AN ESSENTIALLY COMPLETE CLASS OF DECISION FUNCTIONS FOR CERTAIN STANDARD SEQUENTIAL PROBLEMS[1]

By Milton Sobel

Columbia University[2]

Summary. A sequential problem is considered in which independent observations are taken on a chance variable X whose distribution can be represented by

$$dG_\theta(x) = \psi(\theta)e^{\theta x}\,d\mu(x), \tag{1}$$

where the parameter θ belongs to a given interval Ω of the real line but is otherwise unknown. The problem is to test $H_1:\theta \leqq \theta^*$ against $H_2:\theta > \theta^*$, where θ^* is a given point in Ω. Under certain assumptions the following class A is shown to be essentially complete relative to the class of decision rules with bounded risk functions. The decision rule $\delta \varepsilon A$ if and only if after taking n observations

(i) δ depends on the observations only through n and $v_n = \sum_{i=1}^n x_i$ and

(ii) δ specifies a closed interval $J_n:[a_{1n}, a_{2n}]$ for each n and the following rule of action

(a) Stop experimentation as soon as $v_n \not\varepsilon J_n$ and

  (1) accept H_1 if $v_n < a_{1n}$

  (2) accept H_2 if $v_n > a_{2n}$.

(b) If $a_{1n} < a_{2n}$ take another observation if $a_{1n} < v < a_{2n}$.

(c) If $a_{1n} < a_{2n}$ and $v = a_{in}$, accept H_i or take another observation or randomize between these two ($i = 1, 2$).

The Koopman-Darmois family of probability laws given above contains discrete members such as the binomial and Poisson distributions as well as absolutely continuous members such as the normal and exponential. It is interesting to note that the members of the class A can be obtained by starting with the sequential probability ratio test for testing some point $\theta_1^* \leqq \theta^*$ against another point $\theta_2^* > \theta^*$, namely, continue as long as

$$B < \frac{\prod_{i=1}^n \psi(\theta_2^*)e^{\theta_2^* x_i}}{\prod_{i=1}^n \psi(\theta_1^*)e^{\theta_1^* x_i}} < A$$

and replacing the constants B, A by two arbitrary sequences B_n, A_n such that $B_n \leqq A_n$ ($n = 1, 2, \cdots$).

1. Statement of the problem. Independent observations are taken on a chance variable whose distribution is given by

$$dG_\theta(x) = \psi(\theta)e^{x\theta}\,d\mu(x) \tag{2}$$

Received 6/28/51, revised 3/24/53.

[1] Work sponsored by the Office of Naval Research.

[2] Now at Cornell University.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.* www.jstor.org

where the parameter θ is an unknown constant in a given subset Ω of the real line but is otherwise unknown and hence

$$\psi(\theta) = \left[\int_{-\infty}^{\infty} e^{x\theta}\, d\mu(x)\right]^{-1} > 0 \text{ for all } \theta \in \Omega. \tag{3}$$

The measure $\mu(x)$ is given to be absolutely continuous or discrete. The parameter space Ω is given to be an interval $[\underline{\theta}, \bar{\theta}]$ which may be finite or infinite, open or half-open or closed. It will be shown that a certain class A of decision rules is an essentially complete class for testing the hypothesis

$$H_1: \underline{\theta} \leqq \theta^* \text{ against } H_2: \theta > \theta^*$$

where θ^* is a given constant in Ω. There is also given an indifference zone Z in the form of an open interval (θ_1, θ_2) with

$$\underline{\theta} < \theta_1 \leqq \theta^* < \theta_2 < \bar{\theta}, \tag{4}$$

that is, if the true $\theta \in Z$ then the loss incurred by accepting either hypothesis is negligible and can be set equal to zero.

The results of this paper hold also for the case in which (2) is replaced by $\varphi(\theta) \exp \{r(\theta)t(y)\}\, d\nu(y)$ where $r(\theta)$ is strictly monotonic in θ and $t(Y)$ is an absolutely continuous or discrete chance variable. Letting $x = t(y)$ reduces this form to $\varphi(\theta) \exp \{xr(\theta)\}\, d\mu(x)$ where $\mu(x)$ is absolutely continuous or discrete. All the proofs below depend only on the strict monotonicity of θ and therefore hold if θ is replaced by $r(\theta)$ throughout.

If Ω is not an interval we can define a strictly monotonic function $\theta(\tau)$ from an interval Ω^* onto Ω. Then considering τ as the unknown parameter in Ω^* and using the last remark in the previous paragraph the results will still hold.

ASSUMPTIONS.

1. Let $W(\theta, j)$ $(j = 1, 2)$ denote the loss incurred by accepting H_j when θ is the true value. It is assumed that

$$\begin{aligned} W(\theta, 1) &= 0 \text{ for } \theta < \theta_2\,; \qquad & W(\theta, 2) &= 0 \text{ for } \theta > \theta_1 \\ W(\theta, 1) &> 0 \text{ for } \theta > \theta_2\,; \qquad & W(\theta, 2) &> 0 \text{ for } \theta < \theta_1 \end{aligned} \tag{5}$$

and that the two weight or loss functions $W(\theta, j)$ $(j = 1, 2)$ are bounded functions of θ on Ω. It is also assumed that $W(\theta, 1)$ is a nondecreasing and $W(\theta, 2)$ a nonincreasing function of θ on Ω.

2. Let $C(n)$ denote the cost of taking n observations on x. It is assumed that

$$C(n) = c_1 + c_2 + \cdots + c_n \qquad [C(0) = 0] \tag{6}$$

where c_n, the cost of taking the nth observation, is a positive constant which may vary with n. It is also assumed that for some positive constant K

$$\liminf_{n=\infty} nc_n \geqq K. \tag{7}$$

Let $V = \sum_{i=1}^{n} X_i$, $\bar{X} = V/n$ and let x_i denote the observed value of X_i. Let $S_{\bar{x}}$ denote the smallest interval containing all possible values of $\bar{X}$ for all

n $(n = 1, 2, \cdots)$. It is convenient to use $\bar{x}$ to denote an arbitrary point in $S_{\bar{x}}$. We define

$$g_{\bar{x}}(\theta) = \psi(\theta)e^{\bar{x}\theta} \qquad \text{for all } \bar{x} \varepsilon S_{\bar{x}}. \tag{8}$$

It is easily seen that a maximum likelihood estimate $\hat{\theta}$ of θ based on the observed value $\bar{x}$ of $\bar{X}$ is obtained by maximizing (8). The following assumptions are used in Lemma 1 and Theorem 2.

3. (i) For each $\bar{x} \varepsilon S_{\bar{x}}$ there exists a point $\hat{\theta} = \hat{\theta}(\bar{x}) \varepsilon \Omega$ such that $g_{\bar{x}}(\theta)$ is strictly increasing in $[\underline{\theta}, \hat{\theta}]$ and strictly decreasing in $[\hat{\theta}, \bar{\theta}]$.

(ii) There exist points $\bar{x}_1 < \bar{x}_2$ in $S_{\bar{x}}$ such that

$$\hat{\theta}(\bar{x}_1) \neq \hat{\theta}(\bar{x}_2) \text{ and } \theta_1 < \hat{\theta}(\bar{x}_i) < \theta_2 \qquad (i = 1, 2).$$

If $\psi(\theta)$ is differentiable, then for assumptions 3(i) and (ii) to hold it is sufficient that for each $\bar{x} \varepsilon S_{\bar{x}}$ the maximum likelihood equation has a unique solution $\hat{\theta}(\bar{x})$ which takes on all values in Ω as $\bar{x}$ runs through $S_{\bar{x}}$. In the normal, binomial and Poisson cases the reader can easily check that $\hat{\theta}(\bar{x}) = \bar{x}$ and that the assumptions are satisfied.

3. Regular convergence in the space of decision functions. It is assumed that the reader is familiar with the concepts of cost, loss, risk function and Bayes solution. However, the definition of regular convergence in the space of decision functions given by Wald ([1], p. 65) is too general for our purposes and is reviewed here.

Let $x^* = (x_1, x_2, \cdots)$ denote a sequence of independent observations on a random variable X. A (randomized) decision function δ following Wald ([1], p. 6) consists of a set of nonnegative functions $\delta_{jn}(x^*) \geqq 0$, $(j = 0, 1, 2; n = 0, 1, 2, \cdots)$ defined for all x^* and such that for all x^*

$$\sum_{j=0}^{2} \delta_{jn}(x^*) = 1 \qquad (n = 0, 1, 2, \cdots).$$

The quantities $\delta_{jn}(x^*)$ $(j = 0, 1, 2)$ [which depend only on the first n coordinates of x^*] represent, respectively, the probability of taking another observation, accepting H_1 and accepting H_2 when n observations have been taken, the observed values are the first n coordinates of x^* and δ is the decision rule used. A nonrandomized decision function is a special case of the above in which the value of $\delta_{jn}(x^*)$ is zero or one for each n, each j and each x^*.

