The Annals of Statistics
1974, Vol. 2, No. 4, 795-802

THE ASYMPTOTIC SUFFICIENCY OF A RELATIVELY SMALL NUMBER OF ORDER STATISTICS IN TESTS OF FIT[1]

BY LIONEL WEISS

Cornell University

For each n, $X_n(1), \cdots, X_n(n)$ are independent and identically distributed continuous random variables over $(0, 1)$, with common density function equal to $1 + r(x)/n^{\frac{1}{2}}$, $r(x)$ unknown but satisfying certain regularity conditions. The problem is to test the hypothesis that $r(x) = 0$ for all x in $(0, 1)$. $Y_n(1) < \cdots < Y_n(n)$ are the ordered values of $X_n(1), \cdots, X_n(n)$. δ is a fixed value in the open interval $(\frac{3}{4}, 1)$. It is shown that $Y_n([n^\delta]), Y_n(2[n^\delta]), \cdots$ are asymptotically sufficient, and can be assumed to have a joint normal distribution for all asymptotic purposes. Using these facts, a test of the hypothesis is constructed with a good asymptotic power curve.

1. Introduction. For each positive integer n, $X_n(1), \cdots, X_n(n)$ are independent and identically distributed random variables, with an unknown continuous cumulative distribution function $F_n(x)$. The problem is to test the hypothesis that $F_n(x) = G(x)$, where $G(x)$ is a completely specified continuous cumulative distribution function. By replacing $X_n(1), \cdots, X_n(n)$ by $G(X_n(1)), \cdots, G(X_n(n))$ respectively, we transform to independent and identically distributed random variables over the interval $(0, 1)$, and the hypothesis is that the common distribution is the uniform distribution over $(0, 1)$. From now on we assume that this has been done.

We are usually interested in the asymptotic power against alternatives which approach the hypothesis at a rate just rapid enough to keep the asymptotic power in the open interval $(\alpha, 1)$, where α is the asymptotic level of significance. In the present case, letting $f_n(x)$ denote the probability density function for $X_n(i)$, we will be interested in alternatives such that $n^{\frac{1}{2}} \max_{0 \leq x \leq 1} |f_n(x) - 1|$ remains positive and bounded as n increases.

Let $Y_n(1), \cdots, Y_n(n)$ denote the ordered values of $X_n(1), \cdots, X_n(n)$, in increasing order. Since we are assuming that $X_n(i)$ is a continuous random variable over $(0, 1)$, the inequalities $0 < Y_n(1) < \cdots < Y_n(n) < 1$ hold with probability one. For typographical simplicity, if γ is any positive value, n^γ is to be understood to mean the largest integer no greater than n^γ; and if γ is a negative value, n^γ is to be written as $1/n^{-\gamma}$, and $n^{-\gamma}$ is to be understood as the largest integer no greater than $n^{-\gamma}$.

Let ε, δ be fixed values satisfying the conditions $0 < \varepsilon < \frac{1}{2}$, $\frac{3}{4} < \delta < 1$, $3\delta < 2 + \varepsilon$, $2\delta < 1 + 2\varepsilon$. For example, δ could be slightly above $\frac{3}{4}$ and ε slightly

Received December 1970; revised June 1970.

[1] Research supported by National Science Foundation Grants GP-31430X, GP-21184.

AMS 1970 *subject classifications.*

Key words and phrases. Tests of fit, order statistics, asymptotic efficiency.

below $\frac{1}{2}$. Let $k(n)$ denote the largest integer such that $k(n)n^\delta < n$. In Section 2 we show that if

$$\lim_{n\to\infty} n^\varepsilon \max_{0\le x\le 1} |f_n(x) - 1| = 0 \quad \text{and}$$

$$\lim_{n\to\infty} n^\varepsilon \sup_{0<x<1} \left|\frac{d^r}{dx^r} f_n(x)\right| = 0 \qquad \text{for } r = 1, 2, 3,$$

then $\{Y_n(n^\delta), Y_n(2n^\delta), \cdots, Y_n(k(n)n^\delta)\}$ are asymptotically sufficient for all purposes of statistical inference. In Section 3 we show that for all asymptotic purposes we can assume that $\{Y_n(n^\delta), Y_n(2n^\delta), \cdots, Y_n(k(n)n^\delta)\}$ have a joint normal distribution. In Section 4 we use these facts to construct a test based on a quadratic function of $\{Y_n(n^\delta), Y_n(2n^\delta), \cdots, Y_n(k(n)n^\delta)\}$ which has good asymptotic power.

