A REMARK ON SEQUENTIAL DISCRIMINATION[1]

By David A. Freedman

University of California, Berkeley

1. Introduction. You are observing data gathered sequentially. The data is governed by one of a countable number of hypotheses. When all the data is in, these hypotheses are completely distinguishable. At some stage, depending only on data you have in hand, you want to stop and decide which hypothesis is correct, with the probability of error arbitrarily small. You can do this if and only if, for each hypothesis, there is a test based on a bounded amount of data which distinguishes that hypothesis from the set of all the others, with the probability of error arbitrarily small. The object of this note is to state and prove this theorem. (Since countable additivity does not simplify the problem, it will not be assumed, except as noted.)

Here is a special case of the theorem. A coin is to be tossed independently and repeatedly. The probability of heads is unknown, but is known to lie in a countable set Θ. You can stop, and decide what the parameter is with arbitrarily high probability, if and only if each point of Θ is isolated (that is, there is an interval around it free of other points of Θ).

2. Statement of theorem. Let Ω be a set. For $n = 1, 2, \cdots$ let $\mathcal{A}_n$ be a field of subsets of Ω, with $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \cdots$. Let $\mathcal{A} = \bigcup_{n=1}^{\infty} \mathcal{A}_n$, a field. Let Π be a countable set of finitely additive probabilities P on $\mathcal{A}$. A stopping time τ on Ω is a function taking the values $1, 2, \cdots, \infty$, with $\{\tau = n\} \varepsilon \mathcal{A}_n$ for $n = 1, 2, \cdots$. Let $\mathcal{A}_\tau$ be the field of all subsets A of Ω such that $A \cap \{\tau = n\} \varepsilon \mathcal{A}_n$ for $n = 1, 2, \cdots$. Unfortunately, $\mathcal{A}_\tau$ is not in general a subfield of $\mathcal{A}$. Define $P_*\{\tau < \infty\}$ as $\lim_{n\to\infty} P\{\tau \leqq n\}$ and $P_*(A)$ as $\lim_{n\to\infty} P(A \cap \{\tau \leqq n\})$ for $A \varepsilon \mathcal{A}_\tau$.

Consider the following two conditions:

(I) For any $\epsilon > 0$, there is a stopping time $\tau = \tau_\epsilon$, and there are disjoint sets $A_P \varepsilon \mathcal{A}_\tau$ for $P \varepsilon \Pi$, with $P_*\{\tau < \infty\} = 1$ and $P_*(A_P) > 1 - \epsilon$ for all $P \varepsilon \Pi$.

(II) For any $P \varepsilon \Pi$ and $\epsilon > 0$, there is a set $A = A_{P,\epsilon} \varepsilon \mathcal{A}$, such that $P(A) \geqq 1 - \epsilon$ and $Q(A) \leqq \epsilon$ for all $Q \varepsilon \Pi - \{P\}$.

Theorem 1. *Conditions* I *and II are equivalent.*

Condition I was introduced for an example by H. Robbins. I learned it from D. Blackwell. After seeing a draft of this paper, Robbins informed me that he had previously obtained Theorems 1 and 4.

The results of this paper can be viewed as extending some previous work of Hoeffding and Wolfowitz (1958).

3. Proof of Theorem 1. The first step is to prove (I) $\Rightarrow$ (II). Fix $P \varepsilon \Pi$ and $\epsilon > 0$. Find a stopping time τ and disjoint sets $A_Q \varepsilon \mathcal{A}_\tau$ with $Q_*\{\tau < \infty\} = 1$

Received 18 January 1967.

[1] This paper has been supported by the National Science Foundation, Grant GP-2593.

and $Q_*(A_Q) \geqq 1 - \frac{1}{2}\epsilon$ for all $Q \varepsilon \Pi$. Find n so large that $P[A_P \cap \{\tau \leqq n\}] \geqq 1 - \epsilon$. For $Q \neq P$,

