The Annals of Statistics
1989, Vol. 17, No. 1, 107–124

UNIFORMLY POWERFUL GOODNESS OF FIT TESTS[1]

By Andrew R. Barron

University of Illinois at Urbana-Champaign

The simple hypothesis is tested that the distribution of independent random variables $X_1, X_2, \dots, X_n$ is a given probability measure P_0. Let π_n be any sequence of partitions. The alternative hypothesis is the set of probability measures P with $\sum_{A \in \pi_n} |P(A) - P_0(A)| \geq \delta$, where $\delta > 0$. Note the dependence of this set of alternatives on the sample size. It is shown that if the effective cardinality of the partitions is of the same order as the sample size, then sequences of tests exist with uniformly exponentially small probabilities of error. Conversely, if the effective cardinality is of larger order than the sample size, then no such sequence of tests exists. The effective cardinality is the number of sets in the partition which exhaust all but a negligible portion of the probability under the null hypothesis.

1. Introduction. Let $X_1, X_2, \dots$ be independent and identically distributed random variables taking values in a measurable space $(\mathbf{X}, \mathbf{B})$ and let M be the set of all probability measures on the space. An arbitrary probability measure P_0 is entertained as the simple null hypothesis for a sequence of tests based on the data $X_1, X_2, \dots, X_n$. Tests are desired which have small probabilities of error for as large as possible a set of alternative probability measures. A test is *uniformly consistent* if the probabilities of error uniformly converge to 0 as $n \to \infty$; it is said to be *uniformly exponentially consistent* (UEC) if the probabilities of error are uniformly less than e^{-nr} for all large n for some $r > 0$. Exponential bounds on the probability of error are known in the simple versus simple case [Chernoff (1952, 1956) and Csiszár and Longo (1971)], in the discrete case [Hoeffding (1965)], in the case of smooth parametric families [Hoeffding and Wolfowitz (1958), Bahadur (1966) and Brown (1971)] and in the case of convex sets of alternatives defined by "capacities" [Huber and Strassen (1973)].

We consider composite alternative hypotheses of the form

$$\{P \in M: d(P, P_0) \geq \delta\}$$

for various distance functions $d(P, P_0)$ and $\delta > 0$. These sets are generally not convex and nonparametric. Nevertheless, the problem is to determine whether UEC tests exist against $\{P: d(P, P_0) \geq \delta\}$. Of course this problem is motivated by the need to understand the nonlocal behavior of goodness of fit tests. Another motivation comes from the problem of consistency of Bayes procedures. If the prior satisfies a certain natural assumption, then for any distance d for which a UEC test exists, the sequence of posterior distributions asymptotically

Received February 1987; revised March 1988.

[1]Work supported by the Office of Naval Research under Contract N00014-86-K-0670 and by a National Science Foundation Postdoctoral Research Fellowship.

AMS 1980 *subject classifications*. Primary 62G10; secondary 60F10.

Key words and phrases. Uniformly consistent tests, goodness of fit, large deviation inequalities, exponential bounds, Kullback–Leibler divergence, variation distance.

concentrates on $\{P: d(P, P_0) < \delta\}$ with P_0-probability 1 [see Schwartz (1965) and Barron (1986, 1987)]. In the author's dissertation [Barron (1985)], the convergence of minimum complexity estimates of probability measures is established for any distance d for which such tests exist.

For some weak distances, UEC tests are known to exist. Hoeffding and Wolfowitz (1958) observed that if an inequality of the form

$$P^n\{d(\hat{P}_n, P) > \delta/2\} \leq e^{-nr}$$

holds uniformly in P, for some $r > 0$, where $\hat{P}_n$ is the empirical distribution and d satisfies the triangle inequality, then the critical set $\{d(\hat{P}_n, P_0) > \Delta/2\}$ provides a UEC test against $\{P: d(P, P_0) > \delta\}$. Examples of distances which satisfy the Hoeffding–Wolfowitz condition are the Kolmogorov–Smirnov distance, the distances of Vapnik and Chervonenkis (1971) and the variation distances $d_\pi(P, Q) = \sum_{A \in \pi} |P(A) - Q(A)|$ for finite partitions π.

Ideally, we would like the distance function $d(P, P_0)$ to be as sensitive as possible, corresponding to the strongest possible mode of convergence. Unfortunately, for any distance function which dominates the total variation distance ($d(P, P_0) = \sup_\pi d_\pi(P, P_0)$), if the probability measure P_0 is not discrete, then *no* uniformly consistent test exists against $\{P: d(P, P_0) \geq \delta\}$ for small $\delta > 0$ (see Section 5). This negative result has implications for many popular measures of distance or divergence between probability measures including the chi-square, Kullback–Leibler, Hellinger and total variation distance. Csiszár (1967) examined a broad class of measures of divergence (which are characterized by certain natural axioms [Csiszár (1974)]). He discovered that these divergences all dominate the total variation distance. It follows that no uniformly consistent test exists when the hypotheses are defined in terms of any of these distances.

In the narrow gap between the total variation distance $d(P, P_0) = \sup_\pi d_\pi(P, P_0)$ and the variation $d_\pi(P, P_0) = \sum_{A \in \pi} |P(A) - P_0(A)|$ for fixed but arbitrarily fine partitions π, we want distances for which UEC tests do exist. One approach is to consider sequences of distances d_{π_n} corresponding to a sequence of partitions π_n. We let the set of alternatives be $\{P: d_{\pi_n}(P, P_0) \geq \delta\}$, where n is the sample size. If a refining sequence of partitions is chosen which generates the measurable space, then the distances $d_{\pi_n}(P, P_0)$ *increase to* $d(P, P_0)$. By allowing increasing sets of alternatives, we hope to reflect the increasing distinguishability of distributions as sample size increases.

The following question is posed and answered in this paper. For what sequences of partitions π_n do there exist uniformly exponentially consistent tests for P_0 versus $\{P: d_{\pi_n}(P, P_0) \geq \delta\}$ for all $\delta > 0$? The answer is that the $(1 - \delta)$ effective cardinality of π_n must be of the same order as n or smaller for all $\delta > 0$. The $1 - \delta$ effective cardinality is the smallest number of sets in a partition which exhaust at least $1 - \delta$ of the probability under the null hypothesis.

For the case that the effective cardinality is of smaller order than n, a familiar likelihood ratio test is shown to be uniformly exponentially consistent (see Section 3). As a byproduct of our analysis, the optimal exponent properties of the likelihood ratio test established by Tusnády (1977) for partitions of cardinality $o(n/\log n)$ are seen to also be valid for partitions of cardinality $o(n)$. Some properties of the likelihood ratio and chi-square test are known even for partitions of cardinality $O(n)$ [Quine and Robinson (1985) and Holst (1972) determine the Pitman and Bahadur relative efficiencies]; however, it is not known whether either of these popular tests are uniformly exponentially consistent in the $O(n)$ case.

It is surprising that UEC tests do exist when the effective cardinality is $O(n)$. In this case many of the cells will be empty. Without accumulation of observations, the empirical probabilities are inaccurate estimates of the cell probabilities, so the usual test statistics are not expected to have small probabilities of error. In Section 4 we find a test statistic which has the desired property in this case. It is a weighted count of the number of empty cells.

In Section 5 it is shown that UEC tests are not possible if the partition has effective cardinality which is of larger order than the sample size. The proof uses a Bayesian technique involving a prior on the set of alternatives.

