})) = 1$ since $\boldsymbol{\phi} \in \Phi_u(\hat{G})$; so $\phi_{g^{-1}k}(\mathbf{x})(B(\hat{G}(gx), x_{g^{-1}k})) = 1$ since $\boldsymbol{\phi}$ is equivariant. Thus $\phi_r(\mathbf{x})(\bigcap_{g \in \mathscr{P}} B(\hat{G}(g\mathbf{x}), x_r)) = 1$, so $\boldsymbol{\phi} \in \Phi_2(\hat{G}')$ if $\bigcap_{g \in \mathscr{P}} B(\hat{G}(g\mathbf{x}), x_r) \subset B(\hat{G}'(\mathbf{x}), x_r)$. For any $a \in \bigcap_{g \in \mathscr{P}} B(\hat{G}(g\mathbf{x}), x_r)$, $\sum_{\omega=1}^m \hat{G}_\omega(g\mathbf{x})z(\omega, a - b, x_r) \leqq \varepsilon/m$ for all $b \in A$ and $g \in \mathscr{P}$. Hence $(N!)^{-1} \sum_{g \in \mathscr{P}} \sum_{\omega=1}^m \hat{G}_\omega(g\mathbf{x})z(\omega, b - a, x_r) \leqq \varepsilon/m$ for all $b \in A$, so $a \in B(\hat{G}'(\mathbf{x}), x_r)$.

To show $\hat{G}'$ is uniformly consistent, we use the notation of Remark 2 in Section 4. Let $\alpha, \beta, \gamma, \delta$, be arbitrary positive numbers. Then for $N > N(\gamma, \delta)$, $P^\infty[\sum_{\omega=1}^m |G_\omega(\mathbf{Y}(\boldsymbol{\theta})) - G_\omega| > \gamma] < \delta$ for all $\boldsymbol{\theta}$ so that $\int \sum_{\omega=1}^m |\hat{G}_\omega(\mathbf{Y}(\boldsymbol{\theta})) - G_\omega| dP^\infty(\mathbf{Y}) < \gamma + 2\delta$ for all $\boldsymbol{\theta}$, since $\sum_{\omega=1}^m |\hat{G}_\omega - G_\omega| \leqq 2$. Since the $Y_i$ are i.i.d. and $G(g\boldsymbol{\theta}) = G(\boldsymbol{\theta})$ for all $g$ and $\boldsymbol{\theta}$ (where $G(\boldsymbol{\theta})$ is the empirical distribution of $\theta_1, \theta_2, \cdots, \theta_N$) we have

$$\int \sum_{\omega=1}^m |E\hat{G}_\omega(g\mathbf{Y}(g\boldsymbol{\theta})) - G_\omega| \, dP^\infty(\mathbf{Y})$$
$$\leqq E \int \sum_{\omega=1}^m |\hat{G}_\omega(g\mathbf{Y}(g\boldsymbol{\theta})) - G_\omega| \, dP^\infty(\mathbf{Y})$$
$$= E \int \sum_{\omega=1}^m |\hat{G}_\omega(\mathbf{Y}(g\boldsymbol{\theta})) - G_\omega| \, dP^\infty(\mathbf{Y}) < \gamma + 2\delta$$

for all $\boldsymbol{\theta}$ and $g$, where the equality uses the transformation theorem. Thus

$$P^\infty[\sum_{\omega=1}^m |E\hat{G}_\omega(g\mathbf{Y}(g\boldsymbol{\theta}) - G_\omega| > \alpha] < \frac{\gamma + 2\delta}{\alpha}.$$

Since $\hat{G}'(\mathbf{Y}(\boldsymbol{\theta})) = E\hat{G}(g\mathbf{Y}(g\boldsymbol{\theta}))$ for each $\mathbf{Y}$ and $\boldsymbol{\theta}$, we have that if $N > N(\gamma, \delta)$ with $\gamma + 2\delta < \alpha\beta$,

$$P^\infty[\sum_{\omega=1}^m |\hat{G}_\omega'(\mathbf{Y}(\boldsymbol{\theta})) - G_\omega| > \alpha] < \beta \quad \text{for all } \boldsymbol{\theta}.$$

Hence $\hat{G}'$ is uniformly consistent, and we can use it in place of $\hat{G}$.

Let $\boldsymbol{\phi} \in \Phi_2(\hat{G}')$ and let $W_r(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{x}) = L(\theta_r, \phi_r(\mathbf{x}), x_r)$. Then

$$W_{g^{-1}r}(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{x}) = L(\theta_{g^{-1}r}, \phi_{g^{-1}r}(\mathbf{x}), x_{g^{-1}r}) = L(\theta_{g^{-1}r}, \phi_r(g\mathbf{x}), x_{g^{-1}r})$$
$$= W_r(g\boldsymbol{\theta}, \boldsymbol{\phi}, g\mathbf{x}).$$

Since $E(h(g^{-1}N)) = N^{-1} \sum_{r=1}^{N} h(r)$ for any $h$, we have

$$R(\boldsymbol{\theta}, \boldsymbol{\phi}) = \int N^{-1} \sum_{r=1}^{N} W_r(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{x}) \prod_{k=1}^{N} f(\theta_k, x_k) \, d\mu^N(\mathbf{x})$$
$$= \int E[W_{g^{-1}N}(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{x}) \prod_{k=1}^{N} f(\theta_k, x_k)] \, d\mu_N(\mathbf{x})$$
$$= E[\int W_N(g\boldsymbol{\theta}, \boldsymbol{\phi}, g\mathbf{x}) \prod_{k=1}^{N} f(\theta_{g^{-1}k}, x_{g^{-1}k}) \, d\mu_N(x)]$$
$$= E[\int W_N(g\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{x}) \prod_{k=1}^{N} f(\theta_{g^{-1}k}, x_k) \, d\mu^N(x)] \,,$$

with the last step by the transformation theorem, since $\mu^N$ is invariant under permutations of $\mathbf{x}$. Since $E[\theta_{g^{-1}N} = \omega] = N_\omega/N$, we have

$$R(\boldsymbol{\theta}, \boldsymbol{\phi}) = \int E[L(\theta_{g^{-1}N}, \phi_N(\mathbf{x}), x_N) \prod_{k=1}^{N} f(\theta_{g^{-1}k}, x_k)] \, d\mu^N(\mathbf{x})$$

(5.2)
$$= \int E[E[L(\theta_{g^{-1}N}, \phi_N(\mathbf{x}), x_N) \prod_{k=1}^{N} f(\theta_{g^{-1}k}, x_k) \mid \theta_{g^{-1}N} = \omega]] \, d\mu_N(\mathbf{x})$$

$$= \int \sum_{\omega=1}^{m} \frac{N_\omega}{N} L(\omega, \phi_N(\mathbf{x}), x_N) f(\omega, x_N)$$

$$\times E[\prod_{k=1}^{N-1} f(\theta_{g^{-1}k}, x_k) \mid \theta_{g^{-1}N} = \omega] \, d\mu^N(\mathbf{x}) \,.$$