2. The asymptotic sufficiency of $Y_n(n^\delta), \cdots, Y_n(k(n)n^\delta)$. We assume $f_n(x)$ satisfies the assumptions of the last paragraph of Section 1. The joint probability density function for $\{Y_n(1), \cdots, Y_n(n)\}$ is $n! \prod_{i=1}^n f_n(y_n(i))$ if $0 < y_n(1) < \cdots < y_n(n) < 1$ and is zero otherwise. We denote this joint density function by $h_n(y_n(1), \cdots, y_n(n))$, and the corresponding probability measure by H_n.

Now we construct a different joint probability density function for $\{Y_n(1), \cdots, Y_n(n)\}$. This second density function is constructed by assuming that the joint marginal density for $\{Y_n(n^\delta), Y_n(2n^\delta), \cdots, Y_n(k(n)n^\delta)\}$ is what would be given by $h_n(y_n(1), \cdots, y_n(n))$, and that the joint conditional distribution of the other $\{Y_n(i)\}$ is given as follows. The $n^\delta - 1$ random variables in the open interval $(0, Y_n(n^\delta))$ are distributed as the ordered values of $n^\delta - 1$ independent and identically distributed random variables, each with a uniform distribution over $(0, Y_n(n^\delta))$. For $j = 1, \cdots, k(n) - 1$, the $n^\delta - 1$ random variables in the open interval $(Y_n(jn^\delta), Y_n((j+1)n^\delta))$ are distributed as the ordered values of $n^\delta - 1$ independent and identically distributed random variables, each with a uniform distribution over $(Y_n(jn^\delta), Y_n((j+1)n^\delta))$. The $n - k(n)n^\delta$ random variables in the open interval $(Y_n(k(n)n^\delta), 1)$ are distributed as the ordered values of $n - k(n)n^\delta$ independent and identically distributed random variables, each with a uniform distribution over $(Y_n(k(n)n^\delta), 1)$. We denote the resulting joint probability density function for $\{Y_n(1), \cdots, Y_n(n)\}$ by $g_n(y_n(1), \cdots, y_n(n))$, and the corresponding probability measure by G_n.

We now show that if $g_n(y_n(1), \cdots, y_n(n))$ is actually the joint density for $\{Y_n(1), \cdots, Y_n(n)\}$, then $\log[h_n(Y_n(1), \cdots, Y_n(n))/g_n(Y_n(1), \cdots, Y_n(n))]$ converges stochastically to zero as n increases. For typographical simplicity, let $m_n(j)$ denote $n^\delta - 1$ for $j = 1, \cdots, k(n)$, and let $m_n(k(n)+1)$ denote $n - k(n)n^\delta$. Also, let $Y_n(0n^\delta)$ denote zero, and $Y_n((k(n)+1)n^\delta)$ denote unity, Then $F_n(Y_n(0n^\delta)) = 0$ and $F_n(Y_n((k(n)+1)n^\delta)) = 1$. Outside the region where both $h_n(Y_n(1), \cdots, Y_n(n))$ and $g_n(Y_n(1), \cdots, Y_n(n))$ are zero, $\log[h_n(Y_n(1), \cdots, Y_n(n))/g_n(Y_n(1), \cdots, Y_n(n))]$ is equal to

$$(2.1) \qquad \sum_{j=1}^{k(n)+1} \sum_{i=1}^{m_n(j)} \log\left[\frac{\{Y_n(jn^\delta) - Y_n((j-1)n^\delta)\} f_n(Y_n((j-1)n^\delta + i))}{F_n(Y_n(jn^\delta)) - F_n(Y_n((j-1)n^\delta))}\right].$$

