NONPARAMETRIC LIKELIHOOD CONFIDENCE BANDS
FOR A DISTRIBUTION FUNCTION

BY

ART B. OWEN

TECHNICAL REPORT NO. 442
OCTOBER 1993

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT DMS92-04864

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
Nonparametric Likelihood Confidence Bands
For a Distribution Function

By

Art B. Owen

Technical Report No. 442
October 1993

Prepared Under the Auspices
Of
National Science Foundation Grant DMS92-04864

Also Supported by National Science Foundation Grant DMS-9011074.

Department of Statistics
Stanford University
Stanford, California
Nonparametric Likelihood Confidence Bands for a Distribution Function

Art B. Owen

Department of Statistics
Stanford University
Stanford CA 94305
ABSTRACT

Given an i.i.d. sample, one can test whether it comes from a distribution $F$, by using the Kolmogorov-Smirnov statistic. Berk and Jones (1979) show that a certain nonparametric likelihood test for uniformity is more efficient (in Bahadur's sense) than any weighted Kolmogorov-Smirnov test at any alternative to $F$. This test can be inverted to form confidence bands for a distribution function $F$, if we know the null distribution of the test statistic. Berk and Jones give an asymptotic approximation for the null distribution, but it is of extreme value type, based on doubly and triply iterated logarithms, and it takes effect very slowly. In this paper we use a recursion due to Noé (1972) to get exact non-asymptotic results for the null distribution.

Nonparametric likelihood bands are narrower than Kolmogorov-Smirnov bands in the tails and wider in the center. The bands are also skewed, in that the empirical cumulative distribution function is not centered between the upper and lower confidence limits. This skewness is greatest in the tails of the sample.

This paper describes how to compute the likelihood bands and gives a formula for the critical value of the likelihood which attains coverage between 0.95 and 0.9501 for all sample sizes between 2 and 100 inclusive. Another formula covers the range from 101 to 1000. Nonparametric likelihood bands are illustrated on a real data set. Direct calculations show that for $n = 20$, the test underlying the likelihood bands is more powerful than that underlying the Kolmogorov-Smirnov for alternatives of the form $F^\eta$, $\eta > 1$.

KEY WORDS: Bahadur slope, Empirical Distribution.
1. INTRODUCTION

Given an i.i.d. sample from some distribution, the natural estimator of that distribution is the (empirical) cumulative distribution function, or CDF. (This and other terms are defined in the next section.) When plotting the CDF, it is often desirable to add confidence bands; that is, two functions with prespecified probability of containing the true distribution function over the whole real line. Such bands are usually constructed through a test statistic, so that only those distribution functions not rejected by the test statistic lie completely within the bands. The most widely known bands are derived from the Kolmogorov-Smirnov statistic. These bands are parallel to the CDF. It is often desirable to use bands that are narrower in the tails than in the center of the data set, and this can be done through weighted Kolmogorov-Smirnov statistics.

When the sample comes from a continuous distribution function, the construction of confidence bands can be reduced to the case of a sample from the standard uniform distribution. Berk and Jones (1979) give a likelihood based test statistic that is more efficient (in Bahadur’s sense) than any weighted Kolmogorov-Smirnov test statistic at any alternative to the uniform distribution. To make bands from this statistic, we need to know its null distribution. Berk and Jones give asymptotic approximations to the null distribution for large \( n \). In this paper, we compute exact quantiles of the null distribution of Berk and Jones’s statistic for finite \( n \), using a recursion of Noé (1972).

The paper is organized as follows: Section 2 gives notation and reviews some standard results on confidence bands for distribution functions. Section 3 describes the nonparametric likelihood confidence bands. Section 4 describes computation of the bands. A critical value used in constructing 95% bands must be obtained. An approximation to this critical value is given which gives coverage between 95% and 95.01% for \( 1 < n \leq 100 \). Another approximation covers the range \( 100 < n \leq 1000 \). Section 5 compares the likelihood bands to Kolmogorov-Smirnov bands, on a data set, and for alternatives of the form \( F^n \) for \( n = 20 \) and \( \eta > 1 \). Section 6 provides a discussion of when these bands ought to be used. Appendix A sketches two theoretical properties of the tests underlying these bands: their relative optimality as defined by Berk and Jones (1978) and their
increased Bahadur efficiency over any weighted Kolmogorov-Smirnov method at any alternative distribution.