It may be desirable to restrict the hypotheses to probability measures which have a density function (with respect to a fixed dominating measure λ). Recall that the total variation distance between such probabilities may be expressed as the L^1 distance between the density functions, i.e., $d(P, P_0) = \int |p(x) - p_0(x)|\lambda(dx)$. It is seen that uniformly consistent tests do *not* exist against all densities for which the L^1 distance from P_0 is at least δ. In Section 6 we restrict attention to smooth alternatives (e.g., uniformly equicontinuous density functions) and conditions are found such that UEC tests exist. Not surprisingly, less smoothness is required as the sample size increases.

There are several useful methods for evaluating tests of a simple hypothesis which assess the performance separately for each alternative probability or for certain sequences of alternative probabilities (such as Pitman efficiency, Bahadur efficiency, Chernoff efficiency and Tusnády's exponential rate optimality). In contrast a method which is suggested by our formulation is to determine the ratio of the uniform error exponent of a test to the exponent for a minimax test for a large composite alternative hypothesis. However, it is an open problem to determine the exponent of a minimax test for most of the classes of alternatives that we consider. In this paper we only determine whether this exponent is positive.

2. Preliminaries. In this section we give some notation and recall a basic inequality.

Let $(\mathbf{X}^n, \mathbf{B}^n)$ denote the product space of n copies of $(\mathbf{X}, \mathbf{B})$ and let P^n denote the product measure corresponding to any P in M. A nonrandomized test of a sequence of hypotheses $H_{0,n}, H_{1,n} \in M$ is specified by a sequence of $\mathbf{B}^n$ measurable sets $A_{0,n}$. We call $A_{0,n}$ the acceptance set (for $H_{0,n}$) and $A_{1,n} = A_{0,n}^c$ the critical set. If the random sample $X^n = (X_1, X_2, \dots, X_n)$ is in $A_{0,n}$ we accept $H_{0,n}$ and reject otherwise.

A test is uniformly consistent if $\sup\{P^n(A_{j,n}^c)\colon P \in H_{j,n}\} \to 0$ as $n \to \infty$ for $j = 0, 1$. It is uniformly exponentially consistent (UEC) if there exist $r_0 > 0$,

$r_1 > 0$ and n_0 such that for all $n \geq n_0$,

$$\sup_{P \in H_{0,n}} P^n(A_{0,n}^c) \leq e^{-nr_0} \tag{2.1}$$

and

$$\sup_{P \in H_{1,n}} P^n(A_{1,n}^c) \leq e^{-nr_1}. \tag{2.2}$$

For a given test and sequence of hypotheses, the type I and type II asymptotic error exponents are defined by $r_j = -\limsup n^{-1} \ln(\sup_{H_{j,n}} P^n(A_{j,n}^c))$ for $j = 0, 1$, respectively.

More generally, for randomized tests, uniformly (exponentially) consistent tests are defined in the same way with expected values of critical functions replacing probabilities of critical sets. To determine whether consistent tests of hypotheses exist it suffices to restrict attention to nonrandomized test. Indeed, if $0 \leq \phi_n(X^n) \leq 1$ is any critical function, then applying Markov's inequality, $A_{1,n} = \{X^n: \phi_n(X^n) > \frac{1}{2}\}$ is seen to be a critical set for which the probabilities of error $P_0^n(A_{1,n})$ and $P_1^n(A_{1,n}^c)$ are no more than twice $\int \phi_n \, dP_0^n$ and $\int (1 - \phi_n) \, dP_1^n$, respectively.

Schwartz [(1965), Lemma 6.1] and Le Cam [(1973, Lemma 4] have shown that in the presently assumed case of independent random variables, if the two hypotheses H_0 and H_1 are fixed (not allowed to change with n) then the existence of a uniformly consistent test is equivalent to the existence of a uniformly exponentially consistent test. However this equivalence does not persist if the sets $H_{j,n}$ are allowed to grow with n. It is the stronger notion of uniform exponential consistency that we require in this case.

The fundamental work of Le Cam and Schwartz (1960) is acknowledged. They give necessary and sufficient conditions for the existence of uniformly consistent tests for fixed hypotheses in terms of topological properties of a technical sort [see also Le Cam (1973), Lemma 2]. We do not explicitly utilize these topological results here.

Next we mention an inequality from Csiszár (1984). Let $\hat{P}_n(\cdot) = \hat{P}(\cdot\,; X^n)$ denote the empirical probability measure defined by $\hat{P}(A; X^n) = (1/n)\sum_{i=1}^n 1_{\{X_i \in A\}}$ for $A \in \mathbf{B}$ and $X^n \in \mathbf{X}^n$. Let C be any completely convex set of probability measures. Then for all probability measures P and all n,

$$P^n\{\hat{P}_n \in C\} \leq e^{-nD(C\|P)}, \tag{2.3}$$

where $D(C\|P) = \inf\{D(Q\|P): Q \in C\}$ denotes the Kullback–Leibler divergence from C to P. Here $D(Q\|P) = E_Q \ln(q(X)/p(X))$, where $p(x)$ and $q(x)$ are the density functions of P and Q, respectively, with respect to some dominating measure, say $\lambda = P + Q$. Simple lower bounds on probabilities [e.g., as in Bahadur, Gupta and Zabell (1980), (2.12) or Blahut (1987), page 114] show that in many but not all cases the bound in (2.3) is tight (but we will not need these lower bounds here). When $C = \{Q \in M: E_Q f(X) \geq t\}$ for some real valued function f, inequality (2.3) reduces to a well known inequality from Chernoff (1952). In the discrete case with $C = \{Q\}$ a singleton set, (2.3) reduces to a basic inequality from Hoeffding [(1965), equation 2.4)].

A technique for finding uniformly consistent tests is sugested by application of inequality (2.3). Choose the completely convex set C to be a neighborhood of P_0 for which $\lim_n P_0\{\hat{P}_n \in C\} = 1$. Then the sequence of acceptance sets $\{\hat{P}_n \in C\}$ defines a uniformly consistent test for $H_0 = \{P_0\}$ versus $H_1 = \{P: D(C\|P) \geq r_1\}$ for any $r_1 > 0$. Observe that the test will automatically satisfy the uniform exponential bound as in (2.2).

3. Uniformly powerful tests—the $o(n)$ case. In this section we show that the likelihood ratio test has the desired uniform exponential bounds when the cardinality of the partitions is $o(n)$.