Two cases are considered according as X is a discrete or absolutely continuous chance variable. In the discrete case a sequence $\{\delta^i\}$ is said to converge to a limit δ^0 in the regular sense if

$$\lim_{i=\infty} \delta^i_{jn}(x^*) = \delta^0_{jn}(x^*) \tag{9}$$

for each integer $n \geqq 0$, each x^* and each j $(j = 0, 1, 2,)$. In the absolutely continuous case let

$$\delta_{0jn}(x^*) = \left[\prod_{i=0}^{n-1} \delta_{0i}(x^*)\right] \delta_{jn}(x^*) \tag{10}$$

denote the probability, given x^* and δ, of selecting d_j immediately after the nth observation. For each j and n it is assumed that this is a Borel measurable function of $x_1, \cdots, x_n$. For each j and n let

$$\delta_{0jn}(S) = \int_S \delta_{0jn}(x^*)\, dx_1\, dx_2 \cdots dx_n \tag{11}$$

where $S = S_n$ is a Lebesgue measurable set in the space of $x_1, x_2, \cdots x_n$. For $n = 0$ the left member of (11) as well as the integrand are defined to be $\delta_{0j0}(x^*) = \delta_{j0}$ $(j = 0, 1, 2)$. The sequence $\{\delta^i\}$ $(i = 1, 2, \cdots)$ is said to converge to δ^0 in the regular sense if

$$\lim_{i=\infty} \delta^i_{0jn}(S) = \delta^0_{0jn}(S) \tag{12}$$

for all j, all bounded sets S and all integers $n \geqq 0$. It has been shown ([1], Theorem 3.1) that the sequence $\{\delta^i\}$ $(i = 1, 2, \cdots)$ is convergent, that is, a limiting rule δ^0 satisfying (12) exists, if the left member of (12) exists for all j, all bounded, measurable sets S and all integers $n \geqq 0$.

The definition of regular convergence for the absolutely continuous case is weaker than that of the pointwise type which was given for the discrete case. It will be useful to note that the corresponding definition of pointwise convergence for the absolutely continuous case implies regular convergence. This implication is easily seen from the fact that

$$\lim_{i=\infty} \int_S \delta^i_{0jn}(x^*)\, dx_1 \cdots dx_n = \int_S \lim_{i=\infty} \delta^i_{0jn}(x^*)\, dx_1 \cdots dx_n$$

for all j, all bounded, measurable sets S and all integers $n \geqq 0$. This equality follows from the Lebesgue theorem since $0 \leqq \delta_{0jn}(x^*) \leqq 1$ and S is bounded.

4. Relation of this paper to a result of Wald. The purpose of this paper is to give an essentially complete class of decision rules for the problem described above. The following theorem of Wald ([1], Theorem 3.19) is used in the proof. Let D_b denote the class of all decision rules whose risk functions are (uniformly) bounded in Ω. Let ζ denote a class of a priori cdf's such that for any $\xi_0 \in \zeta$ there is a sequence $\{\xi_i\}$ of members of ζ such that

$$\lim_{i=\infty} \xi_i(\omega) = \xi_0(\omega) \tag{13}$$

for any measurable subset ω of Ω. Let B_ζ denote the class of all Bayes solutions relative to members of ζ. Then the closure $\overline{B_\zeta}$ in the regular sense of B_ζ is essentially complete relative to D_b.

To obtain this result, several assumptions are made on the cost function, the loss function, the space D of decisions and the set of decision rules available to the experimenter. The verification of these assumptions under the more restrictive assumptions of this paper is mostly trivial and is omitted. See ([1], chap. 3).

An a priori cdf ξ will be called nondegenerate if it assigns positive probability to every open subset of Ω. It will now be shown that the class ζ of nondegenerate cdf's satisfies the hypothesis of Wald's theorem. Let ξ_0 be any cdf not in ζ. Let ξ^* be any specified cdf in ζ. Let $\{\epsilon_i\}$ $(i = 1, 2, \cdots)$ be a decreasing sequence of numbers such that $0 < \epsilon_i < 1$ and $\lim_{i=\infty} \epsilon_i = 0$. Define

$$\xi_i(\theta) = \epsilon_i \xi^*(\theta) + (1 - \epsilon_i)\xi_0(\theta). \tag{14}$$

Clearly $\xi_i(\theta)$ is a cdf in ζ for each i and we have for any measurable subset ω of Ω

$$\lim_{i=\infty} \xi_i(\omega) = \xi_0(\omega). \tag{15}$$

Hence the class ζ satisfies the hypothesis of Wald's Theorem 3.19.

5. The essentially complete class A and outline of the proof. If a decision rule δ depends on the observations only through $v = \sum_{i=1}^{n} x_i$ for each n $(n = 1, 2, \cdots)$ then $\delta_{jn}(v)$ will be used to denote the probability of accepting H_j given the pair (n, v). The class A of decision rules is defined as follows. The decision rule $\delta \varepsilon A$ if and only if for each positive integer n the following hold.

(i) δ depends on the observations only through v.

(ii) δ specifies a closed interval $J_n(\delta)$: $[a_{1n}(\delta), a_{2n}(\delta)]$ or simply J_n: $[a_{1n}, a_{2n}]$ which determines the action to be taken as follows.

(a) Another observation should be taken (i.e., $\delta_{0n}(v) = 1$) if $a_{1n} < v < a_{2n}$.

(b) Experimentation is stopped as soon as $v \not\varepsilon J_n$ and

(1) H_1 is accepted (i.e., $\delta_{1n}(v) = 1$) if $v < a_{1n}$

(2) H_2 is accepted (i.e., $\delta_{2n}(v) = 1$) if $v > a_{2n}$.

(c) If $a_{1n} < a_{2n}$ then $\delta_{2n}(a_{1n}) = \delta_{1n}(a_{2n}) = 0$.

Condition (c) can be omitted when μ is absolutely continuous since it concerns only a set of G_θ-measure zero for each $\theta \varepsilon \Omega$. The closed interval $[a_{1n}, a_{2n}]$ may reduce to a point but will always be nonempty; however, in the discrete case it need not contain any points of positive probability. Then the principal result of this paper will be that *the class A is essentially complete relative to D_b*. Clearly an immediate consequence will be:

COROLLARY 1. *The class A_b of all decision rules in A, whose risk functions are bounded in Ω, form an essentially complete class relative to D_b.*

The proof of the principal result consists in showing

(a) that A is essentially complete relative to $\bar{A}$, and

(b) that every Bayes solution (if altered at most on a set of measure zero) belongs to A and hence that $\overline{B_\zeta} \subset \bar{A}$. It then follows from Wald's Theorem 3.19 and (b) that $\bar{A}$ is essentially complete relative to D_b. Using (a) it follows that A has the same property. The paper ends with a corollary which shows that the same result holds if no indifference zone is given.

6. A is sequentially compact. In order to show that A is essentially complete relative to $\bar{A}$ it is clearly sufficient to prove the following theorem.

THEOREM 1. *For any $\delta \varepsilon \bar{A}$ there is a $\delta^* \varepsilon A$ which is equivalent to δ, that is, for which*

$$r(\theta, \delta^*) = r(\theta, \delta) \qquad \text{for all } \theta \varepsilon \Omega. \tag{16}$$

Proof. To prove the theorem it is sufficient to show that for any sequence $\{\delta^i\}$ $(i = 1, 2, \cdots)$ of members of A which is convergent in the regular sense

(i) there exists a limit (in the regular sense) δ which is in A and

(ii) any two limits (of the same sequence) are equivalent. In the discrete case a much stronger result holds, namely $A = \bar{A}$. Since in the discrete case the set of quantities $\delta_{jn}(x^*)$ for all possible triples (n, j, x^*) determines the decision function completely then by (9) the convergent sequence $\{\delta^i\}$ must have a unique limit δ^*. It remains only to show that $\delta^* \varepsilon A$.