For $j = 1, \cdots, k(n) + 1$, define $D_n(j-1)$ by the equation $F_n(Y_n(j-1)n^\delta) = n^{-1}(j-1)n^\delta + n^{-\frac{1}{2}}D_n(j-1)$. By the Kolmogorov-Smirnov theorem on the deviation between the empirical and cumulative distribution functions, and the definition of $Y_n(0n^\delta)$, it follows that $\max_{1 \leq j \leq k(n)+1} |D_n(j-1)|$ is bounded with probability one. Writing $Y_n((j-1)n^\delta) = F_n^{-1}(n^{-1}(j-1)n^\delta + n^{-\frac{1}{2}}D_n(j-1))$, expanding $F_n^{-1}(\ )$ around $n^{-1}(j-1)n^\delta$, and using the easily verified fact that $\lim_{n\to\infty} n^\varepsilon \max_{0 \leq u \leq 1} |F_n^{-1}(u) - u| = 0$, we find that with probability approaching one as n increases,

$$|Y_n(jn^\delta) - Y_n((j-1)n^\delta)| < \frac{2n^\delta}{n} \quad \text{for } j = 1, \cdots, k(n)+1\,, \tag{2.2}$$

where the result for $j = k(n) + 1$ comes directly from the definitions of $k(n)$ and $Y_n((k(n)+1)n^\delta)$. Define $\bar{Y}_n(j)$ as $\frac{1}{2}[Y_n(jn^\delta) + Y_n((j-1)n^\delta)]$ for $j = 1, \cdots, k(n)+1$, and $U_n'(i, j)$ as $[Y_n((j-1)n^\delta + i) - \bar{Y}_n(j)]/[Y_n(jn^\delta) - Y_n((j-1)n^\delta)]$. The joint conditional distribution of $\{U_n'(1, j), \cdots, U_n'(m_n(j), j)\}$ given $Y_n((j-1)n^\delta)$ and $Y_n(jn^\delta)$ is that of $m_n(j)$ ordered uniform variables over $(-\frac{1}{2}, \frac{1}{2})$. Let $\{U_n(i, j): i = 1, \cdots, m_n(j); j = 1, \cdots, k(n)+1\}$ denote random variables whose joint conditional distribution given $\{Y_n(n^\delta), \cdots, Y_n(k(n)n^\delta)\}$ is that of independent random variables, each uniform over $(-\frac{1}{2}, \frac{1}{2})$. Then (2.1) has exactly the same distribution as

$$\begin{aligned} \sum_{j=1}^{k(n)+1} \sum_{i=1}^{m_n(j)} \log [(\{Y_n(jn^\delta) - Y_n((j-1)n^\delta)\} \\ \times f_n(\bar{Y}_n(j) + [\bar{Y}_n(jn^\delta) - Y_n((j-1)n^\delta)]U_n(i, j))) \\ \div (F_n(Y_n(jn^\delta)) - F_n(Y_n((j-1)n^\delta)))]\,. \end{aligned} \tag{2.3}$$

Write

$$\begin{aligned} F_n(Y_n(jn^\delta)) = F_n(\bar{Y}_n(j)) &+ (Y_n(jn^\delta) - \bar{Y}_n(j))f_n(\bar{Y}_n(j)) \\ &+ \tfrac{1}{2}[Y_n(jn^\delta) - \bar{Y}_n(j)]^2 f_n'(\bar{Y}_n(j)) \\ &+ \tfrac{1}{6}[Y_n(jn^\delta) - \bar{Y}_n(j)]^3 f_n''(\bar{Y}_n(j)) \\ &+ \tfrac{1}{24}[Y_n(jn^\delta) - \bar{Y}_n(j)]^4 f_n'''(\theta_n(j))\,, \end{aligned}$$

$$\begin{aligned} f_n(\bar{Y}_n(j) &+ [Y_n(jn^\delta) - Y_n((j-1)n^\delta)]U_n(i, j)) \\ &= f_n(\bar{Y}_n(j)) + [Y_n(jn^\delta) - Y_n((j-1)n^\delta)]U_n(i, j)f_n'(\bar{Y}_n(j)) \\ &\quad + \tfrac{1}{2}[Y_n(jn^\delta) - Y_n((j-1)n^\delta)]^2 U_n^2(i, j)f_n''(\bar{Y}_n(j)) \\ &\quad + \tfrac{1}{6}[Y_n(jn^\delta) - Y_n((j-1)n^\delta)]^3 U_n^3(i, j)f_n'''(\bar{\theta}_n(i, j))\,, \end{aligned}$$

where $\theta_n(j)$, $\bar{\theta}_n(i, j)$ are in (0, 1). Substituting these expansions into (2.3) and simplifying, we find that (2.3) can be written as