With $T(\phi_N(\mathbf{x}), \mathbf{x})$ as the integrand of the last expression and $\hat{B}(r) = B(\hat{G}'(\mathbf{x}), x_r)$, we have, for all $\boldsymbol{\phi} \in \Phi_2(\hat{G}')$,

(5.3)
$$R(\boldsymbol{\theta}, \boldsymbol{\phi}) \leqq \int \max_{a \in \hat{B}(N)} T(a, \mathbf{x}) \, d\mu^N(\mathbf{x}) \,.$$

We approximate the right side of (5.3) by the risk of a procedure which is both equivariant and polytope. Let $\boldsymbol{\zeta} \in \Phi_u(\hat{G}')$ be given by $\zeta_r(\mathbf{x})(a) = 1$ for $a = a_{\hat{B}(r)}(x_r)$, where $a_B(x)$ is the first maximizer in $B$ of $\sum_{\omega=1}^{m} G_\omega z(\omega, a, x)$. One might expect $\boldsymbol{\zeta}$ to do about as badly as possible against $\boldsymbol{\theta}$ since it "plays anti-Bayes" against $G$ within the restrictions imposed by membership of $\Phi_u(\hat{G}')$. That $\boldsymbol{\zeta}$ is equivariant follows easily from the fact that $\hat{G}'(g\mathbf{x}) = \hat{G}'(\mathbf{x})$ for each $\mathbf{x}$ and $g$, so from (5.2) and the definition of $T$,

$$\int \max_{a \in \hat{B}(N)} T(a, x) \, d\mu^N(\mathbf{x}) - R(\boldsymbol{\theta}, \boldsymbol{\zeta})$$

(5.4)
$$= \sum_{B \subset A} \int [\hat{B}(N) = B]\{\max_{a \in B} T(a, \mathbf{x}) - T(a_B(x_N), \mathbf{x})\} \, d\mu^N(\mathbf{x})$$

$$\leqq \sum_{B \subset A} \int \max_{a \in B} T(a, \mathbf{x}) - T(a_B(x_N), \mathbf{x}) \, d\mu^N(\mathbf{x}) \,,$$

since each integrand is positive. We now show the right side is $O(N^{-\frac{1}{2}})$.

For each $B$, consider the problem obtained from the present problem by truncating the action space to $B$ and using the loss function $L_B(\omega, a, x) = \sum_{b \in B} L(\omega, b, x) - L(\omega, a, x)$.

Replacing $L$ by $L_B$ in (5.2) and interchanging orders of summation, the risk of an equivariant procedure $\boldsymbol{\phi}$ in this new game is

(5.5)
$$R_B(\boldsymbol{\theta}, \boldsymbol{\phi}) = \int \sum_{b \in B} T(b, \mathbf{x}) - T(\psi_N(\mathbf{x}), \mathbf{x}) \, d\mu^N(\mathbf{x}) \,.$$

In particular, the best equivariant procedure has risk

(5.6)
$$R_B{}^*(G) = \int \sum_{b \in B} T(b, \mathbf{x}) - \max_{a \in B} T(a, \mathbf{x}) \, d\mu^N(\mathbf{x}) \,,$$

and the best simple symmetric procedure has risk

(5.7)
$$R_B(G) = \int \sum_{b \in B} T(b, \mathbf{x}) - T(a_B(x_N), \mathbf{x}) \, d\mu^N(\mathbf{x}) \,.$$

(Since simple symmetric procedures are equivariant we obtain (5.7) from (5.5) by taking $\phi$ to be degenerate at the first minimizer in $B$ of $\sum_{\omega=1}^{m} G_\omega L_B(\omega, a, x_r) \times f(\omega, x_r)$, which is $a_B(x_r)$, the first maximizer of $\sum_{\omega=1}^{m} G_\omega z(\omega, a, x_r)$ in $B$.)

Substituting into the left side of (5.4) the left of (5.3) and replacing the right of (5.4) by the difference of the left sides of (5.7) and (5.6) we obtain, for all $\phi \in \Phi_2(\hat{G}')$,

$$(5.8) \qquad R(\theta, \phi) - R(\theta, \zeta) \leqq \sum_{B \subset A} \{R_B(G) - R_B^*(G)\} .$$

Hannan and Huang [4] have shown that each summand on the right of (5.8) is bounded by $O(N^{-\frac{1}{2}})$ uniformly in $\theta$. Hence

$$(5.9) \qquad \sup_{\phi \in \Phi_2(\hat{G}')} \sup_\theta D(\theta, \phi) \leqq \sup_\theta \{R(\theta, \zeta) - R(G)\} + O(N^{-\frac{1}{2}}) .$$

We now show that $\zeta$ is a polytope procedure. If $\zeta_x$ is the function corresponding to "$\phi_x$" in Definition 6, then

$$\{\zeta_x(y) : y \in \mathscr{H}\} \subset \{a_1, a_2, \cdots, a_n\}$$

(where "$a$" denotes the measure degenerate at $a$) for each $x$. We need to show that $\{y : \zeta_x(y) = a\} \in \mathscr{K}_s^t$, for some $t$ and $s$, for each $x$ and $a$. But $\{y : \zeta_x(y) = a\} = \bigcup_{\{B \subset A : a \in B\}} Q_{FB} \cap R_N$ where $F = \hat{G}(x)$, and $R_N = \{x : a_B(x) = a\} = \bigcap_{b \in B, b < a} \{x : \sum_{\omega=1}^{m} G_\omega z(\omega, b - a, x) < 0\} \cap \bigcap_{b \in B, b > a} \{x : \sum_{\omega=1}^{m} G_\omega z(\omega, b - a, x) \leqq 0\} \in \mathscr{K}_{n-1}$. Since we already have $Q_{FB} \in \mathscr{K}_s^t$ from Lemma 3, the result follows easily.

Hence $\zeta \in \Phi_1(\hat{G})$ with $v = n$, so we can apply the corollary to Theorem 1 to the first term on the right of (5.9), completing the proof of Theorem 2.

**6. Infinite action spaces.** We now weaken the requirement that $A$ be finite to one which allows $A$ to be approximated by a finite subset.

THEOREM 3. *Let $\hat{G}$ be uniformly consistent and let $B = \bigcup_{F \in \mathscr{G}, x \in \mathscr{X}} B(F, x)$ (see Definition 5 and the sentence preceding Definition 6) be totally bounded in the metric $d(a, a') = \sup_{x, \omega} |L(\omega, a - a', x)|$. Then*

$$\sup_{\Phi_2(\hat{G})} \sup_\theta D(\theta, \phi) < \varepsilon + o(1) \quad as \quad N \to \infty .$$

PROOF. For each $\delta > 0$, let $D_\delta = \{a_1, a_2, \cdots, a_k\}$, $k = k(\delta)$, be such that, for any $a \in B$, $d(a, a_j) < \delta$ for some $a_j \in D_\delta$. Let $\phi \in \Phi_2(\hat{G})$, fix $\delta$, and let $\{A_j : j = 1, 2, \cdots, k\}$ be a partition of $B$ such that, for each $j$, $d(a, a_j) < \delta$ for every $a \in A_j$.