In the absolutely continuous case it is easily seen by (11) that any two limits of the same sequence can differ at most on a set of n-dimensional Lebesgue measure zero for $n > 0$ and not at all for $n = 0$. Hence they must be equivalent. It therefore remains in both cases only to identify the limit δ^* and show that $\delta^* \varepsilon A$.

1. Consider the sequences $\{a_{\beta n}(\delta^i)\}$ $(i = 1, 2, \cdots)$ for $\beta = 1, 2$, and $n = 0, 1, 2, \cdots$. Since each element of each of these sequences lies in the same compact set [the real line with $\pm\infty$ added] then by the method of diagonalization there exists a subsequence $\{i_\alpha\}$ of the positive integers such that each of the above sequences converges. Let $a_{\beta n}$ denote the corresponding limits for each β and each n. Then clearly

$$-\infty \leqq a_{1n} \leqq a_{2n} \leqq \infty \qquad (n = 0, 1, 2, \cdots). \tag{17}$$

2. Consider the set of sequences $\{\delta_{jn}^{i_\alpha}(a_{\beta n})\}$ $(\alpha = 1, 2, \cdots)$ for $\beta = 1, 2; j = 0, 1, 2$ and $n = 0, 1, 2, \cdots$. Since each element of each of these sequences lies in the closed interval $[0, 1]$ which is compact, then by diagonalization there exists a subsequence $\{i_\gamma\}$ of $\{i_\alpha\}$ such that each of the above sequences converges. Let $\delta_{jn}(a_{\beta n})$ denote the corresponding limits for each β, each j and each n. Then clearly $0 \leqq \delta_{jn}(a_{\beta n}) \leqq 1$ for each triplet (β, j, n) and

$$\sum_{j=0}^{2} \delta_{jn}(a_{\beta n}) = 1 \quad (\beta = 1, 2; \quad n = 0, 1, 2, \cdots). \tag{18}$$

The symbol $\delta_{j0}(a_{\beta 0})$ denotes the constants δ_{j0} $(j = 0, 1, 2)$. Since $\{i_\gamma\}$ is a subsequence of $\{i_\alpha\}$ the limits $a_{\beta n}$ remain unchanged.

3. Let J_n denote the nonempty closed interval $[a_{1n}, a_{2n}]$ and let δ^* denote the following decision rule.

(i) $\delta_{j0}^* = \delta_{j0}$ $(j = 0, 1, 2)$.

(ii) If $a_{1n} < v < a_{2n}$ then $\delta_{0n}^*(v) = 1$. (Continue experimentation).

(iii) If $v \not\varepsilon J_n$ then $\delta_{0n}^*(v) = 0$ (Stop experimentation and)

(a) $\delta_{1n}^*(v) = 1$ for $v < a_{1n}$ (accept H_1)

(b) $\delta_{2n}^*(v) = 1$ for $v > a_{2n}$ (accept H_2).

(iv) If $v = a_{\beta n}$ then $\delta_{jn}^*(a_{\beta n}) = \delta_{jn}(a_{\beta n})$ $(\beta = 1, 2; \quad j = 0, 1, 2)$.

(Condition (iv) as well as paragraph 2 could be omitted in the absolutely continuous case since they concern only a set of measure zero.)

4. Clearly $\delta^* \varepsilon A$ and it remains only to show that δ^* is a limit of $\{\delta^i\}$ in the regular sense. The above construction was such that

$$\delta_{jn}^*(v) = \lim_{\gamma=\infty} \delta_{jn}^{i_\gamma}(v) \qquad \text{for each triple } (n, j, v), \tag{19}$$

that is, such that the subsequence $\{\delta^{i_\gamma}\}$ converges pointwise to δ^*. Hence by Section 3 $\{\delta^{i_\gamma}\}$ converges to δ^* in the regular sense, that is, for the discrete and absolutely continuous cases respectively

$$\lim_{\gamma=\infty} \delta_{jn}^{i_\gamma}(v) = \delta_{jn}^*(v) \qquad \text{for each triple } (j, n, v) \tag{20}$$

$$\lim_{\gamma=\infty} \delta_{0jn}^{i_\gamma}(S_n) = \delta_{0jn}^*(S_n) \qquad \text{for each triple } (j, n, S_n). \tag{21}$$

By hypothesis the full sequence $\{\delta^i\}$ is convergent in the regular sense so that the left members above converge for the full sequence. They must therefore converge to the same limits as the subsequences, namely, the right-hand members above. Hence δ^* is a limit in the regular sense of $\{\delta^i\}$. This proves Theorem 1.

7. Terminating property of the Bayes solutions in B_ζ. The proof given here that (almost) every Bayes solution with respect to a cdf $\xi \varepsilon \zeta$ belongs to A rests on the fact that every Bayes solution in B_ζ terminates in a finite number of steps with probability one. Assumptions 3 are used only for the latter result which we shall now prove after some definitions and lemmas.

Let ξ denote an arbitrary cdf over Ω. The *a* posterior probability that θ belongs to any Borel measurable subset ω of Ω given the triple (n, v, ξ) is

$$\int_\omega d\xi_v^n(\theta) = \frac{\int_\omega [\psi(\theta)]^n e^{v\theta}\, d\xi(\theta)}{\int_\Omega [\psi(\theta)]^n e^{v\theta}\, d\xi(\theta)}. \tag{22}$$

The notation ξ_v^n for the a posteriori cdf given the triple (n, v, ξ) is justified since the right member of (22) depends on the observations only through n and v. ξ_v^0 will then denote the original a priori cdf ξ for all v.

Following Wald ([1], Section 4.1.1) certain functions ρ_0, ρ_1, ρ_a, ρ_b will be defined which can be used to describe the class B_ξ of Bayes solutions relative to the a priori cdf ξ. Let D^*, D_0 and D_1 denote respectively the class of all decision rules, the class of all decision rules which require that no observations be taken with probability one and the class of all decision rules which require that at least one observation be taken with probability one. Wald defines for any triple (n, v, ξ),

$$\rho_0(\xi_v^n) = \inf_{\delta \varepsilon D_0} r(\xi_v^n, \delta) \tag{23}$$

$$\rho(\xi_v^n) = \inf_{\delta \varepsilon D^*} r(\xi_v^n, \delta) \tag{24}$$

$$\rho_1(\xi_v^n) = \inf_{\delta \varepsilon D_1} r(\xi_v^n, \delta). \tag{25}$$

(Actually the function $\rho(\xi_v^n)$ depends on the sequence $(c_{n+1}, c_{n+2}, \cdots)$ which may vary with n and should be written $\rho^n(\xi_v^n)$ unless all the c_i $(i = 1, 2, \cdots)$ are equal. To simplify notation and since the proofs are not affected it will be assumed that the superscript on ξ also applies to ρ. The same remark applies to (25).) Since $D_0 \subset D^*$ then clearly for any triple (n, v, ξ)

$$f^0(\xi_v^n) = \rho_0(\xi_v^n) - \rho(\xi_v^n) \geqq 0. \tag{26}$$

Wald has shown ([1], Theorem 4.2 and 4.1) that for any cdf ξ_v^n

$$\rho(\xi_v^n) = \min [\rho_0(\xi_v^n), \rho_1(\xi_v^n)] \tag{27}$$

and that

$$\rho_1(\xi_v^n) = \int_\Omega \int_{-\infty}^{\infty} \rho(\xi_{v+y}^{n+1})\, dG_\theta(y)\, d\xi_v^n(\theta) + c_{n+1}. \tag{28}$$

If $g_\theta(x)$ denotes the integrand in (2) then (28) can be written as

$$\rho_1(\xi_v^n) = \int_\Omega \int_{-\infty}^{\infty} \rho(\xi_{v+y}^{n+1}) g_\theta(y)\, d\mu(y)\, d\xi_v^n(\theta) + c_{n+1}. \tag{29}$$

Denote by ω_1, ω_2 respectively the closed intervals $[\theta_2, \bar{\theta}]$, $[\underline{\theta}, \theta_1]$ in Ω. Let

$$\rho_a(\xi_v^n) = \int_\Omega W(\theta, 1)\, d\xi_v^n(\theta) = \int_{\omega_1} W(\theta, 1)\, d\xi_v^n(\theta) \tag{30}$$

$$\rho_b(\xi_v^n) = \int_\Omega W(\theta, 2)\, d\xi_v^n(\theta) = \int_{\omega_2} W(\theta, 2)\, d\xi_v^n(\theta) \tag{31}$$

where the last expression in each case follows from (5). Clearly, $\rho_a(\xi_v^n)$ and $\rho_b(\xi_v^n)$ denote respectively the risks of accepting H_1 and H_2 given the triple (n, v, ξ). Then by (23)

$$\rho_0(\xi_v^n) = \min [\rho_a(\xi_v^n), \rho_b(\xi_v^n)]. \tag{32}$$

The class B_ξ of Bayes solutions relative to the fixed a priori cdf ξ can be described as follows.

