ON THE OPTIMALITY OF SEQUENTIAL PROBABILITY RATIO TESTS[1]

By Theodore K. Matthes[2]

The RAND Corporation

1. Introduction. The sequential probability ratio test (SPRT) for testing a simple hypothesis H_0 against a single alternative H_1 was first proved to be optimal by Wald and Wolfowitz [5] in a sense there defined. A much simpler proof has been given by L. LeCam and appears in Lehmann's book [3]. In the present note a proof of this optimality is given which relies primarily on a simple mapping theorem.

2. Preliminaries. As is customary, we introduce the auxiliary Bayes problem of testing sequentially H_0 against H_1 when ξ is the a priori probability of H_0, w_i is the loss caused by wrong decision under H_i and there is a unit cost per observation. Let $\alpha_i(\delta)$ be the probability of wrong decision and $E_i(\delta)$ the average sample size under H_i using procedure δ. Denote the risk by $r(\xi, w_0, w_1; \delta)$. To show that a given SPRT δ_0 is optimal it suffices to show that for each ξ, $0 < \xi < 1$, losses w_0, w_1 may be chosen (depending on ξ) in such a way that δ_0 is Bayes relative to ξ, w_0, w_1. For if then δ is any procedure with $\alpha_i(\delta) \leqq \alpha_i(\delta_0)$, $i = 0, 1$, it would follow in the usual way that $\xi E_0(\delta) + (1 - \xi)E_1(\delta) \geqq \xi E_0(\delta_0) + (1 - \xi)E_1(\delta_0)$ and letting ξ tend in turn to zero and one would yield $E_i(\delta) \geqq E_i(\delta_0)$, $i = 0, 1$ provided $E_i(\delta) < \infty$.

The Bayes procedures may be characterized by means of the functions

$$\rho_0(\xi, w_0, w_1) = \min [\xi w_0, (1 - \xi)w_1]$$

$$\rho^*(\xi, w_0, w_1) = \inf_{\delta \varepsilon \mathfrak{D}^*} r(\xi, w_0, w_1; \delta)$$

where $\mathfrak{D}^*$ is the class of all procedures which take at least one observation. In fact, for each w_0, w_1 let $g(w_0, w_1)$ and $h(w_0, w_1)$ denote the left and right intersections, respectively, of ρ_0 and ρ^* in $(0, 1)$ provided ρ_0 and ρ^* intersect at all; otherwise, set $g = h = w_1/(w_0 + w_1)$. It is shown in [3] that a Bayes procedure continues sampling beyond n observations only so long as the posteriori probability ξ_n of H_0 satisfies $g(w_0, w_1) \leqq \xi_n \leqq h(w_0, w_1)$ for $n \geqq 0 (\xi_0 \equiv \xi)$.

A SPRT δ_0 characterized by boundaries B, $A (B < A)$ may be described equivalently as taking a first observation and then sampling further so long as

$$g_0 \equiv 1/[1 + A(1 - \xi)/\xi] < \xi_n < 1/[1 + B(1 - \xi)/\xi] \equiv h_0 .$$

Received March 9, 1962.

[1] Any views expressed in this paper are those of the author. They should not be interpreted as reflecting the views of The RAND Corporation or the official opinion or policy of any of its governmental or private research sponsors. Papers are reproduced by The RAND Corporation as a courtesy to members of its staff.

[2] Now at the University of Oregon.

Thus, for δ_0 to be Bayes relative to ξ, w_0, w_1 it must be shown that for $g_0 < h_0$ there is a solution to

$$(2.1) \qquad g(w_0, w_1) = g_0, \qquad h(w_0, w_1) = h_0$$

and that the Bayes procedure associated with ξ and stopping boundaries (2.1) takes at least one observation. The latter requirement is that $g_0 < \xi < h_0$, or $B < 1 < A$. However, it is easy to see that this restriction may be dropped provided in the optimality definition one compares SPRTs with only those procedures which take at least one observation.

3. Proof of optimality. It is shown first that g and h are continuous when $w_0, w_1 > 0$. Let $\epsilon < \xi < 1 - \epsilon$. Then $\rho^*(\xi, w_0, w_1) \leqq 1 + \min(w_0, w_1) < B$ for (w_0, w_1) suitably bounded. It follows that $\rho^*(\xi, w_0, w_1)$ is the infimum of risks over only those procedures that take at least one observation and whose risks are at most B. For such a procedure $E_0(\delta) + E_1(\delta) \leqq B/\epsilon$. The risk functions of all these procedures are equicontinuous when ξ, w_0, w_1 are restricted as above and this fact implies that the infimum $\rho^*(\xi, w_0, w_1)$ is continuous in the open region $0 < \xi < 1$; $w_0, w_1 > 0$. Set $\gamma(\xi, w_0, w_1) = \rho_0(\xi, w_0, w_1) - \rho^*(\xi, w_0, w_1)$. It follows that the function

$$\Psi(\xi, w_0, w_1) = \gamma(\xi, w_0, w_1) \left\{ \left| \xi - \frac{w_1}{w_0 + w_1} \right| + \max\left[0, \gamma\left(\frac{w_1}{w_0 + w_1}, w_0, w_1\right)\right]\right\}$$

is continuous in the same region. By considering the cases $\gamma(w_1/(w_0 + w_1), w_0, w_1) > 0, =0, <0$, it is seen that $g(w_0, w_1)$ is the unique solution in ξ to the equation

$$(3.1) \qquad \Psi(\xi, w_0, w_1) = 0, \qquad \xi \leqq w_1/(w_0 + w_1).$$

Suppose a sequence (w_{0n}, w_{1n}) converges to a point (w_0, w_1) with $w_0, w_1 > 0$ while $g(w_{0n}, w_{1n}) \to g'$. Since always $\min[1/w_0, w_1/(w_0 + w_1)] \leqq g(w_0, w_1) \leqq w_1/(w_0 + w_1)$, it must be that $0 < g' \leqq w_1/(w_0 + w_1)$. From the above,

$$\Psi(g(w_{0n}, w_{1n}), w_{0n}, w_{1n}) = 0 \quad \text{and by continuity,} \quad \Psi(g', w_0, w_1) = 0.$$

By virtue of the uniqueness of the solution to (3.1), it follows that

$$\lim g(w_{0n}, w_{1n}) = g' = g(w_0, w_1).$$

This shows that g (likewise h) is continuous.