Consider the problem obtained by replacing $A$ by $D_\delta$. For this problem, let $\Phi_{u\delta}(\hat{G})$ and $\Phi_{2\delta}(\hat{G})$ satisfy Definitions 6 and 8, respectively, with "$\varepsilon$" replaced by "$\varepsilon + m\delta$"; and let $R_\delta(\cdot)$ be the Bayes envelope.

We observe that $\min_{D_\delta} \sum_{\omega=1}^{m} F_\omega z(\omega, a_j, x) - \inf_A \sum_{\omega=1}^{m} F_\omega z(\omega, a, x) < \delta \max_\omega f(\omega, x) < \delta$ for any $F \in \mathscr{G}$, since $f(\omega, x) \leqq 1$. Integrating with respect to $\mu$ we have, for all $F \in \mathscr{G}$,

$$(6.1) \qquad R_\delta(F) - R(F) < m\delta .$$

Let $\zeta = \zeta(\phi)$ be the procedure in the reduced game given by $\zeta_r(x)(a_j) = \phi_r(x)(A_j)$ for all $r$, $x$ and $j$.

Since $\zeta_r(g\mathbf{x})(a_j) = \phi_r(g\mathbf{x})(A_j) = \phi_{g^{-1}r}(\mathbf{x})(A_j) = \zeta_{g^{-1}r}(x)(a_j)$, $\zeta$ is equivariant. We will show that $\zeta \in \Phi_{u\delta}(\hat{G})$.

Let $\phi_\mathbf{x}$ be the function given by Definition 6. For each $\mathbf{x} \in \mathscr{X}^N$, $y \in \mathscr{X}$ and $a_j \in D_\delta$, let $\zeta_\mathbf{x}(y)(a_j) = \phi_\mathbf{x}(y)(A_j)$. Then, for each $\mathbf{x}$ and $r$, $\zeta_r(\mathbf{x}) = \zeta_\mathbf{x}(x_r)$. If $\zeta_\mathbf{x}(y)(a_j) > 0$ for some $\mathbf{x}$, $y$ and $j$, then $\phi_\mathbf{x}(y)(A_j) > 0$. Since $\phi_\mathbf{x} \in \Phi_u(\hat{G}(\mathbf{x}))$, there is an $a \in A_j$ for which $\sum_{\omega=1}^{m} \hat{G}_\omega(\mathbf{x})z(\omega, a -- b, y) \leqq \varepsilon/m$ for all $b \in A$. By definition of $A_j$ this implies that $\sum_{\omega=1}^{m} \hat{G}_\omega(\mathbf{x})z(\omega, a_j - b, y) \leqq \varepsilon/m + \delta$ for all $b \in A$ (and hence for all $b \in D_\delta$). Hence $\zeta_\mathbf{x}(y)(a_j) > 0$ implies $a_j \in B_\delta(\hat{G}(\mathbf{x}), y, \varepsilon + m\delta)$ where $B_\delta(G, x, \varepsilon)$ satisfies (5.1) when $A$ is replaced by $D_\delta$. Hence $\zeta_\mathbf{x} \in \Phi_{u\delta}(\hat{G}(\mathbf{x}))$ and consequently $\zeta \in \Phi_{2\delta}(\hat{G})$.

By definition of $\zeta$ we have, for all $\theta$ and $\mathbf{x}$,

$$|W(\theta, \phi, \mathbf{x}) - W(\theta, \zeta, \mathbf{x})| \leqq N^{-1} \sum_{r=1}^{N} |W_r(\theta, \phi, \mathbf{x}) - W_r(\theta, \zeta, \mathbf{x})|$$
$$\leqq N^{-1} \sum_{r=1}^{N} \sum_{j=1}^{n} |\int_{A_j} L(\theta_r, a, x_r)\phi_r(\mathbf{x})(da) - L(\theta_r, a_j, x_r)\phi_r(\mathbf{x})(A_j)|$$
$$\leqq N^{-1} \sum_{r=1}^{N} \sum_{j=1}^{n} \int_{A_j} |L(\theta_r, a, x_r) - L(\theta_r, a_j, x_r)|\phi_r(\mathbf{x})(da)$$
$$\leqq N^{-1} \sum_{r=1}^{N} \sum_{j=1}^{n} \delta\phi_r(\mathbf{x})(A_j) = \delta .$$

Integrating this inequality with respect to $P_\theta$, we obtain

(6.2) $$R(\theta, \phi) - R(\theta, \zeta) < \delta \quad \text{for all} \quad \theta .$$

Thus, for any $\delta > 0$,

(6.3) $$\sup_{\Phi_2(\hat{G})} \sup_\theta D(\theta, \phi) \leqq \sup_{\Phi_2(\hat{G})} \sup_\theta \{|R(\theta, \phi) - R(\theta, \zeta(\phi))|$$
$$+ |R(\theta, \zeta(\phi)) - R_\delta(G)| + |R_\delta(G) - R(G)|\}$$
$$< \delta + \sup_{\Phi_{2\delta}(\hat{G})} \sup_\theta |R(\theta, \zeta) - R_\delta(G)| + m\delta$$

by (6.1) and (6.2).

From Theorem 2 there is a function $N_\delta(\gamma) = N(\gamma, \delta)$, such that $N > N(\gamma, \delta)$ implies

$$\sup_{\Phi_{2\delta}(\hat{G})} \sup_\theta |R(\theta, \zeta) - R_\delta(G)| < \gamma + \varepsilon + m\delta .$$

Substituting this in (6.3), with $\delta(\gamma) = \gamma/2(2m + 1)$, we have, for $N > N'(\gamma) = N(\gamma/2, \delta(\gamma))$, $\sup_{\Phi_2(\hat{G})} \sup_\theta D(\theta, \phi) < \gamma + \varepsilon$ as required.

## II. INFINITE STATE SPACES

In this part, $\Omega$ is infinite. We assume the existence of "good" estimators of $G$ and, with a number of conditions on the loss and density functions, we exhibit, for arbitrary $\delta > 0$, procedures which are a slight modification of the "$\varepsilon$-Bayes against the estimate" type and for which $D(\theta, \phi) < \varepsilon + \delta + o(1)$ as $N \to \infty$. We shall show that the conditions we require, except for the conditions concerning estimators, are satisfied in many situations.

**7. Finitely based procedures.** The methods of I are inappropriate when $\Omega$ is infinite because appropriate forms of Lemma 2 are not available, due to the partial failure of the Glivenko-Cantelli theorem in infinite dimensional spaces.

The convergence for which Lemma 2 was used could be expected if the sample space, $\mathscr{X}$, were finite but the problem of estimating $G$ would then be virtually incapable of solution since the distributions $\{P_\omega : \omega \in \Omega\}$ would not be linearly independent, so that the values of $\int P_\omega(A)\, dG(\omega)$ for each $A \in \mathscr{X}$ would not determine $G$ uniquely. Thus a finite sample space seems to be desirable, for convergence results analogous to those of I, and an infinite sample space necessary, to estimate $G$; these considerations motivate the approach of this part.