$$\begin{aligned} \sum_{j=1}^{k(n)+1} \sum_{i=1}^{m_n(j)} \log\Big[1 &+ [Y_n(jn^\delta) - Y_n((j-1)n^\delta)] \frac{f_n'(\bar{Y}_n(j))}{f_n(\bar{Y}_n(j))} U_n(i, j) \\ &+ \tfrac{1}{2}[Y_n(jn^\delta) - Y_n((j-1)n^\delta)]^2 \frac{f_n''(\bar{Y}_n(j))}{f_n(\bar{Y}_n(j))} U_n^2(i, j) \\ &+ \frac{\gamma_n(i, j)}{n^\varepsilon}[Y_n(jn^\delta) - Y_n((j-1)n^\delta)]^3\Big] \end{aligned} \tag{2.4}$$

$$- \sum_{j=1}^{k(n)+1} \sum_{i=1}^{m_n(j)} \log\Big[1 + \tfrac{1}{24}[Y_n(jn^\delta) - Y_n((j-1)n^\delta)]^2 \frac{f_n''(\bar{Y}_n(j))}{f_n(\bar{Y}_n(j))} + \frac{\bar{\gamma}_n(j)}{n^\varepsilon}[Y_n(jn^\delta) - Y_n((j-1)n^\delta)]^3\Big]$$

where $\max_{i,j} |\gamma_n(i, j)|$ and $\max_j |\bar{\gamma}_n(j)|$ converge stochastically to zero as n increases. Expanding the logarithms in (2.4), we find that (2.4) can be written as the sum of the following three expressions:

$$\sum_{j=1}^{k(n)+1} \sum_{i=1}^{m_n(j)} \frac{f_n'(\bar{Y}_n(j))}{f_n(\bar{Y}_n(j))}[Y_n(jn^\delta) - Y_n((j-1)n^\delta)]U_n(i, j), \tag{2.5}$$

$$\tfrac{1}{2} \sum_{j=1}^{k(n)+1} \sum_{i=1}^{m_n(j)} \frac{f_n''(\bar{Y}_n(j))}{f_n(\bar{Y}_n(j))}[Y_n(jn^\delta) - Y_n((j-1)n^\delta)]^2(U_n^2(i, j) - \tfrac{1}{12}), \tag{2.6}$$

$$\begin{aligned}
&\sum_{j=1}^{k(n)+1} \sum_{i=1}^{m_n(j)} \left(\frac{\gamma_n(i, j) - \bar{\gamma}_n(j)}{n^\varepsilon}\right)(Y_n(jn^\delta) - Y_n((j-1)n^\delta))^3 \\
&\quad - \tfrac{1}{2} \sum_{j=1}^{k(n)+1} \sum_{i=1}^{m_n(j)} \frac{1}{(1 + \bar{\beta}_n(i, j))^2} \\
&\qquad \times \Big[[Y_n(jn^\delta) - Y_n((j-1)n^\delta)]\frac{f_n'(\bar{Y}_n(j))}{f_n(\bar{Y}_n(j))} U_n(i, j) \\
&\qquad + \tfrac{1}{2}[Y_n(jn^\delta) - Y_n((j-1)n^\delta)]^2 \frac{f_n''(\bar{Y}_n(j))}{f_n(\bar{Y}_n(j))} U_n^2(i, j) \\
&\qquad + \frac{\gamma_n(i, j)}{n^\varepsilon}[Y_n(jn^\delta) - Y_n(j-1)n^\delta)]^3\Big]^2 \\
&\quad + \tfrac{1}{2} \sum_{j=1}^{k(n)+1} \sum_{i=1}^{m_n(j)} \frac{1}{(1 + \beta_n(j))^2} \\
&\qquad \times \Big[\tfrac{1}{24}[Y_n(jn^\delta) - Y_n(j-1)n^\delta)]^2 \frac{f_n''(\bar{Y}_n(j))}{f_n(\bar{Y}_n(j))} \\
&\qquad + \frac{\bar{\gamma}_n(j)}{n^\varepsilon}[Y_n(jn^\delta) - Y_n((j-1)n^\delta)]^3\Big]^2,
\end{aligned} \tag{2.7}$$

