The Annals of Statistics
1981, Vol. 9, No. 3, 678–682

PROPERTIES OF BAYES SEQUENTIAL TESTS

BY R. H. BERK[1], L. D. BROWN[2] AND ARTHUR COHEN[3]

Rutgers University, Cornell University, and Rutgers University

Consider the problem of sequentially testing composite, contiguous hypotheses where the risk function is a linear combination of the probability of error in the terminal decision and the expected sample size. Assume that the common boundary of the closures of the null and the alternative hypothesis is compact. Observations are independent and identically distributed. We study properties of Bayes tests. One property is the exponential boundedness of the stopping time. Another property is continuity of the risk functions. The continuity property is used to establish complete class theorems as opposed to the essentially complete class theorems in Brown, Cohen and Strawderman.

1. Introduction and Summary. Consider the problem of sequentially testing composite contiguous hypotheses where the risk function is a linear combination of the probability of error in the terminal decision and the expected sample size. Assume that the common boundary of the closures of the null and the alternative hypothesis is compact. The observations are independent and identically distributed. We study properties of Bayes tests. One property which is defined below is the exponential boundedness of the stopping time. This is a desirable property of a test and has been studied for many sequential tests by Wijsman (1979).

Another property to be studied is continuity of the risk functions. This property is of interest in its own right and in addition is used to establish complete classes of tests as opposed to essentially complete classes of tests as discussed in Brown, Cohen, and Strawderman (1980).

The first result obtained is that under mild conditions, for parameter points in the support of the prior distribution, there exists a neighborhood of the point, such that the stopping time for the Bayes test is uniformly exponentially bounded for all parameter points in the neighborhood. This fact can be used to prove that the risk functions are continuous in neighborhoods of points in the support of the priors. This result, in turn, gives the desired improvements in the complete class theorem.

Under additional assumptions it is possible to prove the above results on exponential boundedness and continuity for all points in the parameter space, not only for points in the support of the prior. As examples that satisfy all the assumptions, we can cite most of the one dimensional exponential family distributions in which one-sided or two-sided tests are considered. For the first set of results, (those without the additional assumptions) some multivariate, multiparameter distributions, and a wide variety of hypotheses satisfy the required assumptions.

Preliminaries are given in Section 2 while exponential boundedness and continuity are studied in Section 3.

2. Preliminaries. Let $X, X_1, X_2, \cdots$ be a sample sequence of independent and identically distributed random variables. Let $\mathbf{X} = (X_1, X_2, \cdots)$, $\mathbf{X}_j = (X_1, X_2, \cdots, X_j)$, and

Received June, 1979; revised May, 1980.

[1] Research supported by N.S.F. Grant No. MSC-76-82618-A02.
[2] Research supported by N.S.F. Grant No. MCS-78-24175.
[3] Research supported by N.S.F. Grant No. MCS-78-24167.

AMS 1970 *subject classifications*. Primary 62L10; secondary 62C10, 62L15.

Key words and phrases. Sequential tests, hypothesis testing, Bayes test, exponentially bounded stopping times, exponential family.

denote the realized sequences by $\mathbf{x} = (x_1, x_2, \cdots)$ and $\mathbf{x}_j = (x_1, x_2, \cdots, x_j)$. We assume that there is a σ-finite measure μ which dominates the family $\{P_\theta(\cdot), \theta \in \Theta\}$ of probability measures for $\mathbf{X}$ in the following sense: for each $j = 1, 2, \cdots$, over the σ-field generated by $\mathbf{X}_j$, the measure P_θ is dominated by μ. Write $f_{\theta j} = dP_\theta/d\mu$ relative to this σ-field. Note that if $p(x \mid \theta)$ is the density of x relative to $d\mu$, then $f_{\theta 1} = p(x \mid \theta)$ and $f_{\theta n}(\mathbf{x}_n) = \prod_{i=1}^n p(x_i \mid \theta)$. Assume that the support of the density does not depend on θ. (Some cases where the support depends on θ can be treated with separate special but similar type arguments.) The parameter $\theta \in \Theta$ and Θ is a measurable subset of R^n. The space of the null hypothesis is $\Theta_1 \subset \Theta$ and that of the alternative is $\Theta_2 \subset \Theta$. Assume $\Theta = \Theta_1 \cup \Theta_2$. The closures of Θ_1 and Θ_2 are denoted by $\bar{\Theta}_1$ and $\bar{\Theta}_2$ respectively. Let $\Omega = \bar{\Theta}_1 \cap \bar{\Theta}_2$ and $\Lambda = \Theta \cup \Omega$. Assume $\{P_\theta, \theta \in \Theta\}$ can be extended to a family $\{P_\theta, \theta \in \Lambda\}$ and that the families of densities $f_{\theta n}$, $\theta \in \Lambda$ exist and have the properties to be specified later. A prior probability measure on Θ is denoted by $\Gamma(\cdot)$; $\Gamma_n(\cdot)$ represents the sequence of posterior distributions, where for a measurable set $A \subset \Theta$,