Let π_n be a sequence of partitions of the measurable space $(\mathbf{X}, \mathbf{B})$. Let $X_1, X_2, \ldots, X_n$ be independent random variables with probability measure P. Hoeffding (1965) showed that the likelihood ratio tests for the hypothesis testing problem $P = P_0$ versus $P \neq P_0$ based on the cell counts $\{n\hat{P}_n(A): A \in \pi_n\}$ have acceptance sets

$$A_{0,n} = \left\{D_{\pi_n}\left(\hat{P}_n\|P_0\right) \leq r_0\right\} = \left\{\hat{P}_n \in C^n\right\} \tag{3.1}$$

for some $r_0 > 0$. Here $C^n = C^n_{r_0} = \{Q: D_{\pi_n}(Q\|P_0) \leq r_0\}$, where

$$D_{\pi_n}(Q\|P_0) = \sum_{A \in \pi_n} Q(A) \ln \frac{Q(A)}{P_0(A)}.$$

To examine the set $A_{0,n}$ we may restrict attention to the simplex of probabilities $\mathbf{M}_n = \{(Q(A), A \in \pi_n): Q(A) \geq 0, \Sigma_A Q(A) = 1\}$. The set C^n is a completely convex subset of the simplex, so by Csiszár's inequality (2.3),

$$P^n(A_{0,n}) \leq e^{-nD_{\pi_n}(C^n\|P)} \tag{3.2}$$

for all P and all n, where $D_{\pi_n}(C^n\|P) = \min_{Q \in C^n} D_{\pi_n}(Q\|P)$. This exponent will be strictly positive for any P with $D_{\pi_n}(P\|P_0) > r_0$. Moreover, for any given $r_1 > 0$,

$$P^n(A_{0,n}) \leq e^{-nr_1} \tag{3.3}$$

for all P with $D_{\pi_n}(C^n\|P) \geq r_1$. Thus the set of alternatives for which the probability of error is uniformly exponentially small may be as large as $\{P: D_{\pi_n}(C^n\|P) \geq r_1\}$.

Now let

$$d_{\pi_n}(P_0, P) = \sum_{A \in \pi_n} |P_0(A) - P(A)|.$$

Let $m_n = \|\pi_n\|$ be the number of cells in the partition π_n.

LEMMA 1. *Given any $\delta > 0$, choose $0 < r_0 < \delta^2/2$ and let $r_1 = (\delta - \sqrt{2r_0})^2/2$. We test the hypothesis $H_0 = \{P_0\}$. The likelihood ratio test with acceptance region*

$$A_{0,n} = \left\{D_{\pi_n}\left(\hat{P}_n\|P_0\right) \leq r_0\right\}$$

has uniformly small probability of error $P^n(A_{0,n}) \leq e^{-nr_1}$ for all alternatives in *the set*

$$H_{1,n} = \{P: d_{\pi_n}(P, P_0) \geq \delta\}$$

or, more generally, for all alternatives in the set $\{P: D_{\pi_n}(C_{r_0}^n \| P) \geq r_1\}$.

If the sequence of partitions satisfies

$$\lim_{n\to\infty} \frac{m_n}{n} = 0,$$

then the test has type I *asymptotic error exponent no smaller than* r_0. *Hence the test is uniformly exponentially consistent for* H_0 *versus* $H_{1,n}$.

PROOF. The Kullback–Leibler number is related to the variation distance by an inequality, $D \geq (1/2)d^2$ or equivalently $d \leq \sqrt{2D}$, due to Csiszár (1967) and Kullback (1967). Thus for any Q in $C_{r_0}^n$ we have $d_{\pi_n}(Q, P_0) \leq \sqrt{2D_{\pi_n}(Q\|P_0)} \leq \sqrt{2r_0}$. Now for any P in $H_{1,n}$, we use the triangle inequality for d to obtain

$$\begin{aligned} D_{\pi_n}(Q\|P) &\geq \tfrac{1}{2}\big(d_{\pi_n}(Q, P)\big)^2 \\ &\geq \tfrac{1}{2}\big(d_{\pi_n}(P, P_0) - d_n(Q, P_0)\big)^2 \\ &\geq \tfrac{1}{2}\big(\delta - \sqrt{2r_0}\big)^2 = r_1 \end{aligned}$$

for all Q in $C_{r_0}^n$, whence $D_{\pi_n}(C_{r_0}^n\|P) \geq r_1$. For any such P we have $P^n(A_{0,n}) \leq e^{-nr_1}$ by (3.2) or (3.3). So the type II error probability is uniformly exponentially small.

Now we examine the type I error probability. Let $P_n^{\pi_n} = (P_n(A): A \in \pi_n)$ denote the empirical probability measure restricted to the partition. The number of such probability measures Q for which each $nQ(A)$ is an integer ($A \in \pi_n$) is equal to $\binom{n+m-1}{m-1}$, where $m = m_n$. The probability that any measure Q occurs is bounded by

$$P_0^n\{\hat{P}_n^{\pi_n} = Q\} \leq e^{-nD_{\pi_n}(Q\|P_0)},$$

which is less than e^{-nr_0} whenever $D_{\pi_n}(Q\|P_0) > r_0$. Consequently, the probability of type I error satisfies

$$P_0^n\{D_{\pi_n}(\hat{P}_n\|P_0) > r_0\} \leq \binom{n+m-1}{m-1} e^{-nr_0}. \tag{3.4}$$

This inequality is due to Hoeffding [(1965), (2.8)].

To bound the combinatorial coefficient we use the inequality

$$\binom{N}{k} \leq e^{NH(k/N)}$$

which holds for positive integers $N \geq k$. [This inequality follows from the fact that the probability that a binomial $(N, k/N)$ random variable equals k does not exceed 1.] Here H is the entropy function $H(p) =$

$-p \ln p - (1 - p)\ln(1 - p)$ for $0 \le p \le 1$. We note that $H(p) \to 0$ as $p \to 0$. It follows that

$$\binom{n+m}{m} \le e^{(n+m)H(m/(n+m))} = e^{n\varepsilon_n}, \tag{3.5}$$

where $\varepsilon_n = (1 + m/n)H(m/(n + m)) \to 0$, provided $m/n \to 0$.

Consequently, the probability of type I error is bounded by $e^{-n(r_0 - \varepsilon_n)}$ and hence the type I asymptotic error exponent is at least r_0. This completes the proof of Lemma 1. □

REMARK 1. There are other natural tests which are uniformly exponentially consistent against all P with $d_{\pi_n}(P, P_0) > \delta$. For instance, the test with acceptance region $\{d_{\pi_n}(\hat{P}_n, P_0) \le \sqrt{2r_o}\}$ has the same uniform bounds $e^{-n(r_0 - \varepsilon_n)}$ and e^{-nr_1} for the type I and II errors, respectively, where $r_1 = (\delta - \sqrt{2r_0})^2/2$. However, the likelihood ratio (Kullback–Leibler) test has the advantage that the uniform bounds hold for the larger class of alternatives $\{P\colon D_{\pi_n}(C^n \| P) \ge r_1\}$.

REMARK 2. A more striking advantage of the likelihood ratio test is established in Tusnády (1977). Suppose π_n is an increasing sequence of partitions which generates $\mathbf{B}$ and suppose the cardinality of the partitions is of order $o(n/\log n)$. Tusnády shows that if the type I error exponent is fixed to be r_0, then for each alternative P the type II asymptotic error exponent is equal to $D(C_{r_0} \| P)$ [provided r_0 is a continuity point of $D(C_r \| P)$]. Moreover, Tusnády observed the startling fact that this type II asymptotic exponent for the likelihood ratio test (which does not depend on the alternative P) is the same exponent as is achieved by the best simple versus simple test (i.e., the Neyman–Pearson test which does use the alternative P).

We note that Tusnády's assumption on the cardinality of the partitions may be relaxed to the assumption that $m_n = o(n)$. This follows by applying the bound on $\binom{n+m}{m}$ from inequality (3.5) in place of the weaker bound $(n + 1)^m$ in his inequality (2.11).

REMARK 3. Kallenberg (1985) also has obtained relevant inequalities for the multinomial distribution with $m_n = o(n)$.