Let $\pi$ be a finite measurable partition of $\mathscr{X}$. For each $x \in \mathscr{X}$, let $x'$ be the member of $\pi$ to which $x$ belongs and for each $V \in \pi$, let $L(\omega, a, V)$ be the value of $L(\omega, a, x)$ at a fixed but arbitrary point $x \in V$. As in Part I, let $\mathscr{G}$ be the set of distributions on $\Omega$.

DEFINITION 9. For each $\omega \in \Omega$ and $V \in \pi$, let $P_{\omega\pi}(\{V\}) = P_\omega(V)$. For the component game obtained by replacing $\mathscr{X}$ by $\pi$, $P_\omega$ by $P_{\omega\pi}$ and $L(\omega, a, x)$ by $L(\omega, a, V)$, let $R_\pi(\cdot)$ be the Bayes envelope, and $R_\pi(F, \phi)$ the risk against $F \in \mathscr{G}$ of a procedure $\phi \in \Phi_\pi$, the set of procedures for which $L(\omega, \phi(V), V)$ is $\omega$-measurable for each $V \in \pi$. For each $F \in \mathscr{G}$, let $\Phi_\pi(F)$ be the set of procedures $\varepsilon$-Bayes against $F$.

Component procedures available in the reduced problem are available in the original problem in the sense that, if $\phi \in \Phi_\pi$, the procedure $\psi$ in the original game, given by $\psi(x) = \phi(x')$ for every $x$, can be identified with $\phi$.

For each $E$ and $F \in \mathscr{G}$, let

(7.1)     $\lambda(E, F) = \sup_{\nu, \phi \in \Phi_\pi} \sum_{V \in \pi} \int L(\omega, \nu(V) - \phi(V), V) P_\omega(V)\, d(F - E)(\omega)$.

Then, from Lemma 1 of I, for each $E$ and $F \in \mathscr{G}$ and $\phi \in \Phi_\pi$,

(7.2)     $R_\pi(F, \phi) - R_\pi(F) \leq \lambda(E, F) + R_\pi(E, \phi) - R_\pi(E)$.

The idea in what follows is that, if $\pi$ is a "good" approximation to $\mathscr{X}$ as regards loss and density functions, we might use $x_1, \cdots, x_N$ to estimate $G$ but use only $x_1', x_2', \cdots, x_N'$ when we play "$\varepsilon$-Bayes against the estimate".

**8. Convergence theorems.** In this section we prove convergence results for the following class of procedures:

DEFINITION 10. For $\hat{G}$ an estimator, let $\Phi_\pi(\hat{G}) = \{\phi : \forall \mathbf{x}, \exists \phi_\mathbf{x} \in \Phi_\pi(\hat{G}(\mathbf{x}))$ such that, $\forall r, \phi_r(\mathbf{x}) = \phi_\mathbf{x}(x_r)\}$.

The results of this section are based on the inequality

(8.1)     $|R(\theta, \phi) - R(G)| \leq |R(\theta, \phi) - R_\pi(G)| + |R_\pi(G) - R(G)|$

for any $\phi \in \Phi_\pi(\hat{G})$. In Lemmas 4 and 5 we deal with the first term on the right; the second term is handled in Lemma 6.

LEMMA 4. *Let $\Omega$ be totally bounded in the metric*

$$d(\omega, \omega') = \sup\{|L(\omega, a, V) - L(\omega', a, V)| : a \in A, V \in \pi\}.$$

*Let $L(\omega, a, x) \leqq M < \infty$ for all $\omega$, $a$ and $x$, and let*

$$\delta(\pi) = \sup\{|L(\omega, a, x) - L(\omega, a, y)| : \omega \in \Omega, a \in A, x' = y'\}\,.$$

*Then, with $\delta = \delta(\pi)$, $\eta > 0$ and $\gamma > 0$ there exists $N(\eta, \gamma)$ such that*

$$P_\theta[\sup_{N > N(\eta, \gamma)}\{W(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{x}) - \lambda(\hat{G}(\mathbf{x}), G)\} > R_\pi(G) + \varepsilon + 5\delta + \eta] < \gamma\,,$$

*for all $\boldsymbol{\theta} \in \Omega^\infty$ and all $\boldsymbol{\phi} \in \Omega_\pi(\hat{G})$.*

PROOF. Let $E_1, E_2, \cdots, E_m$ be a partition of $\Omega$ by sets of diameter $< \delta$ so that, for each $i$, $\sup_a |L(\omega, a, x) - L(\omega', a, x')| < 2\delta$ whenever $\omega$, $\omega' \in E_i$. For each $i$ let $\omega_i$ be an arbitrary fixed element of $E_i$.

The proof is similar to that of Theorem 1. Let $\psi \in \Phi_\pi(F)$ for some $F$. Then for any $\boldsymbol{\theta}$, since $|L(\omega, \psi(x'), x) - L(\omega_i, \psi(x'), x')| \leqq 2\delta$,

$$(8.2) \qquad |W(\theta, \psi, \mathbf{x}) - N^{-1} \sum_{i=1}^m \sum_{\{r:\, \theta_r \in E_i\}} L(\omega_i, \psi(x_r'), x_r')| < 2\delta\,.$$

Let $N_i = N(\boldsymbol{\theta}, i) = \sum_{i=1}^N [\theta_r \in E_i]$, $\bar{P}_{N_i} = N_i^{-1} \sum_{r=1}^N [\theta_r \in E_i] P_{\theta_r \pi}$, and for each $D \in \pi$, let $P_{N_i}(D) = N_i^{-1} \sum_{r=1}^N [\theta_r \in E_i][x_r' \in D]$, so that $\bar{P}_{N_i}$ is the "average" distribution on $\pi$ arising from the $\theta_r$'s in $E_i$, and $P_{N_i}$ is the corresponding empirical distribution given by $\{x_r' : \theta_r \in E_i\}$. Then (8.2) becomes

$$(8.3) \qquad \left| W(\boldsymbol{\theta}, \psi, \mathbf{x}) - \sum_{i=1}^m \frac{N_i}{N} \int L(\omega_i, \psi)\, dP_{N_i} \right| < 2\delta$$

which implies $|R(G, \psi) - \sum_{i=1}^m (N_i/N) \int L(\omega_i, \psi)\, d\bar{P}_{N_i}| < 2\delta$. But $|L(\omega, \psi(x'), x) - L(\omega, \psi(x'), x')| \leqq \delta$, so $|R(G, \psi) - R_\pi(G, \psi)| \leqq \delta$. Hence

$$(8.4) \qquad \left| R_\pi(G, \psi) - \sum_{i=1}^N \frac{N_i}{N} \int L(\omega_i, \psi)\, d\bar{P}_{N_i} \right| < 3\delta\,.$$

From (8.3) and (8.4) we have

$$(8.5) \qquad |W(\boldsymbol{\theta}, \psi, \mathbf{x}) - R_\pi(G, \psi)| \leqq \frac{N_i}{N} \int L(\omega_i, \psi)\, d|P_{N_i} - \bar{P}_{N_i}| + 5\delta\,.$$

Let $\boldsymbol{\phi} \in \Phi_\pi(\hat{G})$ and let $\phi_{\mathbf{x}}$ be the function guaranteed by Definition 10. Since $W(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{x}) = W(\boldsymbol{\theta}, \phi_{\mathbf{x}}, \mathbf{x})$ for each $\mathbf{x}$, we have, from (8.5) and using (7.2) and the definition of $\Phi_\pi(F)$ to bound the random variable $R_\pi(G, \phi_{\mathbf{x}})$,

$$(8.6) \qquad |W(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{x}) - R_\pi(G)| \leqq \sum_{i=1}^m \frac{N_i}{N} S(N, \boldsymbol{\theta}, i, \mathbf{x}) + \lambda(\hat{G}(\mathbf{x}), G) + 5\delta + \varepsilon$$

where $S(N, \boldsymbol{\theta}, i, \mathbf{x}) = M \sum_{V \in \pi} |(P_{N_i} - \bar{P}_{N_i})(V)|$.