$$Q[A_P \cap \{\tau \leqq n\}] \leqq 1 - Q_*(A_Q) \leqq \epsilon/2.$$

The second step is to prove (II) $\Rightarrow$ (I). Fix $\epsilon > 0$, and choose a sequence $\epsilon_i > 0$ with $\sum_{i=1}^{\infty} \epsilon_i \leqq \epsilon$. Let $P_1, P_2, \cdots$ be a sequence of elements of Π, in which each element of Π occurs infinitely often. Let $N_0 = 0$. For $k = 1, 2, \cdots$, find an integer $N_k > N_{k-1}$ and a set $A_k \varepsilon \mathcal{A}_{N_k}$, with $P_k(A_k) \geqq 1 - \epsilon_k$ and $Q(A_k) \leqq \epsilon_k$ for all $Q \varepsilon \Pi - \{P_k\}$. Let $\tau = \infty$ on $\Omega - \bigcup_{k=1}^{\infty} A_k$, and let $\tau = N_k$ on $(\Omega - A_1) \cap \cdots \cap (\Omega - A_{k-1}) \cap A_k$. Plainly, τ is a stopping time. Moreover $P_*\{\tau < \infty\} = 1$ for $P \varepsilon \Pi$: if $P = P_k$, then $P(\tau \leqq N_k) \geqq P(A_k) \geqq 1 - \epsilon_k \rightarrow 1$. For $P \varepsilon \Pi$, let $\lambda(P)$ be the least k with $P_k = P$. Let

$$A_P = \{\tau = N_{\lambda(P)}\} = A_{\lambda(P)} \cap (\Omega - A_{\lambda(P)-1}) \cap \cdots \cap (\Omega - A_1).$$

Plainly, as P varies over Π, the A_P are disjoint and $A_P \varepsilon \mathcal{A}_\tau$. (In addition, $A_P \varepsilon \mathcal{A}$). For $k < \lambda(P)$, $P_k \neq P$, so

$$\begin{aligned} P(A_P) &\geqq P(A_{\lambda(P)}) - P(A_{\lambda(P)-1}) - \cdots - P(A_1) \\ &\geqq 1 - \epsilon_{\lambda(P)} - \epsilon_{\lambda(P)-1} - \cdots - \epsilon_1 \\ &\geqq 1 - \textstyle\sum_{i=1}^{\infty} \epsilon_i \geqq 1 - \epsilon. \end{aligned}$$

4. Special cases. For Theorem 2, suppose that each $\mathcal{A}_n$ is finitely generated. Then any pairwise disjoint sequence of sets in $\mathcal{A}$ is empty from some time on, so there is no distinction between finite and countable additivity. Let $\sigma(\mathcal{A})$ be the smallest σ-field of subsets of Ω which includes $\mathcal{A}$. Any probability P on $\mathcal{A}$ extends uniquely to a countably additive probability P on $\sigma(\mathcal{A})$. Give the probabilities on $\mathcal{A}$ the *setwise convergence* topology, as follows: $P_n \rightarrow P$ iff $P_n(A) \rightarrow P(A)$ for each $A \varepsilon \mathcal{A}$. The set of probabilities is now compact and metrizable. Consider the following condition:

(III) For any $P \varepsilon \Pi$, there is a set $A = A_P \varepsilon \sigma(\mathcal{A})$, with $P(A) = 1$ and $Q(A) = 0$ for all Q in the closure of $\Pi - \{P\}$.

In the present situation,

THEOREM 2. *Condition* (II) *is equivalent to condition* (III).

PROOF. (II) $\Rightarrow$ (III). Fix $P \varepsilon \Pi$ and $\delta_n > 0$ with $\delta_n \rightarrow 0$. Let $\delta_{nk} > 0$ and $\sum_k \delta_{nk} \leqq \delta_n$. There is a stopping time $\tau = \tau_{nk}$ and an $A_{nk} \varepsilon \mathcal{A}_\tau$ with $P(A_{nk}) \geqq 1 - \delta_{nk}$ and $Q(A_{nk}) \leqq \delta_{nk}$ for all $Q \varepsilon \Pi - \{P\}$. There is a positive integer $N = N_{nk}$ and a set $B_{nk} \varepsilon \mathcal{A}_N$ with $B_{nk} \subset A_{nk}$ and $P(B_{nk}) \geqq 1 - 2\delta_{nk}$. But $Q(B_{nk}) \leqq \delta_{nk}$ for all $Q \varepsilon \Pi - \{P\}$, so for all $Q \varepsilon K$, the closure of $\Pi - \{P\}$. Let $B_n = \bigcup_k B_{nk}$. Then $P(B_n) = 1$ and $Q(B_n) \leqq \delta_n$ for all $Q \varepsilon K$. Let $B = \bigcap_n B_n$. Then $P(B) = 1$ and $Q(B) = 0$ for all $Q \varepsilon K$.