DEFINITION. The $1 - \varepsilon$ effective cardinality of a partition π with respect to a probability measure P_0 is the minimum number $m = m(\pi, P_0, \varepsilon)$ of sets $A_1, A_2, \dots, A_m$ in π that have total probability $\sum_{i=1}^m P_0(A_i) > 1 - \varepsilon$.

Note that it is natural to order the sets A_i in π such that $P_0(A_1) \ge P_0(A_2) \ge \cdots$. Then a minimal collection of sets to achieve probability greater than $1 - \varepsilon$ will be the first m sets $A_1, A_2, \dots, A_m$ for some m.

The effective cardinality of a *sequence* of partitions π_n is said to be of smaller order than a sequence a_n [written $o(a_n)$] if

$$\lim_{n\to\infty} \frac{m(\pi_n, P_0, \varepsilon)}{a_n} = 0 \quad \text{for all } \varepsilon > 0.$$

The effective cardinality is said to be of order a_n [written $O(a_n)$] if

$$\limsup_{n\to\infty} \frac{m(\pi_n, P_0, \varepsilon)}{a_n} < \infty \quad \text{for all } \varepsilon > 0.$$

Note that if for each $\varepsilon > 0$, there is a set E with $P_0(E) \le \varepsilon$ such that the number of cells which are not in E is of order $o(a_n)$ or $O(a_n)$, then the effective cardinality is of order $o(a_n)$ or $O(a_n)$, respectively.

The following examples illustrate some of the possibilities.

EXAMPLE 1. Let $\pi_n = \pi$ be any fixed countable partition. Then for any probability measure the effective cardinality of the partition is $O(1)$.

EXAMPLE 2. Suppose the probability measure P_0 is discrete [i.e., the probability is concentrated on countably many atoms of $(\mathbf{X}, \mathbf{B})$]. Then any sequence of partitions is of order $O(1)$.

EXAMPLE 3. Let $\mathbf{X}$ be the real line with the Borel sets and let

$$\pi_n = \{[ih, (i+1)h) : i = \ldots, -1, 0, 1, \ldots\}$$

be a uniform partition into intervals of width $h = h_n$. Then for any distribution with an absolutely continuous component, the effective cardinality of π_n is $o(n)$ if and only if $\lim nh_n = \infty$; whereas, the effective cardinality is $O(n)$ if and only if $\liminf nh_n > 0$.

EXAMPLE 4. Let P_0 be a continuous distribution on the real line and let π_n be a partition into n cells determined by uniform quantiles of P_0. Each cell has probability $1/n$ and the effective cardinality of π_n is $O(n)$.

When the distribution P_0 has a continuous component, sequences of partitions of any order are readily constructed.

The following is a useful extension of Lemma 1.

LEMMA 2. *If the effective cardinality of a sequence of partitions is $o(n)$, then for any $\delta > 0$ there exists a UEC test against $H_{1,n} = \{P: d_{\pi_n}(P_0, P) \ge \delta\}$.*

PROOF. Given $\delta > 0$, choose $0 < \varepsilon < \delta/2$. Fix n and let $\pi_n^\varepsilon = \{A_1, A_2, \ldots, A_m, E\}$ be a partition with each A_i in π_n, $P_0(E) \le \varepsilon$ and $m = m(\pi_n, P_0, \varepsilon)$. Let $t = \delta - 2\varepsilon$ and choose $0 < r_0 < t^2/2$. By Lemma 1, the likelihood ratio test with critical set

$$\{D_n^\varepsilon(\hat{P}_n \| P_0) > r_0\}$$

is UEC against $K_n = \{P\colon d_{\pi_n^\varepsilon}(P_0, P) \geq t\}$. Now the variation distance satisfies

$$d_{\pi_n}(P_0, P) = \sum_{A \in \pi_n} |P_0(A) - P(A)| = 2 \sum_{A \in \pi_n} (P_0(A) - P(A))^+.$$

Thus for any P in $H_{1,n}$ we have

$$\begin{aligned}\delta \leq d_{\pi_n}(P_0, P) &\leq 2 \sum_{i=1}^{m_n} (P_0(A_i) - P(A_i))^+ + 2P_0(E_n)\\ &\leq d_{\pi_n^\varepsilon}(P_0, P) + 2\varepsilon\end{aligned}$$

and hence $d_{\pi_n^\varepsilon}(P_0, P) \leq t$. Thus $H_{1,n}$ is a subset of K_n. Therefore, the test is also UEC against $H_{1,n}$. □

4. Uniformly powerful tests—the $O(n)$ case. This section makes use of a new test statistic to handle the case that the effective cardinality is of order n. The following lemma analyzes the behavior of this test. Some restrictions are made to enable us to investigate the essential features of this statistic. Subsequently, a combination of the new test and the likelihood ratio test will be used to prove a more general result.

LEMMA 3. *For each n, let $G_n = \{A_1, \ldots, A_m\}$ be a collection of disjoint measurable subsets of $\mathbf{X}$. Assume that $m \leq cn$ and $P_0(A_k) \leq b/n$ for some constants b and c. Define the test statistic*

$$T_n = \sum_{A \in G_n} \left(e^{nP_0(A)} 1_{\{\hat{P}_n(A) = 0\}} - 1\right).$$

If $0 < t < t_1$, then the test with critical set $\{T_n \geq nt\}$ is UEC against the set of all P with

$$\sum_{A \in G_n} \left(e^{n(P_0(A) - P(A))} - 1\right) \geq nt_1.$$

If also $\pi_n = G_n$ is a partition which exhausts all of $\mathbf{X}$, then this set of alternatives includes $\{P\colon d_{\pi_n}(P, P_0) \geq \sqrt{8ct_1}\}$.

PROOF. We use a Poissonization argument. Fix n and let $A_0 = \mathbf{X} - (A_1 \cup \cdots \cup A_m)$. The (multinomial) distribution of the cell counts $n\hat{P}_n(A_0), \ldots, n\hat{P}_n(A_m)$ is the same as the conditional distribution of $Y_0, \ldots, Y_m$ given that $\sum_{k=0}^m Y_k = n$, where $Y_0, \ldots, Y_m$ are independent Poisson random variables with parameters $\lambda_0, \ldots, \lambda_m$. If the hypothesis P obtains, then the parameter values are $\lambda_k = nP(A_k)$, $k = 0, \ldots, m$ and we denote the joint distribution for $Y_0, \ldots, Y_m$ by $\bar{P}^n$. Define the random variable

$$\bar{T}_n = \sum_{k=1}^m \left(e^{nP_0(A_k)} 1_{\{Y_k = 0\}} - 1\right).$$

Then the probability of type I error satisfies

$$P_0^n\{T_n \geq nt\} = \bar{P}_0^n\left\{\bar{T}_n \geq nt \,\middle|\, \sum_{k=0}^{m} Y_k = n\right\} \leq (2\pi n)^{1/2}\bar{P}_0^n\{\bar{T}_n \geq nt\},$$

where we have used the fact that the Poisson event $\{\sum_{k=0}^m Y_k = n\}$ has probability greater than $(2\pi n)^{-1/2}$. Similarly the probabilities of type II error satisfy

$$P^n\{T_n < nt\} \leq (2\pi n)^{1/2}\bar{P}^n\{\bar{T}_n < nt\}.$$

Consequently, it suffices to provide exponential bounds for the probabilities $\bar{P}_0^n\{\bar{T}_n \geq nt\}$ and $\bar{P}^n\{\bar{T}_n < nt\}$.