$P_{N_i}(V)$ is the average of $N_i$ independent Bernoulli random variables, and $\bar{P}_{N_i}(V)$ is the mean of this average; from this it is easy to show that $E|(P_{N_i} - \bar{P}_{N_i})(V)|^4 = \int |(P_{N_i} - \bar{P}_{N_i})(V)|^4\, dP_\theta < 6N_i^{-2}$, so if $\pi$ has $q$ elements

$$(8.7) \qquad P_\theta[\sup_{N > N'} S(N, \boldsymbol{\theta}, i, \mathbf{x}) > \eta] \leqq \sum_{n=N'}^\infty \sum_{V \in \pi} P_\theta[|(P_{n_i} - \bar{P}_{n_i})(V)| > \eta/qM]$$

$$\leqq \sum_{n=N_i'}^\infty 6q^5 M^4 \eta^{-4} n^{-2} = T(N_i')\,, \qquad \text{say,}$$

by the Markov inequality. Hence

$$P_\theta\left[\sup_{N>N'} \sum_{i=1}^m \frac{N_i}{N} S(N, \boldsymbol{\theta}, i, \mathbf{x}) > \eta\right]$$

$$\leqq \sum_{i=1}^m P_\theta\left[\sup_{N>N'} \frac{N_i}{N} S(N, \boldsymbol{\theta}, i, \mathbf{x}) > \eta/m\right]$$

$$\leqq \sum_{i=1}^m \left\{P_\theta[\sup_{N_i>k} S(N, \boldsymbol{\theta}, i, \mathbf{x}) > \eta/m]\right.$$

$$\left. + P_\theta\left[\frac{k}{N'} \sup_{N_i<k} S(N, \boldsymbol{\theta}, i, \mathbf{x}) > \eta/m\right]\right\}$$

for any $k$. The first summand is $m^4 T(k)$; the second $= 0$ if $N' > 2Mkm/\eta$; so the sum $\to 0$ as $N' \to \infty$, proving the Lemma.

We shall need the following results in the proof of the next lemma. Only outline proofs are given here; details can be found on pages 44–48 of [9]. For both, $h$ is a real-valued function on $\Omega$ and $\alpha_h(\rho) = \sup\{h(\omega) - h(\omega') : d(\omega, \omega') < \rho\}$, where $d$ is a metric on $\Omega$.

(a) The Prohorov metric on $\mathscr{C}$, the set of probability distributions on a metric space $(\Omega, d)$, is given by

$$H^*(E, F) = \inf\{\delta : E(A^\delta) + \delta \geqq E(A) \text{ for all closed } A \subset \Omega\}$$

where $A^\delta = \{\omega : \text{for some } \omega' \in A, d(\omega, \omega') < \delta\}$.

If $a \leqq h(\cdot) \leqq a + M$ then $|\int hd(E - F)| \leqq M\rho + \alpha_h(\rho)$, where $\rho = H^*(E, F)$. The proof is accomplished by showing

$$\int hd(E - F) = \int_0^M Eh^{-1}[x > t] - Fh^{-1}[x > t]\, dt$$

$$\leqq \int_0^M Fh^{-1}[t - \alpha_h(\rho) \leqq x \leqq t] + \rho\, dt,$$

from which the result follows using Fubini's theorem.

(b) The Lévy metric on the set of probability distributions on the real line is given by

$$H(E, F) = \inf\{\delta : E(x - \delta) - \delta < G(x) < F(x + \delta) + \delta \text{ for all } x\}.$$

If $\Omega \subset [a, b]$, $c \leqq h(\cdot) \leqq c + M$ and $\lambda > 2\rho = 2H(E, F)$, then

$$|\int hd(E - F)| < M\left[\frac{b - a}{\lambda} + 1\right]\rho + \alpha_h(\lambda + \rho) + \alpha_n(\lambda + 2\rho),$$

where $[x]$ is the integer part of $x$.

The proof of this is more complicated. The essential details are: construct two partitions of $[a, b]$, $\{x_j\} = \{a + j\sigma, j = 1, \cdots, k\}$, and $x_0' < \cdots < x_k'$, where $k = [(b - a)/\lambda + 1]$, $|x_j - x_j'| \leqq \rho$ and $E(x_j'-) - \rho < F(x_j) < E(x_j') + \rho$. With $y_j = \min\{x_j, x_j'\}$ and $z_j = \max\{x_j, x_j'\}$, let $h_1(x) = \inf\{h(\omega) : \omega \in [y_j, z_{j+1}] \cap \Omega\}$ for $x \in (x_j, x_{j+1})$. Let $h_2(x)$ take on the same value for $x \in (x_j', x_{j+1}')$, with $h_2(x_j') = \max\{h_2(x_j'-), h_2(x_j'+)\}$. It can then be shown that $|h - h_1| \leqq \alpha_h(\lambda + \rho)$, $|h - h_2| \leqq \alpha_h(\lambda + 2\rho)$ and $\int h_2 dE - \int h_1 dF \geqq -Mk\rho$.

LEMMA 5. *If either* (a) $(\Omega, d)$ *is a metric space and* $H^*(\hat{G}(\mathbf{x}), G) \to 0$ *a.e.* $[P_\theta]$ *as* $N \to \infty$ *for each* $\theta \in \Omega$, *where* $H^*$ *is the Prohorov metric, or* (b) $\Omega$ *is a subset of the real line and* $H(\hat{G}(\mathbf{x}), G) \to 0$ *a.e.* $[P_\theta]$ *as* $N \to \infty$ *for each* $\theta \in \Omega^\infty$ *where* $H$ *is the Lévy metric; and if, in addition,* $\Omega$ *and* $A$ *are compact,* $L$ *is jointly continuous in* $\omega$ *and* $a$ *for each fixed* $x$, *and* $P_\omega(V)$ *is continuous in* $\omega$ *for each* $V \in \pi$, *then* $\lambda(\hat{G}(\mathbf{x}), G) \to 0$ *a.e.* $[P_\theta]$ *for each* $\theta$; *and the convergence is uniform in* $\theta$ *if that of* $\hat{G}$ *to* $G$ *is.*

PROOF. For any $\nu$ and $\psi \in \Phi_\pi$ and $V \in \pi$,

$$(8.8) \qquad |L(\omega, \nu(V) - \psi(V), V)P_\omega(V) - L(\omega', \nu(V) - \psi(V), V)P_{\omega'}(V)|$$
$$\leqq 2 \sup_a |L(\omega, a, V) - L(\omega', a, V)| + M|P_{\omega'}(V) - P_\omega(V)|$$

where $M = \sup_{\omega, a, V} L(\omega, a, V) < \infty$. The right side of (8.8) $\to 0$ uniformly as $\omega' \to \omega$, because of the compactness and continuity conditions. Thus, with $\mathscr{H} = \{L(\omega, \nu(V) - \psi(V), V)P_\omega(V) : V \in \pi; \psi, \nu \in \Phi_\pi\}$ we have $\sup_{\mathscr{H}} \alpha_h(\rho) \to 0$ as $\rho \to 0$.