2. NOTATION

Let \( X_1, \ldots, X_n \) be an i.i.d. sample with distribution function \( F(x) = P(X_1 \leq x) \). Unless otherwise stated, we assume that \( F \) is continuous. Let \( F_n \) be the empirical distribution function of the sample, that is

\[
F_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{(-\infty, x]}(X_i)
\]

where \( 1_A(x) \) equals 1 if \( x \in A \) and 0 otherwise.

It is well known that we can write \( X_i = F^{-1}(U_i) \) where \( U_i \) are i.i.d. \( U[0,1] \) random variables and

\[
F^{-1}(u) = \inf \{ x : F(x) \leq u \}
\]

is the usual inverse distribution function. Because \( F \) is continuous we can also write \( U_i = F(X_i) \). See Shorack and Wellner (1986, Chapter 1) for these and related results. The order statistics of the sample are \( X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)} \) and similarly, we may arrange the \( U_i \) in non-decreasing order by taking \( U_{(i)} = F(X_{(i)}) \).

A \( 1 - \alpha \) central confidence band for \( F \) is given by two functions \( L(x) \) and \( H(x) \) where

\[
P \left( L(x) \leq F(x) \leq H(x), \ -\infty < x < \infty \right) = 1 - \alpha.
\]

The randomness in (2.2) comes from \( L(x) \) and \( H(x) \) being functions of the data \( X_1, \ldots, X_n \). This dependence is suppressed for notational convenience. One sided bands correspond to special cases in which \( L(x) \) is identically 0 or \( H(x) \) is identically 1.

The Kolmogorov-Smirnov statistic for testing whether \( X_i \sim F \) is

\[
D_n = \sup_{-\infty < x < \infty} |F_n(x) - F(x)|.
\]

The greatest difference between \( F_n \) and \( F \) has to occur either at an observed value \( X_i \), or in the
limit as \( z \to X_i \) from below. Using this and the continuity of \( F \),

\[
D_n = \max_{1 \leq i \leq n} \max \left( \left| \frac{i-1}{n} - F(X_{(i)}) \right|, \left| \frac{i}{n} - F(X_{(i)}) \right| \right)
= \max_{1 \leq i \leq n} \max \left( \left| \frac{i-1}{n} - U_{(i)} \right|, \left| \frac{i}{n} - U_{(i)} \right| \right).
\]

The \( U_{(i)} \) act as pivots here; they are not observable since they depend on the true unknown \( F \), but their joint distribution is known and it determines the distribution of \( D_n \). Clearly \( D_n \) is distribution-free in that it’s distribution does not depend on which continuous \( F \) generates the data.

If we know that \( P(D_n > \lambda) = 1 - \alpha \), when \( X_i \) really have distribution \( F \), then we can form \( 1 - \alpha \) bands for \( F \) by taking \( H(x), L(x) = F_n(x) \pm \lambda \). Obviously, the bands \( L', H' \) with \( L'(x) = \max(L(x), 0) \) and \( H'(x) = \min(H(x), 1) \) also have coverage \( 1 - \alpha \) and these are the ones that are used. One sided bands may be obtained from one sided test statistics

\[
D_n^+ = \sup_{-\infty < x < \infty} (F_n(x) - F(x))^+
\]

and

\[
D_n^- = \sup_{-\infty < x < \infty} (F_n(x) - F(x))^-.\]

Here \( d_+ \) is \( d \) if \( d > 0 \) and 0 otherwise, and \( d_- \) is \( |d| \) if \( d < 0 \) and 0 otherwise. As for \( D_n \), the maximum of \( D_n^\pm \) takes place at or just left of an observation and the same reduction to the \( U[0,1] \) setting holds. Shorack and Wellner (1986, Chapter 3.8) discuss the asymptotic distributions of \( D_n \) and \( D_n^\pm \).

Test statistics of the form \( D_n \psi = \sup_x \psi(F(x))|F_n(x) - F(x)| \) give rise to the weighted Kolmogorov-Smirnov statistics, considered in Berk and Jones (1979). Here \( \psi(u) \geq 0 \) is a weight function defined on \( 0 \leq u \leq 1 \). Having \( \psi \) depend on \( x \) only through \( F(x) \) makes it possible to reduce any continuous \( F \) to the \( U[0,1] \) case. Confidence bands based on \( D_n \psi \) have width proportional to \( 1/\psi(F(x)) \).

