THE ASYMPTOTIC POWER OF CERTAIN TESTS OF FIT BASED ON SAMPLE SPACINGS[1]

By Lionel Weiss

Cornell University

1. Introduction and summary. Suppose $X_1, X_2, \cdots, X_n$ are independent and identically distributed chance variables, each with density $f(x)$, where $\int_0^1 f(x)\,dx = 1$, $f(x)$ has a finite number of discontinuities, and there are two constants $A, B (0 < A < B < \infty)$ such that $A \leqq f(x) \leqq B$ for all x in $[0, 1]$.

Let Y_0 denote zero, Y_{n+1} denote unity, and let $Y_1 \leqq Y_2 \leqq \cdots \leqq Y_n$ be the ordered values of $X_1, X_2, \cdots, X_n$. Define T_i as $Y_i - Y_{i-1}$ for $i = 1, \cdots, n+1$. Let r be any positive number greater than unity, and let $V(n)$ denote $\sum_{i=1}^{n+1} T_i^r$. The following theorem was proved in [1].

Theorem A. *If $f(x) = 1$ for x in $[0, 1]$, then the distribution of*

$$\frac{n^{r-1/2}V(n) - \sqrt{n}\Gamma(r+1)}{\sqrt{\Gamma(2r+1) - (r^2+1)[\Gamma(r+1)]^2}}$$

approaches the standard normal distribution as n increases. In the present paper, we prove the following generalization of Theorem A:

Theorem 1: *The distribution of*

$$\frac{n^{r-1/2}V(n) - \sqrt{n}\Gamma(r+1)\int_0^1 f^{1-r}(x)\,dx}{\sqrt{[\Gamma(2r+1) - 2r\Gamma^2(r+1)]\int_0^1 f^{1-2r}(x)\,dx - \left[(r-1)\Gamma(r+1)\int_0^1 f^{1-r}(x)\,dx\right]^2}}$$

approaches the standard normal distribution as n increases.

Theorem 1 can be used to compute the asymptotic power of certain tests of fit based on $V(n)$.

2. Proof of Theorem 1 when $f(x)$ is a step function. First we prove Theorem 1 for the case when there are H subintervals $I_1, \cdots, I_H$, $I_1 = [0, z_1)$, $I_2 = [z_1, z_2), \cdots, I_H = [z_{H-1}, 1]$, so that on I_i, $f(x) = a_i$, where $0 < A \leqq a_i \leqq B$. Let N_i denote the number of the values $X_1, \cdots, X_n$ which fall in the interval I_i, and let ${}_iY_1 \leqq {}_iY_2 \leqq \cdots \leqq {}_iY_{N_i}$ be the ordered values of these values in I_i. Denote z_{i-1} by ${}_iY_0$, and z_i by ${}_iY_{N_i+1}$. z_0 is to denote zero, z_H denotes unity. Define ${}_iT_j$ as ${}_iY_j - {}_iY_{j-1}$, for $j = 1, \cdots, N_i + 1$. Define V_i as $\sum_{j=1}^{N_i+1} {}_iT_j^r$. From Theorem A quoted above and from an examination of the conditional distribution of ${}_iY_1, \cdots, {}_iY_{N_i}$ given N_i, it follows that the conditional distribution given N_i of

$$Q_i = \frac{\dfrac{N_i^{r-1/2}V_i}{(z_i - z_{i-1})^r} - \sqrt{N_i}\Gamma(r+1)}{\sqrt{\Gamma(2r+1) - (r^2+1)\Gamma^2(r+1)}}$$

Received July 23, 1956; revised July 26, 1956.

[1] Research under contract with the Office of Naval Research. It may be reproduced in whole or in part for any purpose of the United States Government.

approaches the standard normal distribution as N_i increases. Also, the conditional distribution of ${}_iY_1, \cdots, {}_iY_{N_i}$ given $N_1, \cdots, N_H$ depends only on N_i, while the joint distribution of $N_1, \cdots, N_H$ is multinomial with parameters $n, a_1(z_1 - z_0), \cdots, a_H(z_H - z_{H-1})$. From these facts, it follows that the joint distribution of

$$\left\{\sqrt{n}\left(\frac{N_1}{n} - a_1(z_1 - z_0)\right), \cdots, \sqrt{n}\left(\frac{N_{H-1}}{n} - a_{H-1}(z_{H-1} - z_{H-2})\right), Q_1, \cdots, Q_H\right\}$$

approaches the joint distribution of $\{S_1, \cdots, S_{H-1}, T_1, \cdots, T_H\}$ as n increases, where this last set of chance variables has a joint normal distribution with zero means and covariance matrix $\| a_{ij} \|$ $(i, j = 1, \cdots, 2H - 1)$, where $a_{ii} = 1$ if $i \geqq H$, $a_{ij} = 0$ if i and/or j is $\geqq H$, $a_{ii} = a_i(z_i - z_{i-1})[1 - a_i(z_i - z_{i-1})]$ if $i < H$, and $a_{ij} = -a_i a_j (z_i - z_{i-1})(z_j - z_{j-1})$ if i, j are both $< H$ and $i \neq j$.