where $\max_{i,j} |\bar{\beta}_n(i, j)|$ and $\max_j |\beta_n(j)|$ converge stochastically to zero as n increases. Using (2.2) and the assumptions about $f_n(x)$, ε, and δ, it is easily seen that (2.7) converges stochastically to zero as n increases. The conditional mean and variance (given $\{Y_n(n^\delta), \cdots, Y_n(k(n)n^\delta)\}$) of (2.5) are 0, $\frac{1}{12} \sum_{j=1}^{k(n)+1} m_n(j) \times [f_n'(\bar{Y}_n(j))/f_n(\bar{Y}_n(j))]^2[Y_n(jn^\delta) - Y_n((j-1)n^\delta)]^2$. Using (2.2) and the assumptions about $f_n(x)$, ε, and δ, it is easily shown that this conditional variance converges stochastically to zero as n increases, and this clearly implies that (2.5) converges stochastically to zero as n increases. A similar argument shows that (2.6) converges stochastically to zero as n increases. This completes the proof that $\log[h_n(Y_n(1), \cdots, Y_n(n))/g_n(Y_n(1), \cdots, Y_n(n))]$ converges stochastically to zero as n increases.

It now follows from the argument on pages 261–262 of [3] that if B_n is any measurable region in $(y(1), \cdots, y(n))$-space, then

$$\lim_{n\to\infty} |G_n(B_n) - H_n(B_n)| = 0 .$$

Suppose, for each n, that $X_n(1), \cdots, X_n(n)$ are independent and identically distributed random variables, each with density function $f_n(x)$, where the sequence $\{f_n(x)\}$ satisfies the assumptions of Section 1. Let $Y_n(1) < \cdots < Y_n(n)$ denote the ordered values of $X_n(1), \cdots, X_n(n)$. Suppose there are two statisticians, named A and B respectively. Neither one knows $f_n(x)$, except that each knows that the sequence $\{f_n(x)\}$ satisfies the assumptions of Section 1. Statistician A knows the values of $Y_n(1), \cdots, Y_n(n)$, which have joint density function $h_n(y_n(1), \cdots, y_n(n))$. Statistician B knows only the values of $Y_n(n^\delta), \cdots, Y_n(k(n)n^\delta)$. Even though he does not know $f_n(x)$, by the use of a table of random numbers statistician B is able to generate $n - k(n)$ additional random variables so that the joint density function of his full set of n random variables is $g_n(y_n(1), \cdots, y_n(n))$. Statistician B is able to do this because under g_n the joint conditional distribution of all $\{Y_n(i)\}$ not among $\{Y_n(n^\delta), \cdots, Y_n(k(n)n^\delta)\}$ does not depend on $f_n(x)$, but only on $\{Y_n(n^\delta), \cdots, Y_n(k(n)n^\delta)\}$.

Suppose statistician A uses a region $A(n)$ in $(Y_n(1), \cdots, Y_n(n))$-space for statistical inference. Knowing only $\{Y_n(n_\delta), \cdots, Y_n(k(n)n_\delta)\}$, statistician B can construct a region with the same asymptotic probability as the region $A(n)$. In this sense, $\{Y_n(n^\delta), \cdots, Y_n(k(n)n^\delta)\}$ are asymptotically sufficient. Before applying this asymptotic sufficiency, in the next section we show that for all asymptotic purposes, $\{Y_n(n^\delta), \cdots, Y_n(k(n)n^\delta)\}$ can be considered as jointly normally distributed.

3. The asymptotic normality of $Y_n(n^\delta), \cdots, Y_n(k(n)n^\delta)$. For $j = 1, \cdots, k(n)$, define $Z_n(j)$ as $n^{\frac{1}{2}}(Y_n(jn^\delta) - F_n^{-1}(jn^\delta/n))$. Let $r_n(z_n(1), \cdots, z_n(k(n))$ denote the joint probability function for $Z_n(1), \cdots, Z_n(k(n))$ and let R_n denote the corresponding probability measure. Let $\bar{s}_n(z(1), \cdots, z(k(n))$ denote the following $k(n)$-variate normal probability density function:

$$\left(\frac{1}{2\pi}\right)^{k(n)/2} \left(\frac{n}{n^\delta}\right)^{\frac{1}{2}} \left(\frac{n(n^\delta - 1)}{n^{2\delta}}\right)^{k(n)/2} \times \exp\left[-\frac{n(n^\delta - 1)}{2n^{2\delta}}(z^2(1) + z^2(k(n)) + \textstyle\sum_{j=2}^{k(n)} (z(j) - z(j-1))^2)\right].$$

Again let $\bar{S}_n$ denote the corresponding probability measure.