Motivated by relative error, Rényi (1953) proposes \( \psi = 1/F(x) \). Anderson and Darling (1952) propose

\[
\psi_{AD} = (F(x)(1 - F(x)))^{-1/2}, \tag{2.3}
\]
so that the bands have width at \( x \) proportional to the standard deviation of \( F_n(x) \). The resulting test statistic should not be confused with

\[
A_n^2 = n \int_\infty^{-\infty} \frac{(F_n(x) - F(x))^2}{F(x)(1 - F(x))} dx
\]

(2.4)

which is widely known as the Anderson-Darling statistic.

3. LIKELIHOOD TESTS

The distribution of \( nF_n(x) \) is binomial with parameters \( n \) and \( F(x) \). The Kolmogorov-Smirnov test statistics are based on the departures of \( F_n(x) \) from its expectation \( F(x) \). Anderson and Darling's weight function normalizes proportionately to the standard deviation of \( F_n(x) \). Berk and Jones (1979) consider replacing these moment criteria by a likelihood criterion

\[
R_n = \sup_{-\infty < x < \infty} K(F_n(x), F(x))
\]

(3.1)

where

\[
K(\hat{p}, p) = \hat{p} \log(\hat{p}/p) + (1 - \hat{p}) \log((1 - \hat{p})/(1 - p)).
\]

(3.2)

Impossible combinations, such as \( F(x) = 0 \) and \( F_n(x) > 0 \) at the same \( x \), lead to \( R_n = \infty \). It is understood that \( 0 \log 0 = 0 \) in (3.2). One sided versions \( R_n^+ \) and \( R_n^- \) are defined by only considering \( x \) with \( F_n > F \) and \( F_n < F \) respectively, in the supremum in (3.1). (Should no such \( x \) exist, take \( R_n^+ = 0 \).) The likelihood connection is as follows: \( -nK(\hat{p}, p) \) is the log likelihood ratio for the probability parameter \( p \) based on a binomial observation of \( n\hat{p} \) successes in \( n \) trials. Large values of \( K \) provide evidence that \( P(X \leq x) \neq F(x) \) and \( R_n \) uses the \( x \) at which this evidence is strongest.

Formula (3.1) simplifies to

\[
R_n = \max_{1 \leq i \leq n} \max \left( K\left(\frac{i - 1}{n}, U(i)\right), K\left(\frac{i}{n}, U(i)\right)\right)
\]

(3.3)

for the same reasons that the Kolmogorov-Smirnov statistic simplifies.

If \( P(R_n > \lambda_n) = 1 - \alpha \) then a \( 1 - \alpha \) confidence band may be formed by taking

\[
L(x) = \min \left\{ p \mid K(F_n(x), p) \leq \lambda_n \right\}
\]
and
\[ H(x) = \max \{ p \mid K(F_n(x), p) \leq \lambda_n \}. \]

Like \( F_n \), both \( L(x) \) and \( H(x) \) are step functions that jump only at observed values of \( X_i \). These \( n \) observations divide the real line into \( n + 1 \) intervals, and so \( L \) and \( H \) need only be computed at \( n + 1 \) places.

Nonparametric likelihood confidence bands are formed by making likelihood ratio tests based on the \( \text{Bin}(n, F(x)) \) distribution of \( F_n(x) \). A candidate \( F(x) \) is rejected if it has too small a likelihood at any \( x \). The rejection threshold for the likelihood must account for the fact that an extreme has been taken over \( 2n \) dependent likelihood ratios in (3.3). The resulting bands contain all distribution functions not rejected by the likelihood ratio test.

4. COMPUTATION

Here we describe how to compute the nonparametric likelihood confidence bands. First, we show how to compute the limits \( L(x) \) and \( H(x) \), given a likelihood threshold \( \lambda_n \). Next we show how to find the coverage level \( 1 - \alpha \) corresponding to given values of \( n \) and \( \lambda_n \). Then we use this to solve for \( \lambda_n^{1-\alpha} \) corresponding to a given sample size \( n \) and confidence level \( 1 - \alpha \). This step requires many applications of the first two steps, so it is worth while making the first steps efficient. Finally we provide a simple approximate formula for \( \lambda_n^{0.95} \) that provides coverage between 0.95 and 0.9501 for any \( 1 < n \leq 100 \), and another formula for \( 100 < n \leq 1000 \).