$$\Gamma_n(A) = \int_A f_{\theta n}(\mathbf{x}_n)\Gamma(d\theta) / \int_\Theta f_{\theta n}(\mathbf{x}_n)\Gamma(d\theta). \tag{2.1}$$

Sometimes we express $\Gamma(\cdot) = \pi_1\Gamma_1(\cdot) + \pi_2\Gamma_2(\cdot)$. Here if T is a random variable with distribution Γ, then π_1 is the probability that $T \in \Theta_1$ and Γ_1 represents the conditional probability distribution of T, given $T \in \Theta_1$; Γ_2 has a similar definition.

Let $l_n(\theta) = n^{-1} \sum_1^n \ln[p(X_i \mid \theta)/p(X_i \mid \theta^*)]$ where θ^* is any fixed parameter point, $\lambda_\theta(t) = E_\theta \ln[p(X \mid t)/p(X \mid \theta^*)]$, and $l(x \mid \theta) = \ln[p(x \mid \theta)/p(x \mid \theta^*)]$. For any measurable, real valued function v, on Θ, and measureable set $A \subset \Theta$, let $\Gamma \sup_A v = \sup\{z : \Gamma\{t \in A : v(t) > z\} > 0\}$. When $A = \Theta$ the subscript is omitted. Let $\lambda_\theta^* = \Gamma \sup \lambda_\theta$.

The action space D consists of pairs (n, τ) where n is the value of the stopping time, N, and $\tau = 1$ or 2, depending on whether Θ_1 is accepted or rejected. The loss function, denoted by $L(\theta, (n, \tau)) = cn$ if $\theta \in \Theta_\tau$ and $L(\theta, (n, \tau)) = cn + 1$ if $\theta \notin \Theta_\tau$. Here c represents the cost of each individual observation. A decision function is denoted by δ, the risk function by $R(\theta, \delta) = E_\theta L(\theta, \delta)$ and the expected risk by $EL(\theta, \delta)$. $R_i(\theta, \delta)$ denotes the restriction of the risk when $\theta \in \Theta_i$. A Bayes procedure minimizes $EL(\theta, \delta)$ and for the loss function here a Bayes procedure exists. (See Le Cam (1955).) The Bayes test with respect to Γ will be denoted by δ_Γ and its stopping time by N_Γ.

The support of a probability measure $\Gamma(\cdot)$ is defined as, supp $\Gamma = \cap \{C : \Gamma(C) = 1, C$ closed$\}$. A sequence of events $\{B_n\}$ is said to be exponentially bounded if for some $a > 0$, $0 < \rho < 1$, $P_\theta(B_n) < a\rho^n$. The sequence is uniformly exponentially bounded for all θ lying in a set Q if a and ρ can be chosen independently of $\theta \in Q$. A random variable Y is exponentially bounded if the events $(|Y| > n)$ are.

Let

$$g_{in}(\mathbf{x}_n) = \int_{\Theta_i} f_{\theta n}(\mathbf{x}_n)\Gamma_i(d\theta) \Big/ \left[\pi_1 \int f_{\theta n}(\mathbf{x}_n)\Gamma_1(d\theta) + \pi_2 \int f_{\theta n}(\mathbf{x}_n)\Gamma_2(d\theta)\right], \tag{2.2}$$

so that $\pi_i g_{in}(\mathbf{x}_n)$ represents the conditional probability that $\theta \in \Theta_i$, given $\mathbf{X}_n = \mathbf{x}_n$. Finally let $\tau_n(\mathbf{x}_n) = \min\{\pi_i g_{in}(\mathbf{x}_n) : i = 1, 2\}$ and note that $\pi_1 g_{1n}(\mathbf{x}_n) + \pi_2 g_{2n}(\mathbf{x}_n) = 1$. Also note that if $\tau_n(\mathbf{x}_n) < c$ then the Bayes procedure stops for such $\mathbf{x}_n$ at stage n or before.