The proof is completed by appealing to results (a) and (b) above.

Henceforth we assume that $\Omega$ is separable in the metric $d_1(\omega, \omega') = \sup_{A \in \mathscr{B}} |P_\omega(A) - P_{\omega'}(A)|$, so the set of $\sigma$-finite measures $\mathscr{M} = \{\mu : P_\omega \ll \mu$ for all $\omega \in \Omega\}$ is nonempty.

DEFINITION 11. Let $\alpha(\pi) = \inf_{\mathscr{M}} \alpha_\mu(\pi)$, where $\alpha_\mu(\pi) = \sup_\omega \alpha_{\mu\omega}(\pi) = \sup_\omega \int (f_\omega - f_{\omega\pi})^+ d\mu$, with $f_\omega = dP_\omega/d\mu$ and $f_{\omega\pi}(x) = P_\omega(x')/\mu(x')$ for each $x$.

DEFINITION 12. For each $G \in \mathscr{G}$, let $R_1(G) = \inf R(G, \phi)$, where the infimum is taken over all procedures $\phi$ for which $R(G, \phi)$ exists and $\int\int L(\omega, \phi)f_\omega d\mu \, dG(\omega) = \int\int L(\omega, \phi)f_\omega \, dG(\omega) \, d\mu$.

REMARK. The class of procedures for which this change of order is valid does not depend on $\mu$, since (i) $\mu$ can be taken to be equivalent to $\{P_\omega : \omega \in \Omega\}$ because of the separability, and (ii) if $\mu \ll \nu$, $\int\int dP_\omega/d\nu L(\omega, \phi) \, dG(\omega) \, d\nu = \int d\mu/d\nu \times \int f_\omega L(\omega, \phi) \, dG(\omega) \, d\nu = \int\int f_\omega L(\omega, \phi) \, dG(\omega) \, d\mu$.

LEMMA 6. *Let* $L(\omega, a, x) \leqq M$ *for all* $\omega$, $a$ *and* $x$, *and let* $\delta = \delta(\pi)$ *be as for Lemma 4. Then with* $\alpha = \alpha(\pi)$, $R_\pi(F) - R_1(F) \leqq M\alpha + \delta$ *for all* $F \in \mathscr{G}$.

PROOF. Let $\mu$ be any measure with $P_\omega \ll \mu$ for each $\omega$, let $f_\omega$ and $f_{\omega\pi}$ be as in Definition 11, and let $\phi$ be such that $\int\int L(\omega, \phi)f_\omega d\mu \, dF(\omega) = \int\int L(\omega, \phi)f_\omega \, dF(\omega) \, d\mu$. Then

$$R(F, \phi) \geqq \int\int \{L(\omega, \phi(x), x') - \delta\}f_\omega(x) \, dF(\omega) \, d\mu(x)$$
$$= \int\int L(\omega, \phi(x), x')f_\omega(x) \, dF(\omega) \, d\mu(x) - \delta \,.$$

Since $L \leqq M$ and $f_\omega(x) \geqq f_{\omega\pi}(x) - (f_{\omega\pi}(x) - f_\omega(x))^+$, we have

$$(8.9) \qquad R(F, \phi) + \delta \geqq \int\int L(\omega, \phi(x), x')f_{\omega\pi}(x) \, dF(\omega) \, d\mu(x)$$
$$- M \int\int (f_{\omega\pi} - f_\omega)^+ \, d\mu \, dF(\omega) \,.$$

The first term on the right is bounded below by

(8.10)  $\int \inf_{a \in A} \int L(\omega, a, x') f_{\omega\pi}(x) \, dF(\omega) \, d\mu(x)$
$$\geq R_\pi(F) - M \sum_{\mu(V)=\infty} \int P_\omega(V) \, dF(\omega) \, .$$

To deal with the second term on the right of (8.9) we note that

$$\int (f_\omega - f_{\omega\pi})^+ \, d\mu - \int (f_{\omega\pi} - f_\omega)^+ \, d\mu = \sum_{\mu(V)=\infty} \int V f_\omega \, d\mu + \sum_{\mu(V)<\infty} \int V(f_\omega - f_{\omega\pi}) \, d\mu$$
$$= \sum_{\mu(V)=\infty} P_\omega(V) \, .$$

Thus, from Definition 11, $\alpha_\mu(\pi) - \sum_{\mu(V)=\infty} P_\omega(V) \geq \int (f_{\omega\pi} - f_\omega)^+ \, d\mu$.

Combining this with (8.9) and (8.10), $R(F, \phi) + \delta \geq R_\pi(F) - M\alpha_\mu(\pi)$. Since $\phi$ and $\mu$ are arbitrary, the proof is complete.

COROLLARY. $R(G) \geq R_\pi(G) - M\alpha - \delta$ for all $\boldsymbol{\theta} \in \Omega^\infty$. This follows since $G$ is discrete so the change of order in Definition 12 is valid when either side exists.

We now combine the conditions of Lemmas 4, 5 and 6 to obtain

THEOREM 4. If $\Omega \times A$ is compact in a topology in which $P_\omega(V)$ is continuous and $L(\omega, a, V)$ jointly continuous for each $V \in \pi$ and either (a) or (b) of Lemma 5 holds, then there exists a function $N(\eta, \gamma, \boldsymbol{\theta})$ such that, for each $\boldsymbol{\phi} \in \Phi_\pi(\hat{G})$ and $\boldsymbol{\theta} \in \Omega^\infty$,

$$P_\theta[\sup_{N > N(\eta, \gamma, \theta)} W(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{Y}) - R(G) > \varepsilon + M\alpha + 6\delta + \eta] < \gamma \, .$$

In addition, if the convergence of $\hat{G}$ to $G$ is uniform, $N(\eta, \gamma, \boldsymbol{\theta}) = N(\eta, \gamma)$.

COROLLARY. Under the conditions of Theorem 1, $\sup_{\Phi_\pi(\hat{G})} D(\boldsymbol{\theta}, \boldsymbol{\phi}) < o(1) + \varepsilon + M\alpha + 6\delta$ as $N \to \infty$, for each $\boldsymbol{\theta} \in \Omega^\infty$; this bound is uniform in $\boldsymbol{\theta}$ if the convergence of $\hat{G}$ is uniform.

