DISTRIBUTION-FREE TESTS OF GOODNESS OF FIT

BY

JOHN MARVIN ROGERS

B. S. in Ed., Marion College, Marion, Indiana, 1962

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1966

Approved by:

W J Conover

Major Professor

TABLE OF CONTENTS

0. INTRODUCTION

Frequently a need arises for a test to determine if a random variable is distributed according to an hypothesized distribution function. Several tests have been developed by which an hypothesized distribution can be rejected if the probability of an observed random sample being drawn from this distribution is low. These are called tests of goodness of fit. The different tests of goodness of fit measure different aspects of the deviation of the observed values from what would be expected if the hypothesized distribution were correct.

Distribution free tests of goodness of fit are those which utilize sample statistics which, if the hypothesized distribution is correct, are independent of the distribution of the random variable involved. These tests are valuable because they can be used no matter what the hypothesized distribution may be.

The purpose of this report is to bring together in one place several distribution free tests of goodness of fit. The same notation has been used in presenting each test so that the reader may be able to more easily compare the tests, and to have a basis for choosing among them when a goodness of fit test is needed.

In this report the following conventions were adopted:

1. $F(x)$ is the true distribution function of the random variable X.
2. $X_1, X_2, \ldots, X_n$ is a random sample of size n from the random variable X.
3. $H(x)$ is the hypothesized distribution function of X.

4. $F_n(x)$ is the empirical distribution function of X. That is,

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \varepsilon\,(x - x_i) \text{ where } \varepsilon\,(t) = 1,\ t \geq 0$$

$$= 0,\ t < 0$$

5. $H(x;\theta)$ is the hypothesized class of distribution functions to which $F(x)$ belongs, where θ is a member of the set Θ of all parameter vectors with r elements.

1. CHI SQUARE TEST

Pearson (1900) presented the chi square test of goodness of fit. In the case of continuous one dimensional random variables the test statistic for this test may be written as:

$$(1.1)\quad X^2 = \sum_{i=1}^{k} \frac{\left(n \int_{c_{i-1}}^{c_i} d(F_n(x) - H(x))\right)^2}{n \int_{c_{i-1}}^{c_i} dH(x)}$$

where $c_o < c_1 < c_2 < \ldots < c_k$ and c_o, c_k can be $-\infty$, $+\infty$ respectively. X^2 is written in this form to emphasize the type of deviations that the test is able to detect. Ordinarily for actual use the X^2 statistic is defined as:

$$(1.2)\quad X^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

where $O_i = n \int_{c_{i-1}}^{c_i} dF_n(x)$, $E_i = n \int_{c_{i-1}}^{c_i} dH(x)$, and $-\infty \leq c_o < c_1 < c_2 < \ldots <$

$c_k \leq \infty$.

O_i is the number of observations in the sample which are in the interval (c_{i-1}, c_i) and $E_i = n(P(c_{i-1} < X \leq c_i))$ on the assumption that the hypothesized distribution function is correct. The exact distribution of X^2 is not distribution free (Birnbaum, 1953). However, if $H(x) = F(x)$, and E_i is sufficiently large for every i, the limiting distribution of X^2 as n increases without bound is χ^2 with k-1 degrees of freedom (Pearson, 1900).

In the case where the hypothesized distribution is of the form $H(x; \theta)$ where θ is not known the test can still be used. If θ can be estimated from the sample by certain methods we can let $\hat{\theta}_n$ be the estimate of the parameter vector. Then, if $F(x)$ is of the form $H(x; \theta)$, X^2 is given by

$$(1.3)\quad X^2 = \sum_{i=1}^{k} \frac{\left(n \int_{c_{i-1}}^{c_i} d\,[\,F_n(x) - H(x; \hat{\theta}_n)\,]\right)^2}{n \int_{c_{i-1}}^{c_i} dH(x; \hat{\theta}_n)}$$

where there are r elements of θ to be estimated. X^2 is distributed in the limit as n increases without bound, as a χ^2 variable with k-r-1 degrees of freedom. The estimate of θ must be by any best asymptotically normal estimator based on $\{O_i\}$. Chernoff and Lehman (1954) have studied the distribution of X^2 when θ is estimated by efficient estimators based on the individual sample observations. These estimators based on the individual observations are more efficient than any estimators based on the $\{O_i\}$. If these more efficient estimators are used, the distribution function of X^2 is between the distribution functions of X^2 with parameters known and X^2 with parameters estimated from the $\{O_i\}$. The probability of rejecting a true hypothesis is greater than the level of significance given by tables for χ^2 with k-r-1 degrees of freedom.

Cochran (1952) gave guidelines as to the minimum values for E_i, such that tabled chi square significance levels are "acceptably" close to the true significance levels as calculated from the exact distribution of X^2. Significance levels were deemed "acceptably" close if when tabled significance levels were 0.05 and 0.01 the true significance levels were in the intervals (0.04, 0.06) and (0.007, 0.015) respectively. The choice of these criteria was arbitrary. On the basis of available information, Cochran (1952) concluded that for tests of goodness of fit of bell-shaped curves, such as the normal distribution, there is little disturbance to the 0.05 level when a single E_i is as low as 0.5. This is also true of the 0.01 level if the number of degrees of freedom is greater than 6. In cases where all expectations may be small it appears that tabled significance levels are "acceptably" close to the correct values so long as every E_i is at least 2. Two expectations as low as 1 may be allowed with negligible disturbance to the 0.05 level. On the whole it seems that the restriction that all E_i be greater than 5 or 10 is more restrictive than necessary.