3. Exponentially bounded stopping times and continuity of the risk.

We start with

LEMMA 3.1. *Assume*: (a) *$p(X \mid \theta)$ is continuous in θ* w.p.1. (b) *For each $\theta \in$ supp Γ, there is a compact set $K \subset \Theta$ and a neighborhood $Q = Q(\theta, K)$ of θ so that for some $s >$* 0,

$$\sup_{q \in Q} E_q e^{s\Gamma \sup_{u \notin K} |l(X \mid u)|} < \infty \tag{3.1}$$

and

$$E_\theta \Gamma \sup_{u \notin K} l(X \mid u) < \lambda_\theta^*. \tag{3.2}$$

(c) *For each* $t \in K$, *there is a neighborhood* $U = U_t$ *of* t *so that:*
For all $\theta \in \operatorname{supp} \Gamma$, *there is a neighborhood* $Q = Q(\theta, U_t)$ *of* θ *so that for some* $s > 0$,

$$\sup_{q \in Q} E_q e^{s\Gamma \sup_{u \in U} |l(X \mid u)|} < \infty. \tag{3.3}$$

Then for each $\theta \in \operatorname{supp} \Gamma$ *and neighborhood* A *of* θ, *there is a neighborhood* $Q = Q(\theta, A)$ *of* θ *and a* $\beta > 0$ *so that* $(\Gamma_n A^c > e^{-n\beta})$ *is uniformly exponentially bounded on* Q.

The proof of this lemma utilizes the following:

LEMMA 3.2. *Let* h *be a convex function, finite on a neighborhood of zero, for which* $h(0) = h'(0) = 0$. *Let* $\{h_n\}$ *be a sequence of convex functions for which* $h_n(0) = 0$ *and*

$$\lim_{n \to \infty} h_n(x) = h(x),$$

for all x *in a neighborhood of zero. Then*

$$h_n^+(0) =_{\text{def}} \downarrow \lim_{r \downarrow 0} + h_n(r)/r$$

and

$$h_n^-(0) =_{\text{def}} \uparrow \lim_{r \uparrow 0} - h_n(r)/r$$

both tend to zero as $n \to \infty$.

PROOF. $h_n^+(0) \leq h_n(r)/r$ for $r > 0$, so $\limsup_n h_n^+(0) \leq h(r)/r$ and therefore $\limsup_n h_n^+(0) \leq h'(0) = 0$. Similarly $\liminf_n h_n^-(0) \geq 0$ and, of course, $h_n^-(0) \leq h_n^+(0)$. □

PROOF OF LEMMA 3.1. The argument is similar to the proof of Theorem 5.3 and Corollary 5.4 in Berk (1970a). Some modifications are needed to obtain *uniform* exponential boundedness. □

THEOREM 3.3. *Assume the conditions of Lemma* 3.1 *hold. Also assume that* sup $p(x \mid \theta) < \infty$ *for almost all* x. *Then for each* $\theta \in \operatorname{supp} \Gamma$, *there is a neighborhood* Q, *such that* N_Γ *is uniformly exponentially bounded on* Q.

PROOF. For $\theta \in \operatorname{supp} \Gamma$, let Q be the set whose existence is hypothesized in Lemma 3.1. First suppose that $Q \cap \bar{\Theta}_2 = \varnothing$ (null set), where Q is defined in Lemma 3.1. By shrinking Q if necessary, it follows that

$$P_q\{\pi_1 g_1(\mathbf{X}_n) > 1 - e^{-n\beta}\} > 1 - a\rho^n, \tag{3.4}$$

for all $q \in Q$. The equation (3.4) implies that for all $q \in Q$ and n sufficiently large