Now $V(n)$ is equal to

$$\sum_{i=1}^{H} V_i - \sum_{1}^{H-1} {}_iT_{N_i+1}^r - \sum_{2}^{H} {}_iT_1^r + \sum_{1}^{H-1} [{}_iT_{N_i+1} + {}_{i+1}T_1]^r. \tag{2.1}$$

It can be verified easily from (2.1) and an examination of the distribution of ${}_iT_j$ that $n^{r-1/2}[V(n) - \sum_1^H V_i]$ converges stochastically to zero as n increases. Therefore if

$$\frac{n^{r-1/2}\left[\sum_1^H V_i\right] - \sqrt{n}\Gamma(r+1)\int_0^1 f^{1-r}(x)\,dx}{\sqrt{[\Gamma(2r+1) - 2r\Gamma^2(r+1)]\int_0^1 f^{1-2r}(x)\,dx - \left[(r-1)\Gamma(r+1)\int_0^1 f^{1-r}(x)\,dx\right]^2}} \tag{2.2}$$

has a limiting standard normal distribution as n increases, Theorem 1 is proved when $f(x)$ is a step function. Let us denote $\sqrt{n}[(N_i / n) - a_i(z_i - z_{i-1})]$ by W_i, and note that $W_1 + \cdots + W_H$ is identically equal to zero. The numerator of (2.2) can be written as

$$\begin{aligned}&\sqrt{\Gamma(2r+1) - (r^2+1)\Gamma^2(r+1)} \sum_1^H \frac{(z_i - z_{i-1})^r}{\left(\frac{N_i}{n}\right)^{r-1/2}} Q_i \\ &+ \sqrt{n}\Gamma(r+1) \sum_1^H \frac{(z_i - z_{i-1})}{\left[\frac{a_i N_i}{n}\right]^{r-1}}\left[[a_i(z_i - z_{i-1})]^{r-1} - \left[\frac{W_i}{\sqrt{n}} + a_i(z_i - z_{i-1})\right]^{r-1}\right]\end{aligned} \tag{2.3}$$

and remembering that N_i/n converges to $a_i(z_i - z_{i-1})$ with probability one as n increases, (2.3) has the same limiting distribution as

$$\begin{aligned}&\sqrt{\Gamma(2r+1) - (r^2+1)\Gamma^2(r+1)} \sum_1^H \frac{(z_i - z_{i-1})^{1/2}}{a_i^{r-1/2}} Q_i \\ &\qquad\qquad - (r-1)\Gamma(r+1) \sum_1^H \frac{W_i}{a_i^r}.\end{aligned} \tag{2.4}$$

But from the discussion above, it is easily verified that the distribution of (2.4) approaches a normal distribution with mean zero and variance equal to the square of the denominator of (2.2). This proves Theorem 1 when $f(x)$ is a step function.

3. Proof of Theorem 1 in the general case. The proof in the general case seems to require a great number of details, which we merely outline. In the first place, we may assume that $f(x)$ is continuous on $[0, 1]$, for if it has a finite number of discontinuities, we may handle each subinterval on which it is continuous separately, and then put them together as in Sec. 2. Then, defining λ_i as $| F(Y_i) - i/n |$, and remembering that $f(x) \geqq A > 0$, we find that $| Y_i - F^{-1}(i/n) | \leqq \lambda_i/A$. We have $F(Y_{i+1}) - F(Y_i) = f(\theta_i)[Y_{i+1} - Y_i]$, where $Y_i < \theta_i < Y_{i+1}$, or $F^{-1}(i/n) - (\lambda_i/A) < \theta_i < F^{-1}((i+1)/n) + (\lambda_{i+1}/A)$. Then we may write

$$F(Y_{i+1}) - F(Y_i) = f\left[F^{-1}\left(\frac{i}{n}\right)\right][Y_{i+1} - Y_i] + \gamma_i\,[Y_{i+1} - Y_i],$$

where $\gamma_i = f(\theta_i) - f[F^{-1}(i/n)]$. Due to the uniform continuity of $f(x)$, and the fact that $\max_i n^{1/2-\delta}\lambda_i$ converges stochastically to zero as n increases, we shall be able to ignore the term $\gamma_i[Y_{i+1} - Y_i]$ in certain respects. We denote $F(Y_{i+1}) - F(Y_i)$ by U_{i+1}, and $Y_{i+1} - Y_i$ by T_{i+1}. Then we may write

$$T_{i+1} = \frac{U_{i+1}}{f\left[F^{-1}\left(\frac{i}{n}\right)\right]} - \frac{\gamma_i T_{i+1}}{f\left[F^{-1}\left(\frac{i}{n}\right)\right]}. \tag{3.1}$$