(i) If $\rho_1(\xi_v^n) < \rho_0(\xi_v^n)$ for some pair (n, v) then for any $\delta \not\in D_1$ there exists a decision rule $\delta_1 \in D_1$ such that $r(\xi_v^n, \delta_1) < r(\xi_v^n, \delta)$. Hence for such a pair (n, v) any Bayes solution $\delta_\xi \in D_1$, that is, any Bayes solution will prescribe another observation with probability one.

(ii) If $\rho_a(\xi_v^n) < \min [\rho_b(\xi_v^n), \rho_1(\xi_v^n)]$ for some pair (n, v) then similarly any Bayes solution δ_ξ will accept H_1 with probability one when the pair (n, v) is observed.

(iii) A similar result holds for H_2 when $\rho_b(\xi_v^n) < \min [\rho_a(\xi_v^n), \rho_1(\xi_v^n)]$.

(iv) If $\rho_a(\xi_v^n) = \rho_b(\xi_v^n) < \rho_1(\xi_v^n)$ for some pair (n, v) then any Bayes solution δ_ξ will accept H_1, H_2 with probabilities p, $1 - p$ respectively where $0 \leqq p \leqq 1$ when the pair (n, v) is observed. (It will be a consequence of Lemmas 3 and 5 that when μ is absolutely continuous the equalities in (iv) through (vii) take place at most on a set of v points of G_θ-measure zero for each $\theta \in \Omega$. The Bayes solution can be arbitrary on such a set since it is defined only up to a set of measure zero.)

(v) If $\rho_1(\xi_v^n) = \rho_a(\xi_v^n) < \rho_b(\xi_v^n)$ for some pair (n, v) then any Bayes solution δ_ξ will accept H_1 and take another observation with probabilities p and $1 - p$ respectively where $0 \leqq p \leqq 1$ when the pair (n, v) is observed. (In (v) through (vii) it is assumed that the value $\rho_1(\xi_v^n)$ is attainable with some $\delta \in D_1$ for each pair (n, v). This is actually a consequence of ([1], Theorem 3.2).)

(vi) A similar result holds when $\rho_1(\xi_v^n) = \rho_b(\xi_v^n) < \rho_a(\xi_v^n)$.

(vii) If $\rho_1(\xi_v^n) = \rho_a(\xi_v^n) = \rho_b(\xi_v^n)$ for some pair (n, v) then any randomized or nonrandomized decision rule is a Bayes solution when the pair (n, v) is observed.

Let $\hat{\theta}$ denote the maximum likelihood estimate of θ given the pair (n, v). Since $\hat{\theta}$ depends on v/n it is convenient in Theorem 2 to consider the statistic $\bar{x} = v/n$ instead of v itself. Then $\xi_{\bar{x}}^n$ will denote the a posteriori cdf given the triple $(n, \bar{x}, \xi)$.

LEMMA 1. *The function $\hat{\theta}(\bar{x})$ which is uniquely defined by Assumption* 3(i) *is nondecreasing.*

PROOF. Consider two points $\bar{x}_1 < \bar{x}_2$ and let $\hat{\theta}_1 = \hat{\theta}(\bar{x}_1)$ and $\hat{\theta}_2 = \hat{\theta}(\bar{x}_2)$. By assumption 3(i) we have for all $\theta \varepsilon \Omega$

$$\psi(\hat{\theta}_1)e^{\bar{x}_1\hat{\theta}_1} \geqq \psi(\theta)e^{\bar{x}_1\theta} \tag{33}$$

$$\psi(\hat{\theta}_2)e^{\bar{x}_2\hat{\theta}_2} \geqq \psi(\theta)e^{\bar{x}_2\theta} \tag{34}$$

Hence putting $\theta = \hat{\theta}_2$ in (33) and $\theta = \hat{\theta}_1$ in (34) and dividing yields

$$\hat{\theta}_2(\bar{x}_2 - \bar{x}_1) \geqq \hat{\theta}_1(\bar{x}_2 - \bar{x}_1).$$

Since $\bar{x}_2 > \bar{x}_1$, it follows that

$$\hat{\theta}_2 = \hat{\theta}(\bar{x}_2) \geqq \hat{\theta}(\bar{x}_1) = \hat{\theta}_1 . \qquad \text{Q.E.D.} \tag{35}$$

THEOREM 2. *If $\xi \varepsilon \zeta$ then there exists an integer $N = N(\xi)$ such that any Bayes solution δ_ξ relative to ξ will terminate before $N + 1$ observations with probability one.*

PROOF. Let $\bar{x}_1 < \bar{x}_2$ denote the two points mentioned in Assumption 3. By Lemma 1 and this assumption $\theta_1 < \hat{\theta}(\bar{x}_1) < \hat{\theta}(\bar{x}_2) < \theta_2$. Define

$$\theta_0 = \tfrac{1}{2}[\hat{\theta}(\bar{x}_1) + \hat{\theta}(\bar{x}_2)].$$

Let δ^t denote the following terminal decision rule:

"Accept H_1 if $\hat{\theta}(\bar{x}) \leqq \theta_0$, accept H_2 if $\hat{\theta}(\bar{x}) > \theta_0$."

Let $S_{\bar{x}}^1$ and $S_{\bar{x}}^2$ denote respectively the set of points in $S_{\bar{x}}$ for which $\hat{\theta}(\bar{x}) \leqq \theta_0$ and $\hat{\theta}(\bar{x}) > \theta_0$. These sets are clearly not empty. If δ^t is used the average risk given the triple $(n, \bar{x}, \xi)$ is

$$r(\xi_{\bar{x}}^n, \delta^t) = \begin{cases} \int_{\omega_1} W(\theta, \delta^t)\, d\xi_{\bar{x}}^n(\theta) & \text{if } \bar{x} \varepsilon S_{\bar{x}}^1 \\ \int_{\omega_2} W(\theta, \delta^t)\, d\xi_{\bar{x}}^n(\theta) & \text{if } \bar{x} \varepsilon S_{\bar{x}}^2 \end{cases} \tag{36}$$

In order to show that

$$\lim_{n=\infty} r(\xi_{\bar{x}}^n, \delta^t) = 0 \qquad \text{uniformly for all } \bar{x} \varepsilon S_{\bar{x}} \tag{37}$$

it is sufficient to show that the upper value in (36) tends to zero uniformly for all $\bar{x} \varepsilon S_{\bar{x}}^1$ and a similar result for the lower value. Since the two cases are alike, only the first will be shown. Since by assumption $W(\theta, \delta^t)$ is a bounded function of θ and since $g_{\bar{x}}(\theta) > 0$ then using (22) with v replaced by $n\bar{x}$ it is sufficient for (37) to show that

$$\lim_{n=\infty} \left[\frac{\int_{\omega_1} [g_{\bar{x}}(\theta)]^n \, d\xi(\theta)}{\int_{\omega_0} [g_{\bar{x}}(\theta)]^n \, d\xi(\theta)} \right] = 0 \qquad \text{uniformly for } \bar{x} \,\varepsilon\, S_{\bar{x}}^1 \tag{38}$$

where ω_0 is any subset of Ω. Take for ω_0 the interval $(\theta_0, \hat{\theta}_2)$ where $\hat{\theta}_2 = \hat{\theta}(\bar{x}_2)$. For any n by Assumption 3(i) the function $[g_{\bar{x}}(\theta)]^n$ is strictly decreasing in the interval $(\hat{\theta}(\bar{x}), \bar{\theta})$ and hence also in the subintervals $(\theta_0, \hat{\theta}_2)$ and $(\theta_2, \bar{\theta})$. Hence an upper bound for the expression in brackets in (38) is

$$\frac{[g_{\bar{x}}(\theta_2)]^n}{[g_{\bar{x}}(\hat{\theta}_2)]^n} \cdot \frac{1}{\xi(\omega_0)}. \tag{39}$$

Since $\xi \,\varepsilon\, \zeta$ and $\theta_0 < \hat{\theta}_2$ then $\xi(\omega_0)$ is a positive constant. By (8)

$$\left[\frac{g_{\bar{x}}(\theta_2)}{g_{\bar{x}}(\hat{\theta}_2)} \right]^n = \left[\frac{\psi(\theta_2)}{\psi(\hat{\theta}_2)} e^{\bar{x}(\theta_2 - \hat{\theta}_2)} \right]^n. \tag{40}$$

Since $\theta_0 < \hat{\theta}_2 = \hat{\theta}(\bar{x}_2)$ it follows from Lemma 1 that $\bar{x}_2$ is an upper bound for the set $S_{\bar{x}}^1$. Hence by (40) since $\theta_2 > \hat{\theta}_2$ it is sufficient to show that

$$\lim_{n=\infty} \left[\frac{g_{\bar{x}_2}(\theta_2)}{g_{\bar{x}_2}(\hat{\theta}_2)} \right]^n = 0. \tag{41}$$

The function $g_{\bar{x}_2}(\theta)$ is strictly decreasing in the interval $(\hat{\theta}_2, \theta_2)$. The expression in brackets in (41) is therefore a positive constant less than unity and (41) follows. This proves (37).