(III) $\Rightarrow$ (II). Fix $P \varepsilon \Pi$ and $\epsilon > 0$. Find $A \varepsilon \sigma(\mathcal{A})$ with $P(A) = 1$ and $Q(A) = 0$ for all Q in the closure K of $\Pi - \{P\}$. There is a decreasing sequence $A_n \varepsilon \mathcal{A}$ with $\bigcap_{n=1}^{\infty} A_n \subset A$ and $P(A_n) \geqq 1 - \epsilon$ for all n. As n increases, the continuous function $Q \rightarrow Q(A_n)$ decreases pointwise to 0 on the compact set K, so there is an n with $Q(A_n) \leqq \epsilon$ for all $Q \varepsilon K$, completing the proof.

If each $\mathfrak{A}_n$ is the σ-field generated by a countable partition, and each $P \varepsilon \Pi$ is countably additive, the situation is similar. Condition (II) should be modified by topologizing the countably additive subprobabilities on $\mathfrak{A}$ so $P_j \rightarrow P$ iff $P_j(A) \rightarrow P(A)$ for each atom A of each σ-field $\mathfrak{A}_n$. The countably additive subprobabilities are compact, so the argument remains about the same.

Let Ω be the set of sequences of 0 and 1, and $\mathfrak{A}_n$ the field of subsets of Ω depending only on the first n coordinates. For each $\theta \varepsilon [0, 1]$, let P_θ be the probability on $\mathfrak{A}$ for which the coordinates are independent, 1 with probability θ, 0 with probability $1 - \theta$. Let Θ be a countable subset of [0, 1], with the usual topology.

THEOREM 3. *Condition* I *holds for* $\{P_\theta : \theta \varepsilon \Theta\}$ iff *each point of* Θ *is isolated.*

PROOF. The map $\theta \rightarrow P_\theta$ is a homeomorphism. If $\theta_n \rightarrow \theta$ in Θ, then $P_{\theta_n} \rightarrow P_\theta$, and (III) fails. The set of sequences of 0 and 1 where the relative frequency of 1 is θ has P_θ-measure 1, and P_α-measure 0 for all $\alpha \neq \theta$. If each point of Θ is isolated, (III) holds, completing the proof.

This reasoning can be extended as follows: Let $\mathfrak{X}$ be a set and Σ a σ-field of subsets of $\mathfrak{X}$. Topologize the countably additive probabilities P on Σ as follows: a generalized sequence P_α converges to the limit P iff $P_\alpha(A) \rightarrow P(A)$ for all $A \varepsilon \Sigma$. Let Ω be the set of all $\mathfrak{X}$-sequences, with the product σ-field, and let $\mathfrak{A}_n$ be the field generated by the measurable rectangles depending only on the first n coordinates. For a countably additive probability θ on Σ, let P_θ be the power probability on the product σ-field of Ω. Let Θ be a countable set of countably additive probabilities on Σ.

THEOREM 4. *Condition* (I) *holds for* $\{P_\theta : \theta \varepsilon \Theta\}$ iff *each point of* Θ *is isolated.*

PROOF. Suppose $\theta \varepsilon \Theta$ is isolated from all other $\alpha \varepsilon \theta$. There is a positive integer k, a positive real number δ, and Σ-sets $A_1, \cdots, A_k$, such that for all $\alpha \varepsilon \Theta - \{\theta\}$, for some $j = j(\alpha)$ with $1 \leqq j \leqq k$, $|\theta(A_j) - \alpha(A_j)| \geqq \delta$. Let B_N be the set of $\mathfrak{X}$-sequences where, among the first N coordinates, the relative frequency of visits to A_j is between $\theta(A_j) - \frac{1}{2}\delta$ and $\theta(A_j) + \frac{1}{2}\delta$ for all $j = 1, \cdots, k$. Fix $\epsilon > 0$. Use Chebychev's inequality to verify that for large N, $P_\theta(B_N) \geqq 1 - \epsilon$ and $P_\alpha(B_N) < \epsilon$ for all $\alpha \varepsilon \Theta - \{\theta\}$.

5. A variation. Let $\mathfrak{X}$ be a compact metric, and Ω the set of $\mathfrak{X}$-sequences, with the product structure. Let $\mathfrak{A}_n$ be the σ-field of Borel subsets of Ω depending only on the first n coordinates. Let Π be a countable set of countably additive probabilities on Ω. Consider the following two conditions:

(I*) For any $\epsilon > 0$, there is a stopping time $\tau = \tau_\epsilon$, with $\{\tau = n\}$ open, and there are disjoint open sets $A_P \varepsilon \mathfrak{A}_\tau$ for $P \varepsilon \Pi$, with $P\{\tau < \infty\} = 1$ and $P(A_P) \geqq 1 - \epsilon$ for all $P \varepsilon \Pi$.