To plot the bands we only need \( n + 1 \) values for each of \( L(x) \) and \( H(x) \). Let \( L_i, H_i \), \( i = 0, 1, \ldots, n \) be the value of \( L(x) \) and \( H(x) \) on the open interval \((X_{(i)}, X_{(i+1)})\) taking \( X(0) = -\infty, X_{(n+1)} = \infty \). At a data point, \( H(X_{(i)}) = \max(H_{i-1}, H_i) = H_i \) and \( L(X_{(i)}) = \min(L_{i-1}, L_i) = L_{i-1} \). That is \( H(x) \) is continuous from the right and \( L(x) \) is continuous from the left.

Here is how we might calculate \( H(x) \). For each \( \hat{p} \in (0, 1) \), \( K(\hat{p}, p) \) is a continuous function of \( p \in (0, 1) \), taking a minimum of 0 at \( p = \hat{p} \) and diverging to \( \infty \) as \( p \) tends to 0 or 1. Therefore, for each \( \lambda_n > 0 \), a unique \( H(x) \) exists, and it lies between \( \hat{p} \) and 1. Since we know an interval containing \( H(x) \), the Van Wijngaarden-Decker-Brent method can be used to find \( H(x) \). It has
the reliability of the bisection algorithm but it often converges at a superlinear rate, using inverse quadratic interpolation. This algorithm is implemented as the function \textit{zbrent} in Press, Flannery, Teukolsky and Vetterling (1988, Chapter 9), and as the subroutine \textit{zeroin} in Forsythe, Malcolm and Moler (1977, Chapter 7). If \( \hat{p} = 1 \), take \( H(x) = 1 \), and if \( \hat{p} = 0 \), it is easy to see that \( H(x) = 1 - e^{-\lambda x} \).

An obvious modification works for \( L(x) \) using bounds 0 and \( \hat{p} \). By symmetry

\[
L_i = 1 - H_{n-i}, \quad 0 \leq i \leq n
\]  

(4.1)

so we can reduce the computation by half.

A further speedup may be obtained by using sharper bounds for \( H_i \) than \( \hat{p} \) and 1. For example if \( H_{i+1} \) has been calculated it is a sharper upper bound for \( H_i \) than is 1. The points \((i, H_i)\) always seem to lie on a concave curve and this suggests that for \( i < j < k \) the value \( \eta H_i + (1 - \eta) H_k \), where \( \eta = (k - j)/(k - i) \) might provide a better lower bound than \( \hat{p} \) for \( H_j \). Without a mathematical proof of the convexity of \((i, H_i)\) it is prudent for software to test bounds derived on that basis before using them. These bounds have been tried over many combinations of \( n \) and \( 1 - \alpha \), and have never failed to hold. So far only modest improvements in speed seem to accrue from the sharper bounds.

We want to find \( \lambda_n \) for a given confidence level \( 1 - \alpha \). It turns out to be easier to find the confidence level \( 1 - \alpha \) corresponding to a given \( \lambda_n \). Given \( \lambda_n \), the procedure above gives us \( L_i, H_i \) and so

\[
P(R_n < \lambda_n) = P(L_{i-1} < U(i) < H_i, \quad i = 1, \ldots, n).
\]  

(4.2)

A numerically stable recursion for (4.2) is given by Noé (1972) and is reproduced in Shorack and Wellner (1986, P. 362). We then solve (4.2) for \( \lambda_n^{1-\alpha} \) as that value of \( \lambda \) for which

\[
1 - \alpha = P(R_n < \lambda).
\]

As above, the Van Wijngaarden-Decker-Brent method is applicable provided that we have upper and lower bounds for \( \lambda_n^{1-\alpha} \). Since \( \lambda_n^{1-\alpha} \) increases with \( n \), \( \lambda_n^{1-\alpha} \) is a suitable lower bound for \( \lambda_n^{1-\alpha} \), if one is interested in finding \( \lambda_n^{1-\alpha} \) for all \( n \leq N \) for some \( N \). If one is only interested in one value of
n then $\lambda_i^{1-\alpha} = -n^{-1}\log(1 - \alpha)$ is available. A lower value for $\lambda_n^{1-\alpha}$ may be obtained by observing that the critical log likelihood ratio $-n\lambda_n^{1-\alpha}$ plotted versus $\log(n)$ is convex. The secant through points with $n = 1$ and $n = 2$ provides a lower bound on the critical log likelihood which leads to an upper bound on $\lambda_n^{1-\alpha}$, for $n > 2$. This assumption is tested in the software and has never been violated, though it hasn’t been proved either. A lower bound for $\lambda_i^{1-\alpha}$ may be found by trial and error.