The expected value of $\bar{T}_n$ under the null hypothesis (P_0) is

$$E(\bar{T}_n) = E\sum_k\left(e^{nP_0(A_k)}1_{\{Y_k=0\}} - 1\right) = \sum_k\left(e^{nP_0(A_k)}e^{-nP_0(A_k)} - 1\right) = 0.$$

Applying Hoeffding's inequality for sums of bounded independent random variables [Hoeffding (1963), Theorem 2] yields

$$\bar{P}_0^n\{\bar{T}_n \geq nt\} \leq \exp\left\{\frac{-2(nt)^2}{\sum_k e^{2nP_0(A_k)}}\right\} \leq \exp\left\{-\frac{2nt^2}{(ce^{2b})}\right\}.$$

Hence the probability of type I error $P_0^n\{T_n \geq nt\}$ is exponentially small.

Now suppose that P is an alternative probability measure. Then the expected value of $\bar{T}_n$ is given by

$$E(\bar{T}_n) = \sum_{A\in G_n}\left(e^{n(P_0(A)-P(A))} - 1\right) \geq nt_1.$$

Applying Hoeffding's inequality yields

$$\begin{aligned}\bar{P}^n\{\bar{T}_n < nt\} &\leq \bar{P}^n\{-T_n + E(\bar{T}_n) > n(t_1 - t)\}\\ &\leq \exp\{-2n(t_1 - t)^2/(ce^{2b})\}\end{aligned}$$

uniformly over all P for which $E(\bar{T}_n) \geq nt_1$. Thus the type II error probability is uniformly exponentially small.

Next we derive a lower bound on $E(\bar{T}_n)$ which will be needed later. Use the inequality $e^x - 1 \geq x + (\frac{1}{2})(x^+)^2$, where $x^+ = \max\{0, x\}$, followed by the Cauchy–Schwarz inequality to obtain, with U_n denoting the union of the sets in G_n, that

$$\begin{aligned}E(\bar{T}_n) &\geq n\sum_{A\in G_n}(P_0(A) - P(A)) + \frac{n^2}{2}\sum_{A\in G_n}\left((P_0(A) - P(A))^+\right)^2\\ &\geq n\sum_{A\in G_n}(P_0(A) - P(A)) + \frac{n}{2c}\left(\sum_{A\in G_n}(P_0(A) - P(A))^+\right)^2 \qquad (4.1)\\ &= n(P_0(U_n) - P(U_n)) + \frac{n}{2c}\left(\sum_{A\in G_n}(P_0(A) - P(A))^+\right)^2.\end{aligned}$$

Now if $\pi_n = G_n$ is a partition which exhausts all of $\mathbf{X}$, then $P_0(U_n) - P(U_n) = 0$ and $\sum_{A \in \pi_n} (P_0(A) - P(A))^+ = (1/2)d_{\pi_n}(P_0, P)$. In this case, (4.1) reduces to

$$\frac{n}{8c}\left(d_{\pi_n}(P_0, P)\right)^2,$$

which is not less than nt_1 for all P with $d_{\pi_n}(P_0, P) \geq \sqrt{8ct_1}$. This completes the proof of Lemma 3. □

We are now ready to establish our main result concerning the existence of uniformly powerful tests. The converse to this result will be given in Section 5.

THEOREM 1. *Let π_n be any sequence of partitions with effective cardinality of order n. Then for any $\delta > 0$, there exists a uniformly exponentially consistent test of $H_0 = \{P_0\}$ versus $H_{1,n} = \{P: d_{\pi_n}(P_0, P) \geq \delta\}$.*

PROOF. First we choose some strictly positive constants. Given δ choose $\varepsilon < \delta/2$, let $m_n = m(\pi_n, P_0, \varepsilon)$, and $c = \sup m_n/n$. Choose $t_0 < \delta - 2\varepsilon$ with t_0 small enough that $t_1 = (1/8c)(\delta - 2\varepsilon - t_0)^2 - t_0/2$ is greater than 0. Also choose $r_0 < t_0^2/2$ and $t < t_1$. Finally choose b so large that

$$(1 + 1/b)H(1/(b+1)) < r_0,$$

where $H(\cdot)$ is the binary entropy function, which was defined in Section 3.

Fix n and let $A_1, A_2, \ldots,$ be the sets in π_n ordered so that $P_0(A_1) \geq P_0(A_2) \geq \cdots$. Let $k < n/b$ be the number of these sets which have probability greater than b/n. Let $G_n = \{A_{k+1}, \ldots, A_m\}$, $U_n = A_{k+1} \cup \cdots \cup A_m$, $E_n = A_{m+1} \cup A_{m+2} \cup \cdots$ and $\pi_n^b = \{A_1, A_2, \ldots, A_k, U_n, E_n\}$, where $m = m_n$.

The critical set for the test is

$$C_n = \left\{D_{\pi_n^b}\left(\hat{P}_n \| P_0\right) \geq r_0\right\} \cup \{T_n \geq nt\} = C_{1,n} \cup C_{2,n},$$

where

$$T_n = \sum_{A \in G_n} \left(e^{nP_0(A)} 1_{\{\hat{P}_n(A) = 0\}} - 1\right).$$

The probability of type I error is $P_0^n(C_n) \leq P_0^n(C_{1,n}) + P_0^n(C_{2,n})$. As in the proof of Lemma 1,

$$P_0^n(C_{1,n}) = P_0^n\left\{D_{\pi_n^b}\left(\hat{P}_n \| P_0\right) \geq r_0\right\} \leq e^{-nr_0}\binom{n + n/b}{n/b}$$
$$\leq \exp\{-n(r_0 - (1 + 1/b)H(1/(b+1)))\},$$

which is exponentially small. Next note that by Lemma 3, $P_0^n(C_{2,n}) = P_0^n\{T_n \geq nt\}$ is also exponentially small.

The probabilities of type II error are $P^n(C_n^c) \leq \min\{P^n(C_{1,n}^c), P^n(C_{2,n}^c)\}$. If P is such that $d_{\pi_n^b}(P_0, P) \geq t_0$, then we use

$$P^n(C_n^c) \leq P^n(C_{1,n}^c) = P^n\left\{D_{\pi_n^b}\left(\hat{P}_n \| P_0\right) < r_0\right\}$$

which Lemma 1 shows is less than or equal to e^{-nr_1}, where $r_1 = (t_0 - \sqrt{2r_0})^2/2$.