$$P_q\{\tau_n(\mathbf{X}_n) < c\} > 1 - a\rho^n. \tag{3.5}$$

The equation (3.5) in turn implies that the probability that the Bayes test stops after stage n is less than $a\rho^n$. Hence N_Γ is uniformly exponentially bounded on Q. The same reasoning applies when $Q \cap \bar{\Theta}_1 = \varnothing$. If $\theta \in \Omega$ but is only in the support of Γ_1 or Γ_2 then similar reasoning applies. The last case to consider is when $\theta \in \Omega$ and in the support of both Γ_1 and Γ_2. In this case we apply Lemma 3.1 to Γ_1 and Γ_2 separately to conclude that for ε suitably small

$$P_\theta\{\Gamma_{in}(Q \cap \Theta_i) > 1 - \varepsilon\} > 1 - a\rho^n, \qquad i = 1, 2 \tag{3.6}$$

for all $\theta \in Q$. Now (3.6) implies, by virtue of Lemma 2.2 of Berk, Brown, and Cohen (1980), that the probability of the Bayes test stopping at stage n, exceeds $1 - a\rho^n$. Hence N_Γ is uniformly exponentially bounded on Q. □

COROLLARY 3.4. *Under the conditions of Theorem* 3.3, *for every* $\theta \in \operatorname{supp} \Gamma$, $R_i(\delta_\Gamma, \cdot)$ *is continuous at* θ, $i = 1, 2$.

PROOF. Let m be a positive integer to be chosen later. The risk function for $\theta \in \Theta_1$ is

(3.7) $$R_1(\theta, \delta_\Gamma) = P_\theta \text{ (Reject null hypothesis)} + E_\theta N_\Gamma 1_{(N_\Gamma \le m)} + E_\theta N_\Gamma 1_{(N_\Gamma > m)}.$$

Let $f_{\theta N}(\mathbf{x}) = \sum_{n=1}^{\infty} f_{\theta n}(\mathbf{x}_n) 1_{(N=n)}(\mathbf{x}_n)$, be the density for P_θ relative to μ on the stopped σ-field generated by $N, X_1, X_2, \cdots, X_N$. The density $f_{\theta N}(\mathbf{x})$ is continuous in θ since $p(x|\theta)$ is continuous in θ and only one term in the infinite sum is nonzero for any fixed $\mathbf{x}$. This in turn implies that P_θ (Reject null hypothesis) is continuous in θ. The term $E_\theta N_\Gamma 1_{(N_\Gamma \le m)}$ is continuous in θ since it may be expressed as a finite sum of continuous functions. By virtue of Theorem 3.3 we may choose m sufficiently large so that $E_\theta N_\Gamma 1_{(N_\Gamma > m)}$ is arbitrarily small, uniformly for $\theta \in Q$. This implies the continuity of $R_1(\theta, \delta_\Gamma)$. The same proof applies to $R_2(\theta, \delta_\Gamma)$. □

Corollary 3.4 is of interest both in its own right and because of its applicability. Brown, Cohen, and Strawderman (1980) obtain essentially complete classes for several models of sequential hypothesis testing problems. In their study they find complete classes of tests of the closure of the null hypothesis against the closure of the alternative hypothesis. They point out that such complete classes provide essentially complete classes for testing the corresponding null against alternative hypotheses. We now prove that for problems satisfying the assumption of Corollary 3.4 and the modest assumptions of Theorem 3.2 of Brown, Cohen, and Strawderman (1980), which admit compact Θ_1, the essentially complete class for testing $\bar{\Theta}_1$ vs $\bar{\Theta}_2$ is in fact a complete class (not merely an essentially complete class) for testing $\bar{\Theta}_1$ vs $\bar{\Theta}_2$. (We regard the distinction between essentially complete and complete as important since a procedure outside a complete class is inadmissible, while a procedure outside an essentially complete class can be admissible.)

THEOREM 3.5. *Let* $\bar{\Theta}_1$ *be compact. Under the conditions of Corollary* 3.3 *and the conditions of Theorem* 3.2 *of Brown, Cohen, and Strawderman* (1980), *the complete class of tests of* $\bar{\Theta}_1$ vs $\bar{\Theta}_2$ *is also a complete class for testing* Θ_1 *vs* Θ_2.