Let $s_n(z(1), \cdots, z(k(n))$ denote the $k(n)$-variate normal density function defined by the equation

$$s_n(z(1), \cdots, z(k(n)) = \bar{s}_n\left(f_n\left(F_n^{-1}\left(\frac{n^\delta}{n}\right)\right) z(1), \cdots, f_n\left(F_n^{-1}\left(\frac{k(n)n^\delta}{n}\right)\right) z(k(n))\right) \textstyle\prod_{j=1}^{k(n)} f_n\left(F_n^{-1}\left(\frac{jn^\delta}{n}\right)\right),$$

and denote by S_n the corresponding probability measure.

In this section we show that if B_n is any measurable region in $(z(1), \cdots, z(k(n)))$-space, then

$$\lim_{n\to\infty} |R_n(B_n) - \bar{S}_n(B_n)| = 0 . \tag{3.1}$$

In [3] it was shown that

$$\lim_{n\to\infty} |R_n(B_n) - S_n(B_n)| = 0 ,$$

so (3.1) will be proved if we can show

$$\lim_{n\to\infty} |S_n(B_n) - \bar{S}_n(B_n)| = 0 . \tag{3.2}$$

Using the argument on pages 261–262 of [3], (3.2) will be proved if we can show that if $\bar{s}_n(z_n(1), \cdots, z_n(k(n)))$ is the actual joint density for $Z_n(1), \cdots, Z_n(k(n))$, then $\log [s_n(Z_n(1), \cdots, Z_n(k(n)))/\bar{s}_n(Z_n(1), \cdots, Z_n(k(n)))]$ converges stochastically to zero as n increases. But this last statement is very simply proved, using the easily verified fact that under $\bar{s}_n$, the covariance between $Z_i(n)$ and $Z_j(n)$, where $i \leqq j$, is $n^{3\delta}n^{-2}(n^\delta - 1)^{-1}i(nn^{-\delta} - j)$. Thus (3.1) is true.

In concluding this section, we note that if $\bar{s}_n$ is the joint density for $Z_n(1), \cdots, Z_n(k(n))$, and if we define

$$W_n(1) = \left(\frac{n(n^\delta - 1)}{n^{2\delta}}\right)^{\frac{1}{2}}\left[Z_n(1) - \left(\frac{1 + (1 + k(n))^{\frac{1}{2}}}{k(n)}\right) Z_n(k(n))\right]$$

$$W_n(j) = \left(\frac{n(n^\delta - 1)}{n^{2\delta}}\right)^{\frac{1}{2}}\left[Z_n(j) - Z_n(j-1) - \left(\frac{1 + (1 + k(n))^{\frac{1}{2}}}{k(n)}\right)Z_n(k(n))\right]$$

$$\text{for } j = 2, \cdots, k(n) ,$$

then $W_n(1), \cdots, W_n(k(n))$ are independent standard normal random variables.

4. Application to tests of fit. Throughout this section, we assume that $X_n(1), \cdots, X_n(k(n))$ are independent and identically distributed random variables, with common density $f_n(x) = 1 + n^{-\frac{1}{2}}r(x)$, where $\int_0^1 r(x)\,dx = 0$, and $\sup_{0<x\leqq 1} |r'''(x)|$ is bounded. Then $f_n(x)$ satisfies all the assumptions of Section 1, and $Y_n(n^\delta), \cdots, Y_n(k(n)n^\delta)$ are asymptotically sufficient. Define $Z_n'(j)$ as $n^{\frac{1}{2}}(Y_n(jn^\delta) - jn^\delta/n)$, and define $W_n'(j)$ to be the same function of $\{Z_n'(1), \cdots, Z_n'(k(n))\}$ as $W_n(j)$ is of $\{Z_n(1), \cdots, Z_n(k(n))\}$. We note that $\{W_n'(1), \cdots, W_n'(k(n))\}$ are observable, even if $r(x)$ is unknown, and there is a one-one correspondence between $\{Y_n(n^\delta), \cdots, Y_n(k(n)n^\delta)$ and $W_n'(1), \cdots, W_n'(k(n))\}$. Thus $\{W_n'(1), \cdots, W_n'(k(n))\}$ are asymptotically sufficient, and by the last paragraph of Section 3, for all asymptotic purposes we can assume that the joint distribution for $\{W_n'(1), \cdots, W_n'(k(n))\}$ is that of independent normal random variables, each with variance one, and with $E\{W_n'(j)\}$ given by