Finally, consider P for which $d_{\pi_n}(P_0, P) \geq \delta$ and $d_{\pi_n^b}(P_0, P) < t_0$. We use $P^n(C_n^c) \leq P^n(C_{2,n}^c) = P^n\{T_n < nt\}$ which by Lemma 3 is uniformly exponentially small provided that

$$\sum_{A \in G_n} (e^{n(P_0(A) - P(A))} - 1) \geq nt_1.$$

As in the proof of Lemma 3, it is enough to show that

$$\text{(4.2)} \qquad \frac{1}{2c}\left(\sum_{A \in G_n} (P_0(A) - P(A))^+\right)^2 + P_0(U_n) - P(U_n) \geq t_1.$$

To verify this inequality, observe that

$$\begin{aligned} \frac{\delta}{2} \leq \frac{1}{2} d_{\pi_n}(P_0, P) &= \sum_{A \in \pi_n} (P_0(A) - P(A))^+ \\ &\leq \sum_{A \in G_n} (P_0(A) - P(A))^+ + \sum_{A \in \pi_n^b} (P_0(A) - P(A))^+ + P_0(E_n) \\ &\leq \sum_{A \in G_n} (P_0(A) - P(A))^+ + \frac{t_0}{2} + \varepsilon. \end{aligned}$$

Consequently $\Sigma_{A \in G_n} (P_0 A) - P(A))^+ \geq (1/2)(\delta - 2\varepsilon - t_0)$. Furthermore $P_0(U_n) - P(U_n) \geq t_0/2$. Therefore the left side of (4.2) is greater than or equal to

$$\frac{1}{8c}(\delta - 2\varepsilon - t_0)^2 - \frac{t_0}{2} = t_1$$

as required. This completes the proof of the theorem. □

5. Converse: Nonexistence of uniformly powerful tests. The converse to Theorem 1 is the following result.

THEOREM 2. *If a sequence of partitions π_n does not have effective cardinality of order n, then there does not exist a uniformly exponentially consistent test of $H_0 = \{P_0\}$ versus $H_{1,n} = \{P: d_{\pi_n}(P_0, P) \geq \delta\}$ for some $\delta > 0$.*

PROOF. If the sequence of partitions does not have effective cardinality of order n, then for some $0 < \beta < 1$,

$$\limsup_{n \to \infty} \frac{m(\pi_n, P_0, \beta)}{n} = \infty.$$

Here $m_n = m(\pi_n, P_0, \beta)$ is the minimum number of sets in π_n which exhaust at least $1 - \beta$ of the probability under P_0. We show that no uniformly exponentially consistent test exists against $\{P: d_{\pi_n}(P_0, P) \geq \delta\}$ for any $0 < \delta < \beta$.

Before giving all the particulars, we outline some of the ideas. We find a sequence of finite sets K_n of probability measures such that each P in K_n has

$d_{\pi_n}(P_0, P)$ near β, yet there does not exist a UEC test against K_n. Indeed, it is shown that for any sequence of tests which has type I error tending to 0, the average of the type II error probability (with respect to a uniform prior on K_n) is not exponentially small.

The set of alternatives K_n is obtained from a partition $\tilde{\pi}_n = \{B_0, B_1, \dots, B_{\tilde{m}_n}\}$ which has the properties that $\tilde{\pi}_n$ is coarser than π_n, the cardinality $\tilde{m}_n$ is of larger order than n, the probability $P_0(B_0)$ is near $1 - \beta$, and $P_0(B_1), \dots, P_0(B_{\tilde{m}_n})$ are all near $\beta/\tilde{m}_n$. For each of the ways of selecting one-half of the sets $B_1, B_2, \dots, B_{\tilde{m}_n}$, we construct a P in K_n which has dP/dP_0 near 2 on the selected sets, equal to 0 on the unselected sets and equal to 1 on B_0.

Now for the particulars. Since $\limsup m_n/n = \infty$, there exists a sequence $\tilde{m}_n$ such that $\tilde{m}_n/m_n \to 0$, $\limsup \tilde{m}_n/n = \infty$, and for convenience $\tilde{m}_n$ is an even integer. Set $\varepsilon_n = \tilde{m}_n/(m_n\beta)$ and note that $\lim \varepsilon_n = 0$.

We restrict attention to the subsequence with $n < \tilde{m}_n/2$. Fix n in this subsequence and for notational convenience write m for m_n and $\tilde{m}$ for $\tilde{m}_n$. Order the sets in the partition $\pi_n = \{A_1, A_2, \dots\}$ so that $P_0(A_1) \ge P_0(A_2) \ge \cdots$. Set $k_0 = m - 1$ and define increasing integers $k_1, k_2, \dots,$ recursively as follows: Given k_{j-1}, let k_j be the last k such that $P_0(A_{k_{j-1}+1} \cup \cdots \cup A_k) \le \beta/\tilde{m}$. Set

$$B_j = A_{k_{j-1}+1} \cup \cdots \cup A_{k_j} \quad \text{for } j = 1, 2, \dots, \tilde{m}$$

and

$$B_0 = \mathbf{X} - (B_1 \cup \cdots \cup B_{\tilde{m}}).$$

Now since the $P_0(A_k)$ are decreasing and sum to 1 we have that $P_0(A_k) \le 1/k$ for $k = 1, 2, \dots$. In particular $P_0(A_k) \le 1/m$ for all $k \ge m$. It follows that $P_0(B_j)$ is between $\beta/\tilde{m} - 1/m$ and $\beta/\tilde{m}$ for $j = 1, 2, \dots, m$. Also $1 - P_0(B_0) = \sum_{j=1}^{\tilde{m}} P_0(B_j)$ is between $\beta - \tilde{m}/m = \beta(1 - \varepsilon_n)$ and β.

Let U be a union of $\tilde{m}/2$ of the sets $B_1, B_2, \dots, B_{\tilde{m}}$. Each U has probability $P_0(U) \le \beta/2$. To each U there corresponds a probability measure P with

$$\frac{dP}{dP_0}(x) = \begin{cases} 1, & \text{for } x \in B_0, \\ (1 - P_0(B_0))/P_0(U), & \text{for } x \in U, \\ 0, & \text{otherwise.} \end{cases}$$

We let K_n be the set of all such probability measures. For each P in K_n the distance from P_0 satisfies

$$\begin{aligned} d_{\pi_n}(P_0, P) \ge d_{\tilde{\pi}_n}(P_0, P) &= \sum_{B \in \tilde{\pi}_n} |P_0(B) - P(B)| \\ &= 2(1 - P_0(B_0) - P_0(U)) \\ &\ge \beta(1 - 2\varepsilon_n). \end{aligned}$$

This bound holds uniformly for all P in K_n. For any $0 < \delta < \beta$, it follows that K_n is a subset of $\{P: d_{\pi_n}(P_0, P) > \delta\}$ for all large n.

Let P^n denote the n-fold product of a measure P in K_n. The density of P^n with respect to P_0^n is

$$\frac{dP^n}{dP_0^n}(x) = \prod_{i=1}^{n} \frac{dP}{dP_0}(x_i) = \left(\frac{1 - P_0(B_0)}{P_0(U)}\right)^{n-N(B_0)} 1_{\{S \subset U\}}$$

for $x = (x_1, \ldots, x_n)$ in $\mathbf{X}^n$. Here $N(B) = \sum 1_B(x_i)$ is the observation count for $B \in \tilde{\pi}_n$, and S is the union of those sets B_j, $j = 1, 2, \ldots, \tilde{m}$, for which $N(B_j) > 0$. The number of such occupied B_j is denoted by k. Of course $N(B)$, S and k depend on x. Using the available bounds on $P_0(B_0)$ and $P_0(U)$ we have

$$\frac{dP^n}{dP_0^n}(x) \geq (2(1 - \varepsilon_n))^{n-N(B_0)} 1_{\{S \subset U\}}.$$

Define the measure Q_n on $(\mathbf{X}^n, \mathbf{B}^n)$ to be the uniform mixture of the product measures. Thus

$$Q_n(A) = \frac{1}{\|K_n\|} \sum_{P \in K_n} P^n(A) \quad \text{for } A \in \mathbf{B}^n.$$