It follows from (41) that the approach to zero is at least exponentially fast. On the other hand, by Assumption 2 the constants c_n may approach zero but not fast enough for $\sum_{i=1}^{\infty} c_i$ to converge. It follows from (7) that there exists an integer $N = N(\xi)$ such that for $n \geqq N$

$$0 \leqq \rho_0(\xi_{\bar{x}}^n) \leqq r(\xi_{\bar{x}}^n, \delta^t) < c_{n+1} \qquad \text{for all } \bar{x} \,\varepsilon\, S_{\bar{x}}. \tag{42}$$

By (28) it follows that $\rho_1(\xi_{\bar{x}}^n) \geqq c_{n+1}$ for all pairs $(n, \bar{x})$ and hence for $n \geqq N$

$$\rho_0(\xi_{\bar{x}}^n) < \rho_1(\xi_{\bar{x}}^n) \qquad \text{for all } \bar{x} \,\varepsilon\, S_{\bar{x}}. \tag{43}$$

It follows from the description of B_ξ above that any Bayes solution δ_ξ will terminate experimentation before $N + 1$ observations with probability one. This proves Theorem 2.

The above result does not hold in the case of testing a simple hypothesis against a simple alternative. For example every Wald sequential probability ratio test which consists of a pair of parallel lines does not have the above property. It would be interesting to determine whether we can restrict our attention to tests consisting of pairs of converging straight lines. Although Theorem 2 is useful here as a tool it appears to have some interest in its own right.

Let S_v^n denote the interval generated by $v = n\bar{x}$ as $\bar{x}$ varies over $S_{\bar{x}}$. Since (43) holds for all $\bar{x} \varepsilon S_{\bar{x}}$ it will also hold for all $v \varepsilon S_v^n$, that is,

$$\rho_0(\xi_v^n) < \rho_1(\xi_v^n) \qquad \text{for all } v \varepsilon S_v^n. \tag{44}$$

Hence by (27) for $n \geqq N$

$$\rho(\xi_v^n) = \rho_0(\xi_v^n) \qquad \text{for all } v \varepsilon S_v^n. \tag{45}$$

In the normal case the sets S_v^n are equal to the real line for each n. If the sets S_v^n actually depend on n, as in the binomial case, the range of validity of (44) and (45) varies with n only because ξ_v^n may not be defined for all v. Hence (44) and (45) hold whenever the expressions therein are defined.

8. Other properties of the Bayes solutions in B_ζ. The purpose of this section is to show that (almost) every Bayes solution δ_ξ relative to a $\xi \varepsilon \zeta$ belongs to A. In the discrete case this holds for all Bayes solutions δ_ξ with $\xi \varepsilon \zeta$. In the absolutely continuous case the result is slightly weaker, namely, for each $\xi \varepsilon \zeta$ any Bayes solution δ_ξ' not in A differs from another Bayes solution δ_ξ in A at most on a set of n dimensional Lebesgue measure zero for each positive integer n.

Let

$$f_1^0(\xi_v^n) = \rho_0(\xi_v^n) - \rho_1(\xi_v^n) \tag{46}$$

$$f_a^b(\xi_v^n) = \rho_b(\xi_v^n) - \rho_a(\xi_v^n). \tag{47}$$

Thus $f_1^0(\xi_v^n)$ can be regarded as the advantage of taking another observation over stopping for the given triple (n, v, ξ). Similarly $f_a^b(\xi_v^n)$ is the advantage of accepting H_1 over accepting H_2 for the given triple (n, v, ξ).

It follows from the description of B_ξ above that in order to prove the desired result it is sufficient to show the following for each $\xi \varepsilon \zeta$.

For each positive integer n there exists a nonempty closed interval

$$J_n(\xi): [a_{1n}(\xi), a_{2n}(\xi)] \text{ or simply } J_n : [a_{1n}, a_{2n}]$$

such that

$$\rho_0(\xi_v^n) > \rho_1(\xi_v^n) \text{ [or } f_1^0(\xi_v^n) > 0] \qquad \text{for } a_{1n} < v < a_{2n}, \tag{48}$$

$$\rho_0(\xi_v^n) < \rho_1(\xi_v^n) \text{ [or } f_1^0(\xi_v^n) < 0] \qquad \text{for } v \not\varepsilon J_n, \tag{49}$$

$$\rho_b(\xi_v^n) > \rho_a(\xi_v^n) \text{ [or } f_a^b(\xi_v^n) > 0] \qquad \text{for } v < a_{1n}, \tag{50}$$

$$\rho_b(\xi_v^n) < \rho_a(\xi_v^n) \text{ [or } f_a^b(\xi_v^n) < 0] \qquad \text{for } v > a_{2n}, \tag{51}$$

and if $a_{1n} < a_{2n}$

$$\rho_b(\xi_v^n) > \min [\rho_a(\xi_v^n), \rho_1(\xi_v^n)] \qquad \text{for } v = a_{1n}, \tag{52}$$

$$\rho_a(\xi_v^n) > \min [\rho_b(\xi_v^n), \rho_1(\xi_v^n)] \qquad \text{for } v = a_{2n}. \tag{53}$$

Equations (52) and (53) are superfluous for the absolutely continuous case since they are concerned with a set of G_θ-measure zero for each $\theta \varepsilon \Omega$. In other words, the Bayes solution obtained by pointwise minimization of the average risk is to continue as long as $a_{1n} < v < a_{2n}$. Experimentation is stopped as soon as $v \not\varepsilon J_n$. Then H_1 or H_2 is accepted according to which involves the smallest risk. To show for any $\xi \varepsilon \zeta$ the existence of $J_n(\xi)$ satisfying (48) through (53) for each n will require several lemmas and theorems. Let

$$f_0^b(\xi_v^n) = \rho_b(\xi_v^n) - \rho_0(\xi_v^n) = \max\,[f_a^b(\xi_v^n), 0] \tag{54}$$

where the latter equality follows from (32) and (47). In what follows n will denote an integer, v a point in S_v^n and ϑ a point in Ω even if this is not explicitly stated. Let

$$p_v^n(\theta) = \frac{[\psi(\theta)]^n e^{v\theta}}{\int_\Omega [\psi(\theta)]^n e^{v\theta}\, d\xi(\theta)}. \tag{55}$$

Then $p_v^n(\theta)$ is a probability law relative to the measure $\xi(\theta)$.

LEMMA 2. *Let $\xi(\theta)$ be any cdf on Ω. Let $W(\theta)$ be a bounded, nonincreasing function of θ which is not constant on a set of ξ-measure unity. Then for all $n > 0$ and all pairs $v_1 < v_2$*

$$\int_\Omega W(\theta)\, d\xi_{v_1}^n(\theta) > \int_\Omega W(\theta)\, d\xi_{v_2}^n(\theta). \tag{56}$$

If the last condition on $W(\theta)$ is omitted then the weak inequality holds.