(III*) For any $P \varepsilon \Pi$, there is a Borel set $A = A_P$, with $P(A) = 1$ and $Q(A) = 0$ for all Q in the weak* closure of $\Pi - \{P\}$.

THEOREM 5. *Conditions* (I*) *and* (III*) *are equivalent.*

PROOF. Argue as in Theorems 1 and 2.

6. Examples. Let Ω be the set of sequences of 0 and 1, and $\mathcal{A}_n$ the field of sets determined by the first n coordinates. Let Π be a countable family of probabilities on $\mathcal{A}$, satisfying condition (I). A natural method for constructing τ and A_P is the following: Let μ be a countably additive (prior) probability on Π, with $\mu\{P\} > 0$ for all $P \varepsilon \Pi$. Let $\frac{1}{2} < \alpha(P) < 1$ for all $P \varepsilon \Pi$. For $\omega \varepsilon \Omega$, let ω_n be the set of all $\omega' \varepsilon \Omega$, which agree with ω on the first n coordinates. Compute the (posterior) probability $\mu_{n,\omega}$ on Π as

$$\mu_{n,\omega}\{P\} = \mu\{P\}P(\omega_n)/\textstyle\sum_{Q\varepsilon\Pi} \mu\ \{Q\}Q(\omega_n).$$

Let τ be the least n if any such that, for some $P \varepsilon \Pi$, $\mu_{n,\omega}\{P\} > \alpha(P)$, and let A_P be the set of ω with $\mu_{\tau(\omega),\omega}\{P\} > \alpha[P]$. There may always be a μ for which this procedure works, but

EXAMPLE I. There is a $\Pi = \{P_0, P_1, \cdots\}$ satisfying (I) and a μ for which this procedure fails, namely:

(1) for each j, with P_j-probability 1, $\mu_{n,\cdot}\{P_j\} < 1$ for all n;

(2) for any α, β in $(0, 1)$, for all $N \geqq K(\alpha, \beta)$, for P_N-almost all ω: $\mu_{n,\omega}\{P_0\} > \alpha$ before $\mu_{n,\omega}\{P_N\} > \beta$.

CONSTRUCTION. P_0 puts mass $q_n > 0$ on the sequence whose first n coordinates are 0, and whose remaining coordinates are all 1. Here $\sum_1^\infty q_n = 1$. For $n \geqq 1$, with respect to P_n, the first n coordinates are 0 with probability 1, and the remaining coordinates are independent and identically distributed, being 0 with probability θ_n and 1 with probability $1 - \theta_n$. Here $0 < \theta_n < 1$, and each point of $\{\theta_n\}$ is isolated. The prior μ puts mass $\mu_n > 0$ on P_n, with $\sum_0^\infty \mu_n = 1$. Let $\hat{q}_n = q_n + q_{n+1} + \cdots$, $\hat{\mu}_n = \mu_n + \mu_{n+1} + \cdots$, and $\hat{\rho}_n = \mu_1\theta_1^{n-1} + \mu_2\theta_2^{n-2} + \cdots + \mu_{n-1}\theta_n$. Choose q_n, θ_n, μ_n so

$$\hat{q}_n/\hat{\mu}_n \to \infty \tag{3}$$

and

$$\hat{q}_n/\hat{\rho}_n \to \infty. \tag{4}$$

For instance, let $\theta_n \leqq \frac{1}{2}$, $\mu_n = 2^{-n-1}$, and $q_n = 1/[n(n+1)]$ for all n. Then $\hat{q}_n = 1/n$, $\hat{\mu}_n = 2^{-n}$, $\hat{\rho}_n \leqq (n-1)2^{-n-1}$.

VERIFICATION. Plainly, (III) and (1) hold. To verify (2), suppose $1 \leqq m \leqq n \leqq N$. For P_N-almost all ω, the first N coordinates of ω are 0. For such an ω,

$$\mu_{n,\omega}\{P_0\} = \mu_0\hat{q}_n/(\mu_0\hat{q}_n + \hat{\mu}_n + \hat{\rho}_n) \tag{5}$$

and

$$\mu_{m,\omega}\{P_N\} = \mu_N/(\mu_0\hat{q}_m + \hat{\mu}_m + \hat{\rho}_m). \tag{6}$$

Use (3) and (4) to find n so large that the right side of (5) is more than α. Now choose $K \geqq n$ so large that for all $m = 1, \cdots, n$ and all $N \geqq K$, the right side of (6) is no more than β.