To calculate $\lambda_i^{0.95}, \ldots, \lambda_{100}^{0.95}$ takes roughly 65 seconds of cpu on a DECstation 5000/240 workstation. To construct 95% bands for $n = 1, \ldots, 100$ given $\lambda_n^{0.95}$ takes 0.8 seconds. To calculate $\lambda_i^{0.95}, \ldots, \lambda_{1000}^{0.95}$ takes roughly 1725 minutes. To construct 95% bands for $n = 1, \ldots, 1000$ given $\lambda_n^{0.95}$ takes 82 seconds.

Since constructing bands given $\lambda_n^{0.95}$ is much faster than searching for $\lambda_n^{0.95}$, it is worth while to tabulate $\lambda_n^{0.95}$. A simple formula is even more convenient than a table of critical values. Berk and Jones (1979) give an asymptotic expression in which the critical log likelihood is a linear combination of $\log(\log(n))$ and $\log(\log(\log(n)))$ for large $n$. This asymptote is approached very slowly. Over the range $1 \leq n \leq 100$ the critical log likelihood, $-n\lambda_n^{0.95}$ is almost linear in $\log(n)$, though it shows a slight convexity which may be approximated by introducing polynomial terms in $\log(n)$. One such approximation is

$$\lambda_n^{0.95} = \frac{1}{n} \left( 3.0123 + 0.4835 \log(n) - 0.00957 \log(n)^2 - 0.001488 \log(n)^3 \right), \quad 1 < n \leq 100. \quad (4.3)$$

The right side of (4.3) provides a formula that gives approximate coverage 95% for $1 < n \leq 100$. The actual coverage level attained through (4.3) may be computed by Noé's algorithm and it lies between 95% and 95.01% for $1 < n \leq 100$. That is, (4.3) is very slightly conservative. Formula (4.3) should not be extrapolated to $n > 100$. The case $n = 1$ is trivial, but for completeness $\lambda_i^{0.95} = -\exp(0.95) \doteq 2.9957$.

For $n$ between 100 and 1000, we may replace (4.3) by

$$\lambda_n^{0.95} = \frac{1}{n} \left( 3.0806 + 0.4894 \log(n) - 0.02086 \log(n)^2 \right), \quad 100 < n \leq 1000. \quad (4.4)$$
The actual coverage obtained through (4.4) lies between 95% and 95.01%. Formula (4.4) should not be used for \( n > 1000 \) or for \( n \leq 100 \).

5. EXAMPLES

Figure 1 shows the empirical distribution function of the velocities of 82 galaxies from the Corona Borealis region. The units are thousands of kilometers per second. This data appears in Table 1 of Roeder (1990). Also included in Figure 1 are 95% confidence bands based on the Kolmogorov-Smirnov test statistic. The width of these bands is obtained by adjusting the value for \( n = 80 \) in Table 2 of Birnbaum (1952); \( P(D_{82} < D_{82}^{95}) = .95 \) where \( D_{82}^{95} \approx 0.1496 \times (80/82)^{1/2} \approx 0.1478 \).

Figure 2 shows the nonparametric likelihood confidence bands described in this paper. They are narrower in the tails and only slightly wider near the center of the data range. This data set has small clusters near each extreme. The result is to accentuate the differences between the two sets of bands in the tails, which is where the important differences take place.

Bands based on Anderson and Darling's (1952) weighted distance (2.3) have not been computed. This method has not received widespread use because the asymptotic distribution of the threshold value is intractable, a complication of \( \psi_{AD} \) blowing up at 0 and 1. Chibisov (1966) has shown that \( n^{1/2} D_{n \psi_{AD}} \), diverges to infinity in distribution. That is the greatest departure, as measured by standard deviations, diverges. It diverges very slowly, and Jaeschke (1979) gives asymptotic approximations which in principle allow one to form asymptotic confidence bands. Unfortunately, the asymptotics take hold very slowly. (Berk and Jones (1979) base their asymptotic approximation mentioned above on that of Jaeschke.) The methods of this paper could be adapted to compute finite sample coverage levels for those bands, but this has not been done in view of the superiority of Berk and Jones's test statistic.