PROOF.[3] Since $v_2 > v_1$ the ratio $p_{v_2}^n(\theta)/p_{v_1}^n(\theta)$ is a strictly increasing function of ϑ. Let $\Omega+$ and $\Omega-$ denote respectively the intervals on which $p_{v_2}^n(\theta) >$ and $< p_{v_1}^n(\theta)$. Then

$$\begin{aligned}\int_\Omega W(\theta)\, d[\xi_{v_2}^n(\theta) - \xi_{v_1}^n(\theta)] &= \int_{\Omega+} W(\theta)[p_{v_2}^n(\theta) - p_{v_1}^n(\theta)]\, d\xi(\theta) \\ &\quad - \int_{\Omega-} W(\theta)[p_{v_1}^n(\theta) - p_{v_2}^n(\theta)]\, d\xi(\theta)\end{aligned} \tag{57}$$

Since $W(\theta)$ is nonincreasing there is a constant c such that

$$\int_{\Omega+} W(\theta)[p_{v_2}^n(\theta) - p_{v_1}^n(\theta)]\, d\xi(\theta) \leqq c \int_{\Omega+} [p_{v_2}^n(\theta) - p_{v_1}^n(\theta)]\, d\xi(\theta) \tag{58}$$

$$\int_{\Omega-} W(\theta)[p_{v_1}^n(\theta)] - p_{v_2}^n(\theta)\, d\xi(\theta) \geqq c \int_{\Omega-} [p_{v_1}^n(\theta) - p_{v_2}^n(\theta)]\, d\xi(\theta). \tag{59}$$

If $W(\theta)$ is not constant everywhere (ξ) at least one of the above inequalities can be replaced by the strict inequality. Since $p_{v_1}^n(\theta)$ and $p_{v_2}^n(\theta)$ are both probability laws relative to ξ then

$$\int_{\Omega+} [p_{v_2}^n(\theta) - p_{v_1}^n(\theta)]\, d\xi(\theta) = \int_{\Omega-} [p_{v_1}^n(\theta) - p_{v_2}^n(\theta)]\, d\xi(\theta). \tag{60}$$

---

[3] The proof of this lemma was kindly offered by Professor Erich L. Lehmann.

It follows from (57), by the use of (60) and the revised inequalities (58) and (59), that

$$\int_\Omega W(\theta)\, d[\xi^n_{v_2}(\theta) - \xi^n_{v_1}(\theta) < 0 \tag{61}$$

which proves the lemma.

Clearly if $\xi \varepsilon \zeta$ then it is sufficient in the above lemma to assume that $W(\theta)$ is bounded, nondecreasing and not constant throughout Ω.

COROLLARY 2. *Let $f(y)$ be a bounded nonincreasing function of $y = x_{n+1}$ which is not constant on a set of μ-measure unity. Then for all pairs $\theta_1 < \theta_2$*

$$\int_{-\infty}^{\infty} f(y)\, dG_{\theta_1}(y) > \int_{-\infty}^{\infty} f(y)\, dG_{\theta_2}(y). \tag{62}$$

PROOF. Since for $\theta_2 > \theta_1$ the ratio $g_{\theta_2}(y)/g_{\theta_1}(y)$ is a strictly increasing function of y the proof is exactly as in Lemma 2.

COROLLARY 3. *Let $f_1(y)$ and $f_2(y)$ be bounded functions of $y = x_{n+1}$ such that $f_1(y)$ is nonincreasing and*

$$f_1(y) \geqq f_2(y) \qquad \textit{for all } y. \tag{63}$$

Then for any ξ, all $n > 0$ and all pairs $v_1 < v_2$

$$\int_\Omega \int_{-\infty}^{\infty} f_1(y)\, dG_\theta(y)\, d\xi^n_{v_1}(\theta) \geqq \int_\Omega \int_{-\infty}^{\infty} f_2(y)\, dG_\theta(y)\, d\xi^n_{v_2}(\theta). \tag{64}$$

PROOF. Since $f_1(y)$ is nonincreasing then by Corollary 2 for any pair $\theta_1 > \theta_2$

$$W_1(\theta_1) = \int_{-\infty}^{\infty} f_1(y)\, dG_{\theta_1}(y) \geqq \int_{-\infty}^{\infty} f_1(y)\, dG_{\theta_2}(y) = W_1(\theta_2), \tag{65}$$

that is, $W_1(\theta)$ defined in (65) is a nonincreasing function of θ. If we define $W_2(\theta)$ similarly with f_1 replaced by f_2 then by (63)

$$W_1(\theta) \geqq W_2(\theta) \qquad \text{for all } \theta \varepsilon \Omega. \tag{66}$$

Also since $f_i(y)$ $(i = 1, 2)$ are bounded functions of y then clearly $W_i(\theta)$ are bounded functions of θ $(i = 1, 2)$. Hence for all pairs $v_1 < v_2$ by Lemma 2

$$\int_\Omega W_1(\theta)\, d\xi^n_{v_1}(\theta) \geqq \int_\Omega W_1(\theta)\, d\xi^n_{v_2}(\theta) \tag{67}$$

and by (66)

$$\int_\Omega W_1(\theta)\, d\xi^n_{v_2}(\theta) \geqq \int_\Omega W_2(\theta)\, d\xi^n_{v_2}(\theta). \tag{68}$$

The result (64) follows from (67) and (68).

COROLLARY 4. *If in addition to the hypothesis of Corollary 3 the inequality (63) is strict on a set S^* of positive μ-measure then the inequality (64) is a strict one.*

PROOF. Since $g_\theta(y) > 0$ for all y and all θ the added hypothesis implies that a strict inequality holds in (66) for all $\theta \varepsilon \Omega$. The strict inequality will then hold in (68) and hence also in the final result (64).

LEMMA 3. *For any $\xi \varepsilon \zeta$, all $n > 0$ and all pairs $v_1 < v_2$*

$$f_a^b(\xi_{v_1}^n) > f_a^b(\xi_{v_2}^n), \tag{69}$$

$$f_0^b(\xi_{v_1}^n) \geqq f_0^b(\xi_{v_2}^n) \geqq 0, \tag{70}$$

$$f_0^b(\xi_{v_1}^n) > f_0^b(\xi_{v_2}^n) \qquad \text{if } f_0^b(\xi_{v_1}^n) > 0. \tag{71}$$

In words, as v increases

(i) the advantage of accepting H_2 over accepting H_1 increases,

(ii) the regret in accepting H_2 if we have to make a decision without further observations is nonincreasing. If there is any regret at all it must actually decrease.

PROOF. The functions $-W(\theta, 1)$ and $W(\theta, 2)$ satisfy the assumptions of Lemma 2 for any $\xi \varepsilon \zeta$. Replacing $W(\theta)$ by $W(\theta, 2)$ and $-W(\theta, 1)$ in (56) gives respectively

$$\rho_b(\xi_{v_1}^n) > \rho_b(\xi_{v_2}^n), \tag{72}$$

$$\rho_a(\xi_{v_1}^n) < \rho_a(\xi_{v_2}^n). \tag{73}$$

Subtracting (73) from (72) gives the desired result (69). With the aid of the last equality of (54) the results in (70) are immediate consequences of (69). To prove (71) note that if $f_0^b(\xi_{v_1}^n) > 0$ then by (54) and (69)

$$f_0^b(\xi_{v_1}^n) = f_a^b(\xi_{v_1}^n) > f_a^b(\xi_{v_2}^n). \tag{74}$$

By (54) the value of $f_0^b(\xi_{v_2}^n)$ is either $f_a^b(\xi_{v_2}^n)$ or zero. In the former case (71) follows from (74) and in the latter case the result follows from the assumption that $f_0^b(\xi_{v_1}^n) > 0$. This proves Lemma 3. Let

$$f_1^b(\xi_v^n) = \rho_b(\xi_v^n) - \rho_1(\xi_v^n), \tag{75}$$

$$f_1^a(\xi_v^n) = \rho_a(\xi_v^n) - \rho_1(\xi_v^n). \tag{76}$$

It will be shown that f_1^b is nonincreasing and that f_1^a is nondecreasing in v when $\xi \varepsilon \zeta$ and $n > 0$ are fixed and that these monotonicities are strict whenever the corresponding functions are nonnegative. It will also be shown that if we define

$$f_1^0(\xi_v^n) = \min\,[f_1^b(\xi_v^n), f_1^a(\xi_v^n)] = \rho_0(\xi_v^n) - \rho_1(\xi_v^n) \tag{77}$$

then there exist two points $a_{1n} \leqq a_{2n}$ such that

$$f_1^0(\xi_v^n) \geqq 0 \text{ if and only if } a_{1n} \leqq v \leqq a_{2n}$$

and if $a_{1n} < a_{2n}$ then

$$f_1^0(\xi_v^n) > 0 \qquad \text{for } a_{1n} < v < a_{2n}.$$

These are the points involved in equations (48) through (53). To show that $f_1^b(\xi_v^n)$ is a nonincreasing function of v (Lemma 5) the following result is needed.