PROOF. Let $\mathscr{B}$ denote the class of Bayes tests for testing $\bar{\Theta}_1$ vs $\bar{\Theta}_2$. Brown, Cohen, and Strawderman (1980) prove that $\mathscr{B}$ is a complete class for testing $\bar{\Theta}_1$ vs $\bar{\Theta}_2$. Suppose δ is an admissible test of Θ_1 against Θ_2 but $\delta \notin \mathscr{B}$. Since $\mathscr{B}$ is essentially complete for testing Θ_1 against Θ_2 there exists a prior, Γ, and Bayes test $\delta_\Gamma \in \mathscr{B}$ such that $R_i(\theta, \delta) = R_i(\theta, \delta_\Gamma)$ for all $\theta \in \Theta_i$, $i = 1, 2$. Furthermore the support of Γ must include points in $V = (\bar{\Theta}_1 - \Theta_1) U (\bar{\Theta}_2 - \Theta_2)$, otherwise δ would have to lie in $\mathscr{B}$. In fact the only points where $R(\theta, \delta)$ could possibly be different from $R(\theta, \delta_\Gamma)$ are in $W = \{V \cap \operatorname{supp} \Gamma\}$. Corollary 3.4 implies that $R_i(\theta, \delta_\Gamma)$ is uniformly continuous in neighborhoods of any such points. Also $R_i(\theta, \delta)$ is lower semicontinuous at such points since it may be expressed as a limit of an increasing sequence of continuous functions. Note then that if $R_i(\theta, \delta) \neq R_i(\theta, \delta_\Gamma)$ at points in W, $R_i(\theta, \delta) < R_i(\theta, \delta_\Gamma)$. This contradicts the fact that $\delta_\Gamma \in \mathscr{B}$ but $\delta \notin \mathscr{B}$. □

REMARK 3.6 A result comparable to Theorem 3.5 is true for the other two models treated in Sections 4 and 5 of Brown, Cohen, and Strawderman (1980). One needs the additional assumptions required in Berk's (1970a) Theorem 5.3 to allow for the cases where Γ is not proper.

REMARK 3.7 Under additional assumptions, Theorem 3.3 and Corollary 3.4 can be proved not only for $\theta \in \operatorname{supp} \Gamma$ but for all $\theta \in \Theta$. The additional assumptions required are those stated as Assumptions 3.2 of Berk (1970b), with the additional proviso that in Assumption 3.2 (e), the A_δ form a weak base at $\theta_p \in \Theta$ or the A_δ are contained entirely in either Θ_1 or Θ_2. The essentially step is proving these stronger versions of Theorem 3.3 and

Corollary 3.4 is to recognize that for parameter points not in the support of the prior, the stopping time of the Bayes procedure is bounded above by the stopping time of a weight function procedure using Γ_1 and Γ_2 as weights. (See Berk (1970b) for the definition of a weight function procedure.) Minor modifications of the proof of Berk's (1970b) Theorem 3.4 give the required result for weight function procedures and hence for these Bayes tests.

REFERENCES

BERK, R. H. (1970a). Consistency a posteriori. *Ann. Math. Statist.* **41** 894–906.

BERK, R. H. (1970b). Stopping time of SPRTS based on exchangeable models. *Ann. Math. Statist.* **41** 979–990.

BERK, R. H., BROWN, L. D., COHEN, A. (1980). Bounded stopping times for a class of sequential Bayes tests. To appear in *Ann. Statist.* **9**.

BROWN, L. D., COHEN, A., STRAWDERMAN, W. E. (1980). Complete classes for sequential tests of hypotheses. *Ann. Statist.* **8** 377–398.

LECAM, L. (1955). An extension of Wald's theory of statistical decision functions. *Ann. Math. Statist.* **26** 69–81.

WIJSMAN, R. A. (1979). Stopping time of invariant sequential probability ratio tests. In *Developments in Statistics,* Vol. 2. Academic, New York.

R. H. BERK
ARTHUR COHEN
DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
HILL CENTER FOR THE MATHEMATICAL SCIENCES
BUSCH CAMPUS
NEW BRUNSWICK, NEW JERSEY 08903

L. D. BROWN,
DEPARTMENT OF MATHEMATICS
CORNELL UNIVERSITY
ITHACA, NEW YORK 14850