Figures 3 and 4 compare 95% confidence bands for a smaller sample size, \( n = 20 \). They are for a hypothetical sample with equispaced observations. For a real sample, the horizontal axis of the two plots would be shrunk and stretched to match the order statistics but the vertical axis would
be unchanged. By spacing the sample points equally, we get a clearer comparison of the shapes of the bands themselves as distinct from the shape of the data set. Figure 3 uses a critical value of 0.2939 for $D_n$, obtained from Birnbaum (1952, Table 2). (By Noé’s algorithm, the coverage level is 0.949781.)

For both sample sizes, 20 and 82, the nonparametric likelihood bands are narrower in the tails and wider in the middle than the Kolmogorov-Smirnov bands. The tradeoff seems favorable to the likelihood bands: For $n = 82$ the likelihood bands are 38.7% as wide as the Kolmogorov-Smirnov bands at the extreme tails and 113% as wide at the center. For $n = 20$ these ratios are 66% and 101% respectively.

Suppose that $X_1, \ldots, X_{20}$ come from a continuous distribution $F$. The distribution $F^n$ for $n > 1$ is skewed to the right compared to $F$. The comparable left-skewed alternative is $1 - (1 - F)^n$. The distributions $F$ and $F^n$ differ over their whole range, but the greatest difference takes place near the middle of the range, between $F^{-1}(e^{-1}) = F^{-1}(0.37)$ (for small $n$) and $F^{-1}(1-e^{-1}) = F^{-1}(0.63)$ (for large $n$). Using Noé’s algorithm we can calculate the probability that $F^n$ is contained in the confidence bands for both Kolmogorov-Smirnov and nonparametric likelihood bands. Figure 5 plots these probabilities versus $n$. It is clear that even for samples as small as $n = 20$, the nonparametric likelihood test has greater power than the Kolmogorov-Smirnov test, at least for alternatives $F^n$.

6. DISCUSSION

The empirical CDF gives a view of the sample that complements that found in histograms and density estimates. The latter are generally better at showing features such as modes, clusters or the shape of the distribution’s tails. The empirical CDF is better if one is interested in reading quantiles $F^{-1}(u)$ or tail probabilities $F(x), 1 - F(x)$ off the plot. The empirical CDF has two other advantages: it is uniquely defined without having to select a bandwidth or bin size, and non-asymptotic distribution free probability calculations for given values of $n$ are available for it.

The empirical CDF has considerable sampling uncertainty associated with it, especially in small samples. Confidence bands are perhaps the best way to display this variability. Given two
sets of confidence bands, one that has a greater chance of excluding some distribution other than the true distribution is better. The nonparametric likelihood bands studied here have better asymptotic power, in Bahadur's sense, than any weighted Kolmogorov-Smirnov bands at any alternative distribution. Appendix A describes Bahadur's notions of optimality.

This paper focuses on confidence bands constructed by tests. If one is interested in testing per se, and not in confidence bands, there are often better tests available, such as those based on the Anderson-Darling statistic (2.4). See D'Agostino and Stephens (1986). The set of distributions with small values of (2.4) is not plottable as a confidence band. Without getting into too much detail, confidence bands describe "infinite dimensional rectangles" whereas "infinite dimensional ellipsoids" may have better power for testing.

APPENDIX A: OPTIMALITY

This appendix sketches the optimality properties that Berk and Jones obtained for their tests based on $R_n$ and $R_n^\pm$. The story has been simplified, and some readers may wish to consult the references cited here, for a more precise treatment.