LEMMA 4. *For any cdf ξ, all $n \geqq 0$ and all v*

$$\rho_b(\xi_v^n) = \int_\Omega \int_{-\infty}^{\infty} \rho_b(\xi_{v+y}^{n+1})\, dG_\theta(y)\, d\xi_v^n(\theta) \tag{78}$$

where $y = x_{n+1}$.

The proof is omitted. Intuitively the lemma says "if there is no cost of taking observations and H_2 is going to be accepted after the next observation then nothing is gained or lost by accepting H_2 now."

LEMMA 5. *For each $\xi \varepsilon \zeta$, all $n > 0$ and all pairs $v_1 < v_2$*

$$f_1^b(\xi_{v_1}^n) \geqq f_1^b(\xi_{v_2}^n) \tag{79}$$

$$f_1^b(\xi_{v_1}^n) > f_1^b(\xi_{v_2}^n) \qquad \textit{if } f_1^b(\xi_{v_1}^n) \geqq 0. \tag{80}$$

PROOF. An induction on n will be used. The theorem is first shown for $n \geqq N = N(\xi)$ defined in Theorem 2 for each $\xi \varepsilon \zeta$. Substituting (78) and (28) in (75) and using the fact that (45) holds for $n \geqq N$ gives

$$f_1^b(\xi_{v_1}^n) = \int_\Omega \int_{-\infty}^{\infty} f_0^b(\xi_{v_1+y}^{n+1})\, dG_\theta(y)\, d\xi_{v_1}^n(\theta) - c_{n+1} \tag{81}$$

$$f_1^b(\xi_{v_2}^n) = \int_\Omega \int_{-\infty}^{\infty} f_0^b(\xi_{v_2+y}^{n+1})\, dG_\theta(y)\, d\xi_{v_2}^n(\theta) - c_{n+1} \tag{82}$$

where $y = x_{n+1}$.

(i) Corollary 3 will now be used with $f_1(y)$ and $f_2(y)$ replaced by $f_0^b(\xi_{v_1+y}^{n+1})$ and $f_0^b(\xi_{v_2+y}^{n+1})$ respectively. The functions f_0^b are clearly bounded by (54). By (70) $f_0^b(\xi_{v_1+y}^{n+1})$ is a nonincreasing function of y and (63) also follows from (70). It therefore follows by Corollary 3 that the double integral in (81) is not less than the double integral in (82). This proves (79) for $n \geqq N$.

(ii) If v_1 is such that $f_1^b(\xi_{v_1}^n) \geqq 0$ than by (81)

$$\int_{-\infty}^{\infty} \{f_0^b(\xi_{v_1+y}^{n+1}) - c_{n+1}\} \left[\int_\Omega g_\theta(y)\, d\xi_{v_1}^n(\theta)\right] d\mu(y) \geqq 0 \tag{83}$$

where $g_\theta(y)$ is the integrand in (2). Since for each y the function $g_\theta(y) > 0$ for each θ the expression in brackets must also be positive for each y. Hence

$$f_0^b(\xi_{v_1+y}^{n+1}) \geqq c_{n+1} > 0 \tag{84}$$

on a set S^* of positive μ-measure. Then by (71) since $v_1 < v_2$

$$f_0^b(\xi_{v_1+y}^{n+1}) > f_0^b(\xi_{v_2+y}^{n+1}) \qquad \text{for } y \varepsilon S^*. \tag{85}$$

It follows from Corollary 4 that (80) holds for $n \geqq N$.

If $N = 1$ the theorem is proved. Otherwise assuming the theorem holds for $n = k + 1, k + 2, \cdots, N$ $(N \geqq k + 1 > 1)$ it will be shown to hold for $n = k$. Let

$$f^b(\xi_v^n) = \rho_b(\xi_v^n) - \rho(\xi_v^n) = \max\, [f_0^b(\xi_v^n), f_1^b(\xi_v^n)] \tag{86}$$

where the last equality follows from (27). Substituting (78) and (28) in (75) and using (86) gives for any n (and in particular for $n = k$)

$$f_1^b(\xi_{v_1}^k) = \int_\Omega \int_{-\infty}^{\infty} f^b(\xi_{v_1+y}^{k+1})\, dG_\theta(y)\, d\xi_{v_1}^k(\theta) - c_{k+1} \tag{87}$$

$$f_1^b(\xi_{v_2}^k) = \int_\Omega \int_{-\infty}^{\infty} f^b(\xi_{v_2+y}^{k+1})\, dG_\theta(y)\, d\xi_{v_2}^k(\theta) - c_{k+1} \tag{88}$$

where $y = x_{k+1}$.

(i) Corollary 3 will again be used with $f_1(y)$ and $f_2(y)$ replaced by $f^b(\xi_{v_1+y}^{k+1})$ and $f^b(\xi_{v_2+y}^{k+1})$ respectively. The boundedness of these functions follows from (86) and (27). It remains only to show that $f^b(\xi_{v_1+y}^{k+1})$ is a nonincreasing function of y since (63) follows easily from this result. Let $y_1 < y_2$ denote two possible values of y. By (70)

$$f_0^b(\xi_{v_1+y_1}^{k+1}) \geqq f_0^b(\xi_{v_1+y_2}^{k+1}), \tag{89}$$

and by the induction hypothesis on (79)

$$f_1^b(\xi_{v_1+y_1}^{k+1}) \geqq f_1^b(\xi_{v_1+y_2}^{k+1}). \tag{90}$$

Hence using the last equality of (86)

$$f^b(\xi_{v_1+y_1}^{k+1}) \geqq f^b(\xi_{v_1+y_2}^{k+1}). \tag{91}$$

It therefore follows from Corollary 3 that the right member of (87) is not less than the right member of (88). This proves (79) for $n = k$ and hence for all $n > 0$.

(ii) If v_1 is such that $f_1^b(\xi_{v_1}^k) \geqq 0$ then proceeding as in (83) and (84) we obtain from (87) and (86) the result

$$\max\,[f_0^b(\xi_{v_1+y}^{k+1}), f_1^b(\xi_{v_1+y}^{k+1})] = f^b(\xi_{v_1+y}^{k+1}) \geqq c_{k+1} > 0 \tag{92}$$

on a set S^* of positive μ-measure. Consider any $y \,\varepsilon\, S^*$.

Case 1. If f_0^b is the larger in (92) then by (71), (92) and (79)

$$f^b(\xi_{v_1+y}^{k+1}) = f_0^b(\xi_{v_1+y}^{k+1}) > f_0^b(\xi_{v_2+y}^{k+1}), \tag{93}$$

$$f^b(\xi_{v_1+y}^{k+1}) > f_1^b(\xi_{v_1+y}^{k+1}) \geqq f_1^b(\xi_{v_2+y}^{k+1}). \tag{94}$$

Case 2. If f_1^b is the larger in (92) then by (70), (92) and the induction hypothesis for (80)

$$f^b(\xi_{v_1+y}^{k+1}) = f_1^b(\xi_{v_1+y}^{k+1}) > f_1^b(\xi_{v_2+y}^{k+1}), \tag{95}$$

$$f^b(\xi_{v_1+y}^{k+1}) > f_0^b(\xi_{v_1+y}^{k+1}) \geqq f_0^b(\xi_{v_2+y}^{k+1}). \tag{96}$$

Case 3. If $f_1^b = f_0^b$ in (92) then both (93) and (95) hold. Hence in each case for all $y \,\varepsilon\, S^*$

$$f^b(\xi_{v_1+y}^{k+1}) > \max\,[f_0^b(\xi_{v_2+y}^{k+1}), f_1^b(\xi_{v_2+y}^{k+1})] = f^b(\xi_{v_2+y}^{k+1}). \tag{97}$$

It follows from Corollary 4 that the right member of (87) is greater than the right member of (88). This proves (80) for $n = k$ and hence for all $n > 0$.

If we define f_b^a, f_0^a, f_1^a and f^a in exactly the same manner as f_a^b, f_0^b, f_1^b and f^b respetively except that ρ_a and ρ_b are interchanged, then the following theorem can be proved in a manner completely analogous to that above.

LEMMA 6. *For each $\xi \varepsilon \zeta$, all $n > 0$ and all pairs $v_1 < v_2$*

$$f_1^a(\xi_{v_1}^n) \leqq f_1^a(\xi_{v_2}^n) \tag{98}$$

$$f_1^a(\xi_{v_1}^n) < f_1^a(\xi_{v_2}^n) \qquad \text{if } f_1^a(\xi_{v_2}^n) \geqq 0. \tag{99}$$

Just to indicate why the inequalities are reversed the reader will note that the analogue of (69) holds with reversed sign since $f_a^b = -f_b^a$.