$$-n^{\frac{1}{2}}\left(\frac{n(n^\delta - 1)}{n^{2\delta}}\right)^{\frac{1}{2}}\left[\frac{n^\delta}{n} - F_n^{-1}\left(\frac{jn^\delta}{n}\right) + F_n^{-1}\left(\frac{(j-1)n^\delta}{n}\right)\right.$$
$$\left. - \left(\frac{1 + (1 + k(n))^{\frac{1}{2}}}{k(n)}\right)\left[\frac{k(n)n^\delta}{n} - F_n^{-1}\left(\frac{k(n)n^\delta}{n}\right)\right]\right].$$

Denote the resulting joint density by $\bar{t}_n(w_n'(1), \cdots, w_n'(k(n)))$, and the probability measure by $\bar{T}_n$. Denote by $t_n(w_n'(1), \cdots, w_n'(k(n)))$ the joint density given by assuming $W_n'(1), \cdots, W_n'(k(n))$ are independent normal random variables, each with variance one and with $E\{W_n'(j)\} = -(n^\delta/n)^{\frac{1}{2}}r(jn^\delta/n)$. Again denote the corresponding probability measure by T_n. It is easily shown that if B_n is any measurable region in $(w_n'(1), \cdots, w_n'(k(n))$-space, then

$$\lim_{n\to\infty} |\bar{T}_n(B_n) - T_n(B_n)| = 0 .$$

For the rest of this paper we make the following "Normality Assumption": for each n, $W_n'(1), \cdots, W_n'(k(n))$ are independent normal random variables, each with variance one, and $E\{W_n'(j)\} = -(n^\delta/n)^{\frac{1}{2}}r(jn^\delta/n)$. Of course, this Normality Assumption is not true, but the results of the preceding paragraph show that this assumption gives the correct results for all asymptotic purposes, and we are only interested in asymptotic theory in this paper. Making the assumption avoids some circumlocutions below.

Since $\{W_n'(1), \cdots, W_n'(k(n))\}$ are asymptotically sufficient, there must be a test of the hypothesis $r(x) = 0$ which is based on $\{W_n'(1), \cdots, W_n'(k(n))\}$ and has asymptotic power against all $r(x)$ which is at least as good as the asymptotic power of any given test. For example, a reasonable conjecture would be that the test which rejects the hypothesis if $\max_{1\leq i\leq k(n)} |W_n'(1) + \cdots + W_n'(i)| > c_n(\alpha)$, where $c_n(\alpha)$ is chosen to give the desired level of significance α, is asymptotically at least as good as the familiar Kolmogorov–Smirnov test. Even if we knew the asymptotic power of the proposed test, however, the asymptotic power of the Kolmogorov–Smirnov test is not known in any form that would allow comparison of the asymptotic powers. See [1] for a discussion of the computation of the asymptotic power of the Kolmogorov–Smirnov test.

Let $\phi(t)$ denote $\int_t^\infty (2\pi)^{-\frac{1}{2}}e^{-y^2/2}\,dy$, and for any α in $(0, 1)$, let z_α be defined by $\phi(z_\alpha) = \alpha$. It is easily shown that if we test the hypothesis that $r(x) = 0$ against one specific alternative, say $\bar{r}(x)$, the maximum possible asymptotic power of a test of level of significance α is $\phi(z_\alpha - (\int_0^1 \bar{r}^2(x)\,dx)^{\frac{1}{2}})$. This means that a "natural" measure of distance between the uniform density over $(0, 1)$ and the density $f_n(x) = 1 + n^{-\frac{1}{2}}r(x)$ is $\int_0^1 r^2(x)\,dx$. The rest of this paper is devoted to investigating tests with good asymptotic power with respect to this distance.