PROOF. Given $\eta > 0$ we have, by Theorem 4,

$$R(\boldsymbol{\theta}, \boldsymbol{\phi}) = \int W(\boldsymbol{\theta}, \boldsymbol{\phi}) \, dP_\theta \leq R(G) + \varepsilon + \eta + M\alpha + 6\delta$$

if $N > (\eta/2, \eta/2M, \boldsymbol{\theta})$, since $W \leq M$.

We remark, in concluding this section, that none of the results depend essentially on the particular determination of $L(\omega, a, x')$, so any determination will suffice.

**9. Approximating the sample space by a finite partition.** The usefulness of Theorem 4 and its corollary depends on the availability of partitions for which $\alpha(\pi)$ and $\delta(\pi)$ are arbitrarily small.

LEMMA 7. Let $P_\omega \ll \mu$ with $dP_\omega/d\mu = f_\omega$. Then for each $\omega$ and $\alpha > 0$ there is a partition $\pi_\omega$ such that $\alpha_{\mu\omega}(\pi) \leq \alpha$ whenever $\pi$ is a sub-partition of $\pi_\omega$.

PROOF. Choose $a_1$ so that $P_\omega(f_\omega^{-1}[0, a_1]) < \alpha/3$. (We assume $\alpha < 3$.) Let $a_0 = 0$, and $a_j = 3a_{j-1}/(3 - \alpha) = (3/(3 - \alpha))^{j-1} a_1$ for $1 < j \leq k$ where $k = \min \{j : P_\omega(f_\omega^{-1}[a_j, \infty)) < \alpha/3\}$. Let $a_{k+1} = \infty$ and let $\pi_\omega = \{V_0 \cdot V_1, \cdots, V_k\}$ where $V_j = f_\omega^{-1}[a_j, a_{j+1}]$.

Now let $V \subset V_j$, $1 < j < k$; and $x \in V$. Then

$$(f_\omega(x) - P_\omega(V)/\mu(V))^+ \leqq f_\omega(x)(1 - a_{j-1}/a_j) \leqq \alpha f_\omega(x)/3 .$$

Hence if $\pi$ is a sub-partition of $\pi_\omega$,

$$\alpha_{\mu\omega}(\pi) = \sum_{V \in \pi} \int V(f_\omega - P_\omega(V)/\mu(V))^+ \, d\mu$$
$$= \sum_{j=1}^k \sum_{V \subset V_j} \int Vf_\omega \alpha/3 \, d\mu + d\mu + \sum_{V \subset V_0 \cup V_k} \int Vf_\omega d\mu < \alpha .$$

LEMMA 8. *If $\Omega$ is totally bounded in the metric $d(\omega, \omega') = \sup |P_\omega(A) - P_{\omega'}(A)|$, then for any $\alpha > 0$ there is a partition, $\pi$, of $\mathscr{H}$ for which $\alpha(\pi) \leqq \alpha$.*

PROOF. Since $\Omega$ is totally bounded in $d$ it is separable, so for some $\sigma$-finite $\mu$, $P_\omega \ll \mu$ for all $\omega \in \Omega$. Let $f_\omega = dP_\omega/d\mu$. Then $d(\omega, \omega') = \frac{1}{2} \int |f_\omega - f_{\omega'}| \, d\mu$.

By Lemma 7 we can find, for each $\omega$, a partition $\pi_\omega$ such that $\alpha_{\mu\omega}(\pi) \leqq \alpha/2$ whenever $\pi$ is a sub-partition of $\pi_\omega$. For any $\omega$, $\omega'$ and any $\pi$

$$\sum_\pi \int V|(f_\omega - P_\omega(V)/\mu(V))^+ - (f_{\omega'} - P_{\omega'}(V)/\mu(V))^+| \, d\mu$$
$$< \sum_\pi \{ \int V|f_\omega - f_{\omega'}| \, d\mu + \int V|P_\omega(V)/\mu(V) - P_{\omega'}(V)/\mu(V)| \, d\mu \} ,$$

the second summand on the right being bounded by $|P_\omega(V) - P_{\omega'}(V)|$ which is $\leqq \int V|f_\omega - f_{\omega'}| \, d\mu$. Hence $\alpha_{\mu\omega}(\pi) - \alpha_{\mu\omega'}(\pi) \leqq 2 \int |f_\omega - f_{\omega'}| \, d\mu = 4d(\omega, \omega')$.

Let $U_1, \cdots, U_k$ be a covering of $\Omega$ by spheres of $d$-diameter $\leqq \alpha/8$ with $\omega_i \in U_i$, $i = 1, 2, \cdots, k$, arbitrary. Then if $\pi$ is any finite sub-partition of $\pi_{\omega_1}, \cdots, \pi_{\omega_k}$, we have $\alpha_{\omega\mu}(\mu) \leqq \alpha_{\omega_i\mu}(\pi) + 4d(\omega, \omega_i) \leqq \alpha$ for $\omega \in U_i$. Hence $\alpha_\mu(\pi) \leqq \alpha$. Since $\alpha(\pi) \leqq \alpha_\mu(\pi)$ the lemma is proved.

REMARK 1. If $\alpha(\pi) \leqq \alpha$ for some $\pi$ and $\delta(\pi_1) \leqq \delta$ for some $\pi_1$, then $\alpha(\pi_2) \leqq \alpha$ and $\delta(\pi_2) \leqq \delta$ for any sub-partition, $\pi_2$, of $\pi$ and $\pi_1$. Obviously $\delta(\pi) = 0$ for all $\pi$ if $L$ is independent of $x$.

REMARK 2. The condition of Lemma 8 holds for many families. Scheffé's theorem, that $\int |f_\omega - f_{\omega'}| \, d\mu \to 0$ if, for every $F$ with $\mu(F) < \infty$, $\mu[F \cap \{x : |f_\omega(\omega) - f_{\omega'}(x)| > \eta\}] \to 0$ for all $\eta > 0$, serves to establish total boundedness—usually compactness—in many cases.

For example, let $T$ map $\mathscr{H}$ into $E^k$, $\mu$ be a $\sigma$-finite measure on $\mathscr{H}$ and $\Theta = \{\omega \in E^k : \int e^{\omega T} \, d\mu < \infty\}$ where $\omega T$ is an inner product. The class of densities $\{C(\omega)e^{\omega T(x)} : \omega \in \Theta\}$, where $C(\omega) = [\int e^{\omega T} \, d\mu]^{-1}$, is the exponential family on $\mathscr{H}$ generated by $T$ and $\mu$. The well-known continuity of $C$ in $\Theta^0$, the interior of $\Theta$, implies $f_\omega(x) \to f_{\omega'}(x)$ for all $x$ and all $\omega' \in \Theta^0$, so by Scheffé's theorem any compact subset of $\Theta^0$ satisfies the condition of Lemma 8.

Again, let $\int p \, d\nu = 1$ where, for $\mu$ Lebesgue measure on $E^k$, $\nu \ll \mu$ and $d\nu/d\mu$ is bounded. Let $f_\omega(x) = p(x - \omega)$, $\omega \in E^k$. Then $\int |f_\omega - f_{\omega'}| \, d\nu \to 0$ uniformly as $\omega \to \omega'$, so our condition is satisfied if $\Omega$ is any bounded subset of $E^k$.