In the $w_0 - w_1$ plane consider the square path consisting of the line segments $C_1 : w_0 = 1, 1 \leqq w_1 \leqq K$ (K will be chosen shortly); $C_2 : 1 \leqq w_0 \leqq K, w_1 = 1$; $C_3 : w_0 = K, 1 \leqq w_1 \leqq K$; and, $C_4 : 1 \leqq w_0 \leqq K, w_1 = K$. Define the complex-valued function $\Phi(w_0, w_1) = g(w_0, w_1) + ih(w_0, w_1)$. When either $w_0 = 1$ or $w_1 = 1$ a Bayes procedure (not unique in this case) is to take no observations and guess that hypothesis which is associated with the smaller loss. Then

$g = h = w_1/(w_0 + w_1)$ since otherwise there would be a range of ξ for which it would pay to sample. Hence, as w_0, w_1 traces out $C_1 \cup C_2$, Φ moves along the line segment joining $(1 + i)K/(K + 1)$ and $(1 + i)/(K + 1)$.

To see the behavior of Φ on C_3 and C_4 we shall compare the risk of a Bayes procedure δ_B relative to ξ, w_0, w_1 with the risk of a fixed sample procedure δ_F which takes N observations and has error probabilities α_i. Since

$$\rho_0(g, w_0, w_1) \leqq \rho^*(g, w_0, w_1) \quad \text{and}$$
$$r(\xi, w_0, w_1; \delta_B) = \min [\rho_0(\xi, w_0, w_1), \rho^*(\xi, w_0, w_1)],$$

one has

$$\begin{aligned} w_0 g &= \rho_0(g, w_0, w_1) \\ &= r(g, w_0, w_1; \delta_B) \\ &\leqq r(g, w_0, w_1; \delta_F) \\ &= N + g\alpha_0 w_0 + (1 - g)\alpha_1 w_1 . \end{aligned} \tag{3.2}$$

On C_3, $w_0 = K$, $w_1 \leqq K$ and (3.2) gives $g(K, w_1) \leqq N/K + \alpha_0 + \alpha_1$. Quite analogously, $h(w_0, K) \geqq 1 - (N/K + \alpha_0 + \alpha_1)$ on C_4. Choose, now, first α_0, α_1 small enough and then K large enough so that $N/K + \alpha_0 + \alpha_1 < \min(g_0, 1 - h_0)$. With this choice of K, $\mathcal{R}e[\Phi] = g(K, w_1) < g_0$ on C_3 and $\mathcal{I}m[\Phi] = h(w_0, K) > h_0$ on C_4. Recalling that $g_0 < h_0$, it is clear (a diagram may help) that Φ encircles the point $g_0 + ih_0$ as w_0, w_1 traces out once its whole path.

LEMMA (Rado and Reichelderfer [4], p. 390): *Let R be a bounded simply connected Jordan region in the plane and let C be its arbitrarily oriented boundary curve. If Φ is a complex-valued function continuous in R and $\Phi \neq 0$ in R, then* Var_C [*argument* Φ] $= 0$.

Application of this Lemma to the continuous function $\Phi - (g_0 + ih_0)$ shows that (2.1) has a solution.

REMARKS. It should be pointed out that the novelty of the above proof of optimality consists in applying the mapping lemma to show directly that the pair of equations in (2.1) has a solution. Wald and Wolfowitz [5] originally proved the solvability of (2.1) (they actually treat a closely related pair of equations) by choosing w_1 depending on w_0 so that one of the equations is satisfied and then varying w_0 to satisfy the other equation as well. Arrow, Blackwell and Girshick [1] obtain a pair of equations for w_0 and w_1 that are linear in these variables with coefficients depending on g and h in a complicated way. The possibility of solving for w_1, w_2 in terms of g, h is not fully settled, however, inasmuch as the simultaneous equations might be incompatible for some g_0, h_0. Indeed, compatibility is assured only for g_0, h_0 in the range of the mapping (g, h). Proceeding similarly, LeCam in [3] effectively proves that these equations are solvable by varying losses, again one at a time. Burkholder and Wijsman [2] make use of the fact that to prove optimality it suffices to show only that there exist certain g_0, h_0 sufficiently close to 0 (and 1) for which (2.1) is solvable.

REFERENCES

[1] ARROW, K. J., BLACKWELL, D. and GIRSHICK, M. A. (1949). Bayes and minimax solutions of sequential decision problems. *Econometrica* **17** 213–244.

[2] BURKHOLDER, D. L. and WIJSMAN, R. A. (1963). Optimum properties and admissibility of sequential tests. *Ann. Math. Statist.* **34** 1–17.

[3] LEHMANN, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

[4] RADO, T. and REICHELDERFER, P. V. (1955). *Continuous Transformations in Analysis.* Springer-Verlag, Berlin.

[5] WALD, A. and WOLFOWITZ, J. (1948). Optimum character of the sequential probability ratio test. *Ann. Math. Statist.* **19** 326–339.