THEOREM 3. *For each $\xi \varepsilon \zeta$ and each $n > 0$ there exists a nonempty closed interval*

$$J_n(\xi): [a_{1n}(\xi), a_{2n}(\xi)]$$

such that equations (48) *through* (53) *hold.*

PROOF. Define $a_{2n} = a_{2n}(\xi)$ as the greatest lower bound of values of v for which

$$\rho_b(\xi_v^n) \leqq \min [\rho_a(\xi_v^n), \rho_1(\xi_v^n)] \tag{100}$$

if any such values exist and define it as ∞ otherwise. Define $a_{1n} = a_{1n}(\xi)$ as the least upper bound of values of v for which

$$\rho_a(\xi_v^n) \leqq \min [\rho_b(\xi_v^n), \rho_1(\xi_v^n)] \tag{101}$$

if any such values exist and define it as $-\infty$ otherwise.

(i) To show that $a_{1n} \leqq a_{2n}$ we note that for any v_0 satisfying (100) by (69)

$$f_a^b(\xi_v^n) < f_a^b(\xi_{v_0}^n) \leqq 0 \qquad \text{for } v > v_0. \tag{102}$$

Since this contradicts (101) it follows that the least upper bound of v values satisfying (101) is at most equal to the greatest lower bound of v values satisfying (100). Hence $a_{1n} \leqq a_{2n}$ and a nonempty closed interval $[a_{1n}, a_{2n}]$ is therefore defined for each $n > 0$.

(ii) For any v points such that $a_{1n} < v < a_{2n}$ neither (100) nor (101) holds and hence using (32)

$$\rho_1(\xi_v^n) < \min [\rho_a(\xi_v^n), \rho_b(\xi_v^n)] = \rho_0(\xi_v^n). \tag{103}$$

This proves (48).

(iii) For $v_0 > a_{2n}$ (if such values exist) by (100) and (79)

$$f_1^b(\xi_{v_0}^n) \leqq 0. \tag{104}$$

If strict inequality holds we apply (79) and if equality holds we apply (80), but the two results are the same, namely that

$$f_1^b(\xi_v^n) < 0 \qquad \text{for every } v > v_0. \tag{105}$$

Since v_0 can be taken arbitrarily close to the greatest lower bound a_{2n} then

$$f_1^b(\xi_v^n) < 0 \qquad \text{for every } v > a_{2n}. \tag{106}$$

(The reason for introducing v_0 is to avoid the proof of continuity of f_1^b in v. Although the continuity holds it is not needed for our purposes.) Similarly by (101) and Lemma 6 it follows that

$$f_1^a(\xi_v^n) < 0 \qquad \text{for } v < a_{1n}. \tag{107}$$

Hence by (106) and (107)

$$\rho_0(\xi_v^n) = \min [\rho_a(\xi_v^n), \rho_b(\xi_v^n)] < \rho_1(\xi_v^n) \qquad \text{for every } v \varepsilon J_n. \tag{108}$$

This proves (49).

(iv) By (100), (101) and (69)

$$f_a^b(\xi_v^n) < 0 \qquad \text{for } v > a_{2n}, \tag{109}$$

$$f_b^a(\xi_v^n) < 0 \qquad \text{for } v < a_{1n}. \tag{110}$$

This proves (50) and (51).

(v) By definition of a_{1n}, a_{2n} it follows that

$$\rho_b(\xi_v^n) > \min [\rho_a(\xi_v^n), \rho_1(\xi_v^n)] \qquad \text{for } v < a_{2n}, \tag{111}$$

$$\rho_a(\xi_v^n) > \min [\rho_b(\xi_v^n), \rho_1(\xi_v^n)] \qquad \text{for } v > a_{1n}. \tag{112}$$

Hence if $a_{1n} < a_{2n}$ then (111) holds for $v = a_{1n}$ and (112) holds for $v = a_{2n}$. This proves (52) and (53) and completes the proof of Theorem 3.

As mentioned above, it follows in the discrete case that for each $\xi \varepsilon \zeta$ the class $B_\xi \subset A$ and hence $B_\zeta \subset A$. Hence by the definition of closure $\overline{B_\zeta} \subset \bar{A}$. In the absolutely continuous case Theorem 3 shows that for $\xi \varepsilon \zeta$ there exists a Bayes solution $\delta = \delta_\xi$ which belongs to A and that any other Bayes solution $\delta' = \delta'_\xi$ relative to the same ξ can differ from δ at most on a set of n-dimensional Lebesgue measure zero for each positive integer n. Hence by (11)

$$\delta_{0nj}(S_n) = \delta'_{0nj}(S_n) \tag{113}$$

for each $n \geqq 0$, each j and every Lebesgue measurable set S_n in the space of $x_1, x_2, \cdots x_n$. Hence, given any sequence $\{\delta^i\}$ $(i = 1, 2, \cdots)$ of members of B_ζ which converges in the regular sense (see Section 3), we can for each i replace δ^i by a member of A without altering the convergence or the limit. Clearly then, since any member of $\overline{B_\zeta}$ is a limit of members of B_ζ it is also a limit of members of A and must therefore be in $\bar{A}$, that is, $\overline{B_\zeta} \subset \bar{A}$.

It now follows from Wald's Theorem 3.19 (see Section 4 above) that $\bar{A}$ is essentially complete relative to D_b in both the discrete and absolutely continuous cases. Then using the result of Section 6 the final result is obtained, namely that A is essentially complete relative to D_b.

Corollary 5. *If in the above problem the indifference zone is the empty set and $\hat{\theta}(\bar{x})$ takes on values between θ^* and $\theta^* + \epsilon$ for every $\epsilon > 0$ then the class A remains essentially complete relative to D_b.*

Proof. Consider the sequence of problems P^i $(i = 1, 2, \cdots)$ in which the parameter space Ω remains fixed and the indifference zone Z^i is the open interval

(θ^*, θ_i) where $\theta^* < \theta_{i+1} < \theta_i < \bar{\theta}$ for each i $(i = 1, 2, \cdots)$ and $\lim_{i=\infty} \theta_i = \theta^*$. Then the limiting problem P^0 has an empty indifference zone. In order that Assumption 3(ii) hold for each problem P^i it is sufficient that among the values assumed by the function $\hat{\theta}(\bar{x})$ there is a strictly decreasing sequence approaching θ^*. By the above the class A is then essentially complete relative to D_b for each problem P^i $(i = 1, 2, \cdots)$. Let $\bar{\delta}$ be any decision rule for P^0. Then $\bar{\delta}$ is also a decision rule for each problem P^i. For each i $(i = 1, 2, \cdots)$ there exists a decision rule $\delta_i \,\varepsilon\, A$ such that

$$r_i(\theta, \delta_i) \leqq r_i(\theta, \bar{\delta}) \qquad \text{for all } \theta \,\varepsilon\, \Omega \tag{114}$$

where $r_i(\theta, \delta)$ is the risk function when δ is the decision rule used and P^i is the problem. Since $r_i(\theta, \delta)$ differs from $r_0(\theta, \delta)$ only on the open interval (θ^*, θ_i) then by (114)

$$r_0(\theta, \delta_i) \leqq r_0(\theta, \bar{\delta}) \qquad \text{for } \theta \,\varepsilon\, \Omega - Z^i. \tag{115}$$

Wald has shown ([1], Theorem 3.2) that we can extract from the sequence $\{\delta_i\}$ of members of A a subsequence $\{\delta_{i_\alpha}\}$ which converges in the regular sense to a limit δ_0 and which is such that

$$\liminf_{\alpha=\infty} r_0(\theta, \delta_{i_\alpha}) \geqq r_0(\theta, \delta_0) \qquad \text{for all } \theta \,\varepsilon\, \Omega. \tag{116}$$

By Theorem 1 it can be assumed that $\delta_0 \,\varepsilon\, A$ since otherwise we can find an equivalent rule in A. Consider any point $\theta \,\varepsilon\, \Omega$. For sufficiently large i the point $\theta \,\varepsilon\, \Omega - Z^i$ so that (115) holds. Hence by (115) and (116)

$$r_0(\theta, \delta_0) \leqq r_0(\theta, \bar{\delta}). \tag{117}$$

Since θ is arbitrary (117) holds for all $\theta \,\varepsilon\, \Omega$. This proves the corollary.

REFERENCE

[1] A. Wald, *Statistical Decision Functions*, John Wiley and Sons, 1950.