REMARK 3. In both of the above examples we also have $|P_\omega(B) - P_{\omega'}(B)| \to 0$ uniformly as $\omega' \to \omega$. In Lemma 5 this was obtained from the compactness of $\Omega$ and used in showing $\lambda(F, G) \to 0$ as $H(F, G)$(or $H^*(F, G)) \to 0$. However, the

requirement that $\sup_{a,V} |L(\omega, a, V) - L(\omega', a, V)| \to 0$ uniformly still seems to need separate treatment.

REMARK 4. The problem of estimating $G$ is, in the main, still outstanding. Robbins [12] discusses the general problem of estimating a prior distribution function $G$ and, under some reasonable assumptions, demonstrates the convergence on the continuity set of $G$ of a certain type of estimator. He gives no explicit method for obtaining this type of estimator, but Deely and Kruse [1], with the additional assumption that $F_\omega(x)$ is continuous in $x$ for each $\omega$, exhibit a method of finding an estimator satisfying Robbins' condition. Finally, Fox [2], in the cases (a) $\Omega = (0, \infty)$ and $P_\omega$ uniform on $(0, \omega)$ and (b) $\Omega = (-\infty, \infty)$ and $P_\omega$ uniform on $(\omega, \omega + 1)$, exhibits estimates of $G^{(N)}$ which converge $[P_\theta]$ in the Lévy sense under conditions considerably weaker than the boundedness of $\Omega$ (he gives similar results for the estimation of a prior).

REMARK 5. Apart from the estimation problem, the procedures discussed here present the practical problems of choosing an appropriate partition, $\pi$, and of obtaining the values $P_\omega(V)$ for each $V \in \pi$ and $\omega \in \Omega$.

However there are many practical situations in which it is necessary either to round off the observations or to round off the values of $f_\omega(x)$ when $x$ is observed. This implies a partition which may be adequate. Let $\pi = \{V_1, \cdots, V_k\}$ be, as in Lemma 8, a sub-partition of each of the $\pi_\omega$ of Lemma 7. For each $\omega \in \Omega$ and $V \in \pi$, let

$$f_\omega^*(V) = \inf_{x \in V} f_\omega(x) \quad \text{if} \quad \sup_{x \in V} f_\omega(x) > 3 \inf_{x \in V} f_\omega(x)/(3 - \alpha)$$
$$= \text{arbitrary, with} \quad \inf_V f_\omega(x) \leq f_\omega^*(V) \leq \sup_V f_\omega(x) \quad \text{otherwise}.$$

Let $\Phi_\pi^*(F)$ be the collection of procedures which, when $x$ is observed, assign mass 1 to the set

$$\{a : \int L(\omega, a - b, x') f_\omega^*(x') \mu(x') \, dF(\omega) < \eta/k \quad \text{for all} \quad b \in A\},$$

assigning arbitrarily if $f_\omega(x') = 0$ for all $\omega$.

Suppose $\phi \in \Phi_\pi^*(F)$ and that $R_\pi(F, \psi_n) \leq R(F) + 1/n$. Then

$$R_\pi(F, \phi) = \sum_{V \in \pi} \int L(\omega, \phi(V), V) P_\omega(V) \, dF(\omega)$$
$$\leq \sum_{V \in \pi} \int L(\omega, \phi(V), V) f_\omega^*(V) \mu(V) \, dF(\omega)$$
$$+ M \sum_{V \in \pi} \int (P_\omega(V) - f_\omega^*(V) \mu(V))^+ \, dF(\omega).$$

The first term on the right is bounded by

$$\sum_{V \in \pi} \int L(\omega, \psi_n(V), V) f_\omega^*(V) \mu(V) \, dF(\omega) + \eta$$
$$\leq \sum_{V \in \pi} \int L(\omega, \psi_n(V), V) P_\omega(V) \, dF(\omega)$$
$$+ M \sum_{V \in \pi} \int (P_\omega(V) - f_\omega^*(V) \mu(V))^- \, dF(\omega).$$

Thus, letting $n \to \infty$,

$$R_\pi(F, \phi) \leq R_\pi(F) + \eta + M \sum_{V \in \pi} \int |P_\omega(V) - f_\omega^*(V) \mu(V)| \, dF(\omega).$$

The last term on the right is bounded by

$$M \sup_\omega \sum_{V \in \pi} |P_\omega(V) - f_\omega{}^*(V)\mu(V)| \leq M \sup_\omega \int |f_\omega(x) - f_\omega{}^*(x')| \, d\mu ,$$

which, by an argument parallel to that of Lemma 7, is bounded by $M\alpha$. Thus $R_\pi(F, \phi) \leq R_\pi(F) + \eta + M$ so that $\phi \in \Phi(F)$ for $\eta$ and $\alpha$ sufficiently small. Hence our results apply to procedures which are based on $f_\omega{}^*(x')$ instead of $P_\omega(x')$.

## REFERENCES

[1] DEELY, J. J. and KRUSE, R. L. (1968). Construction of sequences estimating the mixing distribution. *Ann. Math. Statist.* **39** 286-288.

[2] FOX, RICHARD J. (1970). Estimating the empirical distribution function of a parameter sequence. *Ann. Math. Statist.* **41** 1845-1852.

[3] GILLILAND, DENNIS C. (1968). Sequential compound estimation. *Ann. Math. Statist.* **39** 1890-1904.

[4] HANNAN, JAMES F. and HUANG, F. S. (1972 a). Equivariant procedures in the compound decision problem with finite state component problem. *Ann. Math. Statist.* **43** 102-112.

[5] HANNAN, JAMES F. and HUANG, F. S. (1972 b). A Stability of symmetrization of product measures with few distinct factors. *Ann. Math Statist.* **43** 308-319.

[6] HANNAN, JAMES F. and ROBBINS, HERBERT (1955). Asymptotic solutions of the compound decision problem for two completely specified distributions. *Ann. Math. Statist.* **26** 37-51.

[7] HANNAN, JAMES F. and VAN RYZIN, J. R. (1965). Rate of convergence in the compound decision problem for two completely specified distributions. *Ann. Math. Statist.* **36** 1743-1752.

[8] HORN, SUSAN CADAKIS (1968). The optimality criterion for compound decision problems. Technical Report No. 10, Department of Statistics, Stanford Univ.

[9] OATEN, ALLAN (1969). Approximation to Bayes risk in compound decision problems. RM–233, Department of Statistics and Probability, Michigan State Univ.

[10] RAO, R. RANGA (1962). Relations between weak and uniform convergence of measures with applications. *Ann. Math. Statist.* **33** 659-680.

[11] ROBBINS, HERBERT (1951). Asymptotically sub-minimax solutions of compound decision problems. *Proc. Second Berkeley Symp. Math. Statist. Prob.* 131-148. Univ. of California Press.

[12] ROBBINS, HERBERT (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35** 1-20.

[13] SAMUEL, ESTER (1963). Asymptotic solutions of the sequential compound decision problem. *Ann. Math. Statist.* **34** 1079-1094.

[14] VAN RYZIN, J. R. (1966). The compound decision problem with $m \times n$ finite loss matrix. *Ann. Math. Statist.* **37** 412-424.