Berk and Jones use Bahadur's theory of exact slopes. See Bahadur (1971). Let $H_0$ be a null hypothesis and $H_A$ be an alternative hypothesis. For example $H_0$ could be that $F(0.5) = 0.5$ and $H_A$ could be that $F(0.5) \geq 0.7$. Suppose that in truth $F(0.5) = .75$. Let $P_n$ be the power of a test statistic based on $n$ observations. That is, $P_n$ gives the probability that $H_0$ is rejected in favor of $H_A$. It is common for $P_n$ to approach 1 exponentially fast as $n$ increases, if $H_A$ is true. That is \[ \lim_n n^{-1} \log(1 - P_n) = -\frac{1}{2}c \] for $0 < c < \infty$. The quantity $c$ is Bahadur's exact slope, and it usually depends on which point of $H_A$ obtains. Larger $c$ means greater asymptotic power. Notice that for Bahadur's exact slope, the alternative hypothesis is held fixed as $n$ increases and the power goes to 1. This differs from the usual setting in which the alternative gets closer to the null hypothesis as $n$ increases and the power is held constant. For example one might hold power fixed at 0.8 and find that the corresponding alternative is some multiple of $n^{-1/2}$ away from the null hypothesis, smaller multiples being better.
The tests were derived using the principle of relative optimality in Berk and Jones (1978). They call a combination of test statistics “relatively optimal” if it has exact slope at least as great as any of the test statistics being combined. This has to hold at every alternative. Given a finite set of possibly dependent test statistics, Berk and Jones (1978) consider using the minimum attained $p$ value. This statistic then has to be adjusted to account for the minimization. Berk and Jones find that the resulting statistic is relatively optimal under fairly general conditions. That is, searching over a finite set of test statistics cannot degrade the Bahadur exact slope.

For testing whether a sample comes from $F$, the test statistics are dependent binomial likelihood ratio statistics, and the combination is based on the most extreme of them. More interestingly, the number of these test statistics grows with $n$. The statistic $R_n$ is based on the most significant of $2n$ $p$ values based on the same $n$ data points. One might have expected that the minimum attained $p$ value would have to be adjusted considerably to account for the large number of tests and that as a result using the one most sensitive test would have a big advantage. But Berk and Jones (1979) show that $R_n$ is relatively optimal. That weighted Kolmogorov-Smirnov statistics (both one and two sided) are not relatively optimal, follows from expressions in sections 3 and 5 of Berk and Jones (1979). That is, for some alternatives to $F$, one can have $0 < y < 1$ and a test based on $|F_n(y) - F(y)|$ with greater exact slope than that of $D_n$.

Berk and Jones (1979) show that the exact slope of their nonparametric likelihood ratio statistic is greater than or equal to the exact slope of any weighted Kolmogorov-Smirnov statistic. This means that for any alternative to $F$, the nonparametric likelihood test has asymptotic power at least as great as that of any weighted Kolmogorov-Smirnov test.

The confidence bands based on these tests contain every distribution not rejected by the tests. Therefore for any alternative to $F$, confidence bands derived from the nonparametric likelihood ratio test have at least as great an asymptotic probability of excluding it as do bands based on any weighted Kolmogorov-Smirnov test.
REFERENCES


CAPTIONS

Figure 1: The solid line is the empirical CDF of the velocities of 82 galaxies in the Corona Borealis region. The dotted lines are 95% confidence bands for the underlying distribution function, obtained by inverting the Kolmogorov-Smirnov test statistic.

Figure 2: The solid line is the empirical CDF of the velocities of 82 galaxies in the Corona Borealis region. The dotted lines are 95% confidence bands for the underlying distribution function, obtained by inverting the nonparametric likelihood test statistic.

Figure 3: The solid line is an empirical CDF for 20 data points. The dotted lines are 95% confidence bands obtained by inverting the Kolmogorov-Smirnov test statistic. Equispaced order statistics are used to show the shape of the bands.

Figure 4: The solid line is an empirical CDF for 20 data points. The dotted lines are 95% confidence bands obtained by inverting the nonparametric likelihood test statistic. Equispaced order statistics are used to show the shape of the bands.

Figure 5: The vertical axis is the probability that the alternative distribution \( F^\eta \), is not included in the confidence bands for \( F \) based on a sample of \( n = 20 \) observations from a continuous distribution \( F \). The horizontal axis gives \( \eta \). The curve for the Kolmogorov-Smirnov bands lies above that for the nonparametric likelihood bands indicating that the latter are better at excluding alternatives of the form \( F^\eta \) when \( n = 20 \).
CDF and 95% Kolmogorov-Smirnov Bands
82 Observations

Velocity of Galaxy
CDF and 95\% Nonparametric Likelihood Bands
82 Observations

Velocity of Galaxy
CDF and 95\% Kolmogorov-Smirnov Bands
20 Equispaced Points
CDF and 95% Nonparametric Likelihood Bands
20 Equispaced Points

Equispaced Data
Probability That $F^\eta$ is Included in 95\% K-S and N-P L Bands (n=20)