We are going to examine the moments of the chance variable $W = \sum n^r T_i^r$, and it is clear from an examination of (3.1) that the leading terms of these moments will be the corresponding moments of, say,

$$\sum\left\{\frac{nU_i}{f\left[F^{-1}\left(\frac{i-1}{n}\right)\right]}\right\}^r = Q.$$

Let $V_1, \cdots, V_{n+1}$ be independent chance variables, each with density e^{-v} for $v > 0$. Then $E\{V_{i_1}^{a_1}V_{i_2}^{a_2} \cdots V_{i_k}^{a_k}\} = \Gamma(a_1+1)\Gamma(a_2+1)\cdots\Gamma(a_k+1)$. Also, it is well known that

$$E\{(nU_{i_1})^{a_1}(nU_{i_2})^{a_2}\cdots(nU_{i_k})^{a_k}\} = \frac{(n^{a_1+\cdots+a_k})\Gamma(n+1)\Gamma(a_1+1)\cdots\Gamma(a_k+1)}{\Gamma(n+a_1+\cdots+a_k+1)},$$

and this last expression approaches $\Gamma(a_1+1)\cdots\Gamma(a_k+1)$ as n increases. That is, with respect to their moments, the chance variables $nU_1, \cdots, nU_{n+1}$ act like the independent chance variables $V_1, \cdots, V_{n+1}$.

Defining the chance variable Q' as

$$\sum\left\{\frac{V_i}{f\left[F^{-1}\left(\frac{i-1}{n}\right)\right]}\right\}^r,$$

it is known that $E\{[(Q' - EQ') / \sigma_{Q'}]^k\}$ approaches μ_k, the kth moment of a standard normal chance variable, for any positive integral k. From the discussion above, one might expect the same to hold for $E\{[Q - EQ) / \sigma_Q]^k\}$, and a detailed examination shows that this is so. It is also so for $E\{[(W - EW) / \sigma_W]^k\}$, since the terms in this not given by the corresponding terms with W replaced by Q approach zero in the limit, due to the properties of γ_t defined above. This completes the proof.

4. The asymptotic power of certain tests of fit. To test the hypothesis that $f(x) = 1$ for $0 \leqq x \leqq 1$, the test that rejects when $V(n) \geqq C_n(\alpha)$ has been suggested, where $C_n(\alpha)$ is a constant depending on the sample size n and on the desired level of significance α. Denote $(1/\sqrt{2\pi})\int_v^\infty e^{-(t^2/2)}\, dt$ by $\phi(v)$, and let $k(\alpha)$ denote the value such that $\phi(k(\alpha)) = \alpha$. Then Theorem A shows that for large n, $C_n(\alpha)$ is approximately equal to

$$n^{-r+1/2}[\sqrt{n}\Gamma(r + 1) + k(\alpha)\sqrt{\Gamma(2r + 1) - (r^2 + 1)\Gamma^2(r + 1)}],$$

while if the true common density is $f(x)$, then the large-sample power of the test is approximately equal to

$$\phi\left(\frac{n^{r-1/2}C_n(\alpha) - \sqrt{n}\Gamma(r + 1)\int_0^1 f^{1-r}(x)\, dx}{\sqrt{[\Gamma(2r + 1) - 2r\Gamma^2(r + 1)]\int_0^1 f^{1-2r}(x)\, dx - \left[(r - 1)\Gamma(r + 1)\int_0^1 f^{1-r}(x)\, dx\right]^2}}\right).$$

REFERENCE

[1] D. A. Darling, "On a class of problems related to the random division of an interval," *Ann. Math. Stat.*, Vol. 24 (1953), pp. 239–253.

THE DISTRIBUTION OF THE NUMBER OF LOCALLY MAXIMAL ELEMENTS IN A RANDOM SAMPLE

By T. Austin, R. Fagen, T. Lehrer, and W. Penney

Washington, D. C.

0. Summary. The distribution of the number of different locally maximal elements in a random sample is found, where the sampling is from a continuous population of real numbers. This distribution has application in certain nonparametric tests; the problem of finding the distribution may be regarded as identical with the enumeration of permutations according to the number of distinct locally maximal elements.

1. Introduction. An ordered sample of n real numbers is drawn at random from a population having a continuous distribution. For a given integer k, an element of the sample is called locally maximal if it is the largest of some k consecutive elements of the sample. The distribution of the number of different

Received September 4, 1956.