This measure has density with respect to P_0^n given by

$$\frac{dQ_n}{dP_0^n} = \frac{1}{\|K_n\|} \sum_{P \in K_n} \frac{dP^n}{dP_0^n} \geq \frac{1}{\|K_n\|} (2(1 - \varepsilon_n))^{n-N(B_0)} \sum_{P \in K_n} 1_{\{S \subset U\}}.$$

Now the sum $\sum 1_{\{S \subset U\}}$ is the number of choices for U which cover S. This number is seen to equal $\binom{\tilde{m} - k}{\tilde{m}/2 - k}$. Also the cardinality of K_n is $\binom{\tilde{m}}{\tilde{m}/2}$. Consequently, the lower bound on the density is

$$\frac{dQ_n}{dP_0^n} \geq (2(1 - \varepsilon_n))^{n-N(B_0)} \left(\binom{\tilde{m} - k}{\tilde{m}/2 - k} \Big/ \binom{\tilde{m}}{\tilde{m}/2}\right).$$

The ratio of binomial coefficients simplifies to a product of k fractions each of which exceeds $(\tilde{m}/2 - k)/\tilde{m} \geq (1/2) - (n/\tilde{m})$. Using $k \leq n - N(B_0)$ we have

$$\frac{dQ_n}{dP_0^n} \geq (2(1 - \varepsilon_n))^{n-N(B_0)} \left(\frac{1}{2} - \frac{n}{\tilde{m}}\right)^{n-N(B_0)}$$

$$= \left((1 - \varepsilon_n)\left(1 - \frac{2n}{\tilde{m}}\right)\right)^{n-N(B_0)} \tag{5.1}$$

$$\geq \left((1 - \varepsilon_n)\left(1 - \frac{2n}{\tilde{m}}\right)\right)^{n}.$$

This lower bound holds uniformly for all x in X^n. Now set $r_n = -\ln(1 - \varepsilon_n) - \ln(1 - 2n/\tilde{m}_n)$ for $n < \tilde{m}/2$ and $r_n = \infty$ for $n \geq \tilde{m}/2$. Note that as $n \to \infty$ we have $\liminf r_n = 0$. Thus we have

$$\frac{dQ_n}{dP_0^n} \geq e^{-nr_n},$$

which is not exponentially small.

Finally, if $A_n \in \mathbf{B}^n$ is any acceptance set for P_0 with $P_0^n(A_n^c) \leq \alpha < 1$, then the maximum probability of type II error satisfies

$$\begin{aligned}
\sup_{P \in K_n} P^n(A_n) &\geq \frac{1}{\|K_n\|} \sum_{P \in K_n} P^n(A_n) \\
&= Q^n(A_n) = \int_{A_n} \left(\frac{dQ_n}{dP_0^n} \right) dP_0^n \qquad (5.2) \\
&\geq e^{-nr_n} P_0^n(A_n) \\
&\geq e^{-nr_n}(1 - \alpha),
\end{aligned}$$

which is not exponentially small. Since $K_n \subset \{P: d_{\pi_n}(P_0, P) > \delta\}$ for all large n, this completes the proof of Theorem 2. □

A consequence of the calculations in Theorem 2 is the following result which is proved by other means in Le Cam [(1973), Proposition 2].

THEOREM 3. *Let P_0 be any nondiscrete probability measure on a measurable space. Then for all sufficiently small $\delta > 0$, there does not exist a uniformly consistent test for P_0 versus $\{P: d(P, P_0) > \delta\}$, where $d(P, P_0)$ is the total variation distance.*

PROOF. Since P_0 is not discrete, there exists an event A which can be carved into disjoint measurable sets of arbitrarily small positive probability [let $\gamma = P_0(A)$]. Then for every $0 < \beta < \gamma$, there exists partitions with arbitrarily large $(1 - \beta)$ effective cardinality. Given any $0 < \delta < \gamma$, choose $\delta < \beta < \gamma$ and let π_n be a sequence of partitions with $(1 - \beta)$ effective cardinality equal to $m_n = n^3$.

If in the proof of Theorem 2, we set $\tilde{m}_n = n^2$, then $\varepsilon_n = 1/(n\beta)$ and (5.1) gives a lower bound of

$$e^{-nr_n} = \left(\left(1 - \frac{1}{n\beta}\right)\left(1 - \frac{2}{n}\right) \right)^n,$$

which does not tend to 0. Let $H_1 = \{P = d(P, P_0) > \delta\}$ denote the set of alternatives. Now since $d(P, P_0) \geq d_{\pi_n}(P, P_0)$ we have that H_1 contains the set $\{P: d_{\pi_n}(P, P_0) > \delta\}$. Using (5.2), if A_n is any sequence of acceptance sets with $P_0^n(A_n^c) \leq \alpha < 1$, then

$$\liminf_{n \to \infty} \sup_{P \in H_1} P^n(A_n) \geq \liminf e^{-nr_n}(1 - \alpha) > 0.$$

Therefore, no uniformly consistent test exists for P_0 versus $\{P: d(P, P_0) > \delta\}$. □

REMARK. It is interesting to note that for any given n, we may chose a partition with effective cardinality sufficiently large that the bound in (5.1) is arbitrarily close to 1. Consequently, for every n, $\sup_{H_1} P^n(A_n) \geq (1 - \alpha)$, whenever $P_0^n(A_n^c) \leq \alpha$. It follows that a minimax test for these hypotheses is to ignore the data and flip a Bernoulli (α) coin.

Of course, the same result on the nonexistence of uniformly consistent tests holds if the total variation distance is replaced by any measure of distance or divergence which dominates the total variation.

6. Smooth alternatives. Uniformly consistent tests are shown to exist against all sufficiently smooth alternatives which are outside a total variation distance neighborhood of the null hypothesis. We restrict attention to the case that the measurable space is the real line with the Borel sets. The hypothesized probability measures are required to have density functions with respect to Lebesgue measure. In which case the total variation distance between the probability measures reduces to the L^1 distance between the density functions, $d(p, p_0) = \int |p(x) - p_0(x)|\,dx$. Let $\delta(\varepsilon, p)$, $\varepsilon > 0$, denote the inverse of the modulus of continuity of a function p [i.e., the supremum of δ for which $|x - y| < \delta$ implies $|p(x) - p(y)| < \varepsilon$]. For a set K of functions, let $\delta(\varepsilon, K) = \inf_K \delta(\varepsilon, p)$. Recall that a set K of functions is uniformly equicontinuous if $\delta(\varepsilon, K)$ is positive for all $\varepsilon > 0$.

THEOREM 4. *Let K be any set of uniformly equicontinuous density functions. Then for any probability density function p_0 and any $\varepsilon > 0$, there exists a uniformly exponentially consistent test for p_0 versus $\{p: d(p, p_0) \geq \varepsilon\} \cap K$.*

Moreover, if K_n is a sequence of sets of density functions with $\liminf_{n \to \infty} n\delta(\varepsilon, K_n) > 0$ for all $\varepsilon > 0$, then there exists a uniformly exponentially consistent test for p_0 versus $\{p: d(p, p_0) \geq \varepsilon\} \cap K_n$.

Thus the degree of smoothness $\delta(\varepsilon, K_n)$ may be as small as order $1/n$.