Expand $r(x)$ in a Fourier cosine series over $(0, 1)$: $r(x) = \sum_{j=1}^\infty A_j 2^{\frac{1}{2}} \cos j\pi x$. There is no constant term, because $\int_0^1 r(x)\,dx = 0$. $\int_0^1 r^2(x)\,dx = \sum_{j=1}^\infty A_j^2$.

Temporarily, we limit the class of alternatives by assuming that $A_j = 0$ for all $j > T$, where T is a fixed and known positive integer. Define $S_n(j)$ as $(n^\delta/n)^{\frac{1}{2}} \sum_{i=1}^{k(n)} W_n'(i)2^{\frac{1}{2}} \cos j\pi(n^{-1}in^\delta)$. Under our Normality Assumption, $\{S_n(1), \cdots, S_n(T)\}$ have a joint normal distribution, with covariance matrix approaching the identity matrix as n increases, and $E\{S_n(j)\}$ approaching $-A_j$ as n increases. Using an a priori distribution which assigns probability b to the point $A_1 = A_2 = \cdots = A_T = 0$, and probability $1 - b$ spread uniformly over $A_1^2 + \cdots + A_T^2 = \bar{c} > 0$, it is easily shown that the following test maximizes the minimum asymptotic power against alternatives $A_1^2 + \cdots + A_T^2 \geqq \bar{c}$, among all tests with asymptotic level of significance α: Reject if $\sum_{j=1}^{T} S_n^2(j) > c(\alpha; T)$ where $c(\alpha; T)$ is chosen to give asymptotic level of significance α. The asymptotic distribution of $\sum_{j=1}^{T} S_n^2(j)$ is noncentral chi-square, with T degrees of freedom and noncentrality parameter $\sum_{j=1}^{T} A_j^2$, so $c(\alpha; T)$ is found from the central chi-square table. It is easily verified that for $\sum_{j=1}^{T} A_j^2$ fixed, the asymptotic power of the test approaches α as T increases. This means that there is no test procedure which has power staying above α against all alternatives $r(x)$ subject to the sole restriction $\int_0^1 r^2(x)\,dx \geqq \bar{c} > 0$.

In order to keep the asymptotic power above the asymptotic level of sigificance, we must limit the class of alternatives in some way. One reasonable way is to assume that if the null hypothesis $\sum_{j=1}^{\infty} A_j^2 = 0$ is not true, then $\int_0^1 r^2(x)\,dx \equiv \sum_{j=1}^{\infty} A_j^2 \geqq c_1$ and $\sum_{j=1}^{\infty} (jA_j)^2 \leqq c_2$, where c_1, c_2 are given positive values with $c_1 < c_2$. (We note that formally, $r'(x) = -\pi \sum_{j=1}^{\infty} (jA_j)2^{\frac{1}{2}} \sin j\pi x$, which motivates the restriction on $\sum_{j=1}^{\infty} (jA_j)^2$.) If we use $\sum_{j=1}^{T} S_n^2(j)$ as our test statistic, the minimum possible noncentrality parameter when the null hypothesis is not true is equal to 0 if $c_2 \geqq (T+1)^2 c_1$, and is equal to $[c_1(T+1)^2 - c_2]/[(T+1)^2 - 1]$ if $c_2 \leqq c_1(T+1)^2$. Since we are given c_1 and c_2, we can use a table of noncentral chi-square to choose the value of T to maximize $P[\sum_{j=1}^{T} S_n^2(j) > c(\alpha; T)]$, assuming the worst possible noncentrality parameter (that is, the smallest possible noncentrality parameter computed in the preceding sentence). This test is similar to Neyman's "smooth" test of fit [2].

REFERENCES

[1] DURBIN, J. (1971). Boundary-crossing probabilities for the Brownian motion and Poisson processes and techniques for computing the power of the Kolmogorov-Smirnov test. *J. Appl. Probability* **8** 431-453.

[2] NEYMAN, J. (1937). Smooth test for goodness of fit. *Skand. Aktuarietidskr.* **20** 149-199.

[3] WEISS, L. (1969). The asymptotic joint distribution of an increasing number of sample quantiles. *Ann. Inst. Statist. Math.* **21** 257-263.

DEPARTMENT OF OPERATIONS RESEARCH
CORNELL UNIVERSITY
ITHACA, NEW YORK 14850