Another plausible method is the following: For each $P \varepsilon \Pi$, compute the likelihood ratios $P(\omega_n)/Q(\omega_n)$. Fix numbers $1 < K(P) < \infty$. Stop when for the first time, for some P, $P(\omega_n)/Q(\omega_n) > K(P)$ for all $Q \varepsilon \Pi - \{P\}$, and decide on that P. Blackwell showed me an example where this method fails too.

Example 2. There is a Π satisfying condition (I), such that for some $Q \varepsilon \Pi$, for any $\omega \varepsilon \Omega$ and $n = 1, 2, \cdots$, there is a $P \varepsilon \Pi$ with $Q(\omega_n)/P(\omega_n) \leqq 1$.

Notation. Ω is the set of sequences of 0, 1, 2, and $\mathfrak{A}_n$ is the field of sets determined by the first n coordinates. Q is the probability on $\mathfrak{A}$ for which the coordinates are independent and identically distributed, being 0 with probability $\frac{1}{2}$ and 1 with probability $\frac{1}{2}$. Let Σ be the set of finite non-empty sequences of 0 and 1; let $|\sigma|$ be the length of $\sigma \varepsilon \Sigma$.

Construction. Let P_σ, a probability on Ω, assign mass $2^{-|\sigma|}$ to the sequence consisting of σ followed by all 2's, and let P_σ assign the remaining mass $1 - 2^{-|\sigma|}$ to the sequence which consists of $|\sigma|$ 2's followed by σ followed by all 2's. Then Π is $\{Q, P_\sigma : \sigma \varepsilon \Sigma\}$. The verification is easy and is omitted.

At one time, I thought that (III) might be equivalent to saying that each P in Π was orthogonal to each Q in the closure of $\Pi - \{P\}$. This is false, as Blackwell showed by

Example 3. There is a set Π which does not satisfy (III), although each $P \varepsilon \Pi$ is orthogonal to each $Q \varepsilon$ closure $\Pi - \{P\}$.

Construction. Use the notation of Example 2. Let Q_σ assign mass 1 to the sequence σ followed by all 2's. Then $\Sigma = \{Q, Q_\sigma : \sigma \varepsilon \Sigma\}$. The verification is easy.

Finally, in view of Theorems 2 and 4, it is reasonable to ask whether, in the countably additive case, (I) is equivalent to this condition: each $P \varepsilon \Pi$ is distinguishable from all the countably additive probabilities in the closure of $\Pi - \{P\}$, with respect to the topology of setwise convergence on $\mathfrak{A}$. This too is false, as indicated by the discussion following Theorem 2.

Example 4. There is a set Π of countably additive probabilities which does not satisfy (I), although for each $P \varepsilon \Pi$ there is an $A \varepsilon \sigma(\mathfrak{A})$ with $P(A) = 1$ and $Q(A) = 0$, for all countably additive Q in the closure of $\Pi - \{P\}$ relative to the topology of setwise convergence on $\mathfrak{A}$.

Construction. Ω is the set of sequences of positive integers. $\mathfrak{A}_n$ is the σ-field of subsets depending only on the first n coordinates. For $n = 1, 2, \cdots$ with respect to the probability P_n on $\mathfrak{A}$, the coordinates are independent the first coordinate is 1 with probability $\frac{1}{2}$ and n with probability $\frac{1}{2}$; each other coordinate is 1 with probability $\frac{1}{2} + 1/(n + 2)$ and 2 with probability $\frac{1}{2} - 1/(n + 2)$. With respect to the probability P_0 on $\mathfrak{A}$, the coordinates are independent; the first coordinate is 1, and each other coordinate is 1 with probability $\frac{1}{2}$ and 2 with probability $\frac{1}{2}$. Finally, $\Pi = \{P_0, P_1, \cdots\}$.

Verification. Any subset of Π is closed under setwise convergence, and any pair of elements of Π are orthogonal. But P_0 cannot be separated in the sense of condition (II) from $\{P_1, P_2, \cdots\}$.

REFERENCE

Hoeffding, W. and Wolfowitz, J. (1958). Distinguishability of sets of distributions. (The case of independent and identically distributed chance variables). *Ann. Math. Statist.* **29** 700–718