PROOF. Let $\pi = \pi^h$ be a uniform partition of the line into intervals of width h. For any probability density function p, let p^π denote the histogram with $p^\pi(x) = \int_A p(y)\,dy/h$ for $x \in A$ and $A \in \pi$. It is known that $\lim_{h \to 0} d(p, p^\pi) = 0$ for any measurable density function p [Abou-Jaoude (1976)], so given $\varepsilon > 0$ we may choose $h_0 > 0$ such that $d(p_0, p_0^\pi) \leq \varepsilon/4$ for all $0 < h \leq h_0$.

Now choose c sufficiently large that $P_0[\{|X| \geq c\} < \varepsilon/8$. Let $h = h_n$ be the minimum of $\delta_n(\varepsilon/16c, K_n)$ and h_0 and let $\pi = \pi^h$. If $p(x)$ is continuous then by the mean value theorem, for every x there is a y with $|x - y| < h$ such that $p^\pi(x) = p(y)$. It follows that for all p in K_n and all x, $|p^\pi(x) - p(x)| \leq \varepsilon/16c$.

Write the L^1 distance between the probability densities in terms of the positive part $d(p_0, p) = 2\int (p_0(x) - p(x))^+\,dx$. Noting that the integrand is dominated by $p_0(x)$ we bound the contribution in the tails,

$$d(p_0, p) \le 2\int_{-c}^{c} (p_0(x) - p(x))^+ \, dx + \frac{\varepsilon}{4}.$$

Using the triangle inequality we obtain for all p in K_n that

$$d(p_0, p) \le 2\int_{-c}^{c} (p_0 - p_0^\pi)^+ + 2\int_{-c}^{c} (p_0^\pi - p^\pi)^+ + 2\int_{-c}^{c} (p^\pi - p)^+ + \frac{\varepsilon}{4}$$

$$\le d_\pi(P_0, P) + \frac{3\varepsilon}{4}.$$

Thus for any such p with $d(p_0, p) \ge \varepsilon$ we have $d_\pi(P_0, P) \ge \varepsilon/4$, so the set of alternatives is a subset of $\{P: d_\pi(P_0, P) \ge \varepsilon/4\}$. Now $\liminf nh_n > 0$ so the sequence of partitions has effective cardinality of order n. Therefore, a UEC test exists against $\{p: d(p_0, p) \ge \varepsilon\} \cap K_n$. □

7. An extension. For clarity this paper has restricted attention to a simple null hypothesis $H_0 = \{P_0\}$ which does not depend on the sample size. Nevertheless, it is seen that the proofs allow simple null hypotheses $H_{0,n} = \{P_{0,n}\}$ which may depend on n. The necessary and sufficient condition on the sequence of pairs $\{P_{0,n}, \pi_n\}$ for the existence of UEC tests against $\{P: d_{\pi_n}(P_{0,n}, P) \ge \delta\}$ for all $\delta > 0$ is that the $1 - \delta$ effective cardinality satisfy $\limsup(m(\pi_n, P_{0,n}, \delta)/n) < \infty$ for all $\delta > 0$. Note in particular that instead of fixing P_0 and regarding the condition as a restriction imposed on π_n, we could fix a countable partition π (which amounts to restricting attention to a discrete space) and then regard $m(\pi, P_{0,n}, \delta) = O(n)$ as a restriction imposed on the sequence of simple null hypotheses. This different setting for the problem was suggested by a referee.

Acknowledgments. Portions of the work described in this paper appeared in the author's dissertation [Barron (1985)]. The author acknowledges his gratitude for the guidance of Professor Thomas M. Cover.

REFERENCES

ABOU-JAOUDE, S. (1976). Conditions nécessaires et suffisantes de convergence L^1 en probabilité de l'histogramme pour une densité. *Ann. Inst. H. Poincaré Sect. B (N.S.)* **12** 213–231.

BAHADUR, R. R. (1966). An optimal property of the likelihood ratio statistic. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 13–26. Univ. California Press.

BAHADUR, R. R., GUPTA, J. C. and ZABELL, S. L. (1980). Large deviations, tests and estimates. In *Asymptotic Theory of Statistical Tests and Estimation* (I. M. Chakravarti, ed.) 33–64. Academic, New York.

BARRON, A. R. (1985). Logically smooth density estimation. Ph.D. dissertation, Stanford Univ.

BARRON, A. R. (1986). Comment on "On the consistency of Bayes estimates", by P. Diaconis and D. Freedman. *Ann. Statist.* **14** 26–30.

BARRON, A. R. (1987). The exponential convergence of posterior probabilities with implications for Bayes estimators of density functions. Technical Report 7, Dept. Statistics, Univ. Illinois.

BLAHUT, R. E. (1987). *Information Theory.* Addison-Wesley, Reading, Mass.

BROWN, L. D. (1971). Non-local asymptotic optimality of appropriate likelihood ratio tests. *Ann. Math. Statist.* **42** 1206–1240.

CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on a sum of observations. *Ann. Math. Statist.* **23** 493–507.

CHERNOFF, H. (1956). Large sample theory: Parametric case. *Ann. Math. Statist.* **27** 1–22.

CSISZÁR, I. (1967). Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.* **2** 299–318.

CSISZÁR, I. (1974). Information measures: A critical survey. In *Trans. Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes* 73–86. Academia, Prague.

CSISZÁR, I. (1984). Sanov property, generalized *I*-projection and a conditional limit theorem. *Ann. Probab.* **12** 768–793.

CSISZÁR, I. and LONGO, G. (1971). On the error exponent for source coding for testing simple statistical hypotheses. *Studia Sci. Math. Hungar.* **6** 181–191.

HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

HOEFFDING, W. (1965). Asymptotically optimal tests for multinnomial distributions. *Ann. Math. Statist.* **36** 369–401.

HOEFFDING, W. and WOLFOWITZ, J. (1958). Distinguishability of sets of distributions. *Ann. Math. Statist.* **29** 700–718.

HOLST, L. (1972). Asymptotic normality and efficiency for certain goodness-of-fit tests. *Biometrika* **59** 137–145.

HUBER, P. J. and STRASSEN, V. (1973). Minimax tests and the Neyman–Pearson lemma for capacities. *Ann. Statist.* **1** 251–263.

KALLENBERG, W. C. M. (1985). On moderate and large deviations in multinomial distributions. *Ann. Statist.* **13** 1554–1580.

KULLBACK, S. (1967). A lower bound for discrimination information in terms of variation. *IEEE Trans. Inform. Theory* **IT-13** 126–127.

LE CAM, L. (1973). Convergence of estimates under dimensionality restrictions. *Ann. Statist.* **1** 38–53.

LE CAM, L. and SCHWARTZ, L. (1960). A necessary and sufficient condition for the existence of consistent estimates. *Ann. Math. Statist.* **31** 140–150.

QUINE, M. P. and ROBINSON, J. (1985). Efficiencies of chi-square and likelihood ratio goodness-of-fit tests. *Ann. Statist.* **13** 727–742.

SCHWARTZ, L. (1965). On Bayes' procedures. *Z. Wahrsch. verw. Gebiete* **4** 10–26.

TUSNÁDY, G. (1977). On asymptotically optimal tests. *Ann. Statist.* **5** 385–393.

VAPNIK, V. N. and CHERVONENKIS, A. Y. (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory Probab. Appl.* **16** 264–280.

DEPARTMENT OF STATISTICS AND
DEPARTMENT OF ELECTRICAL AND
COMPUTER ENGINEERING
101 ILLINI HALL
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
725 SOUTH WRIGHT STREET
CHAMPAIGN, ILLINOIS 61820