Two numerical examples of the use of the χ^2 test of goodness of fit will be given. To illustrate the use of this test against a completely specified hypothesis it will be tested that the following set of values is a random sample from a normal distribution with mean 30 and variance 100.

TABLE 1.1

Data from Tate and Clelland (1957)

16.7	18.8	24.0	35.1	39.8
17.4	19.3	24.7	35.8	42.1
18.1	22.4	25.9	36.5	43.2
18.2	22.5	27.0	37.6	46.2

Class intervals were chosen so that the expectations, E_i, of all classes were equal, as recommended by Mann and Wald (1942). In this case the expectations will be chosen so that each is equal to 5. The classes are then $(-\infty < x < 23.25)$, $(23.25 < x < 30)$, $(30 < x < 36.25)$, and $(36.25 < x < \infty)$ $0_1 = 8$, $0_2 = 4$, $0_3 = 2$, and $0_4 = 6$

$$X^2 = \frac{9 + 1 + 9 + 1}{5} = 4 .$$

If the set of values were a random sample from a normal distribution with mean 30 and variance 100, X^2 would be approximately distributed as a χ^2 variable with 3 degrees of freedom. The probability of a value of χ^2 with 3 degrees of freedom being as large or larger than 4 is about .25 . Therefore we have little justification for concluding that these numbers are not a sample from a normal distribution with mean 30 and variance 100.

To illustrate the use of the χ^2 test of goodness of fit in the case where it is necessary to estimate some parameters, the hypothesis that the following 55 observations are a random sample from a normal distribution will be tested.

TABLE 1.2

Data from Tate and Clelland (1957)

12.4	15.3	17.2	17.9	19.1	20.4	23.1	26.3	28.5	33.1	38.8
13.1	15.6	17.5	18.1	19.2	20.8	23.2	27.2	28.7	33.7	38.9
13.6	16.2	17.5	18.3	19.2	21.0	23.9	27.4	29.2	33.9	43.0
15.0	16.5	17.6	18.5	19.4	22.0	25.2	28.2	31.0	36.2	66.2
15.2	17.0	17.8	18.8	20.1	22.5	25.4	28.4	32.4	37.4	73.4

In problems of this type, if one bases his choice of class intervals on the sample observations he influences the probability of rejection to a certain degree. If one estimates the parameters of the hypothesized distribution from the individual sample observations the probability of rejection is influenced by the effect mentioned above of altering the distribution of X^2. It appears that in the usual uses of this test it would be as good to underestimate the level of significance by estimating parameters from individual sample observations as to influence the level of significance by an unknown amount by an arbitrary choice of class intervals. Therefore, even though it is recognized that the level of significance as given by tables will be only approximate, the mean and variance of the normal distribution will be estimated from the individual sample observations by the minimum-variance unbiased estimators. According to Mood and Graybill (1963), these are, respectively,

$$\bar{x} = \frac{\sum_{i=1}^{n} X_i}{n} = 25.00 \text{ and } s^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{x})^2}{n-1} = 128.14.$$

Class Boundaries, chosen so that expectations will each be 5.5, are 10.49, 15.47, 19.04, 23.11, 25.00, 27.89, 30.96, 34.53, and 39.51. Observed values are: 0, 6, 14, 11, 2, 5, 5, 5, 4, and 3 .

$$(1.4) \quad X^2 = \frac{(0-5.5)^2 + (6-5.5)^2 + \dots + (3-5.5)^2}{5.5} = 124.30 .$$

If X^2 were distributed as a χ^2 variable with 7 degrees of freedom $P(X^2 \geq 18.475)$ would be 0.01. If X^2 were distributed as a χ^2 variable with 9 degrees of freedom $P(X^2 \geq 21.666)$ would be 0.01. Since the distribution of X^2 under the null hypothesis is between these two distributions it is quite safe to conclude that this set of observations is not a sample from a normal distribution.

2. SMIRNOV'S W_n^2 TESTS

Smirnov (1936) studied the statistic

$$(2.1) \quad W_n^2 = n \int_{-\infty}^{\infty} (F_n(x) - H(x))^2 \, \Psi(H(x)) \, dH(x)$$

where $\Psi(t) \geq 0$ for $0 \leq t \leq 1$ is a given weight function chosen on the basis of what types of deviations are considered important. If $\Psi(t) = 1$ we have as a special case

$$(2.2) \quad \omega^2 = n \int_{-\infty}^{\infty} (F_n(x) - H(x))^2 \, dH(x) .$$

If the $\{X_i\}$ are considered ordered so that $X_i \leq X_{i+1}$ for all i this is

$$\omega^2 = n \sum_{i=1}^{n+1} \int_{X_{i-1}}^{X_i} (H(x) - F_n(x))^2 \, dH(x)$$

where $X_{n+1} = +\infty$, $X_o = -\infty$.

$$\omega^2 = n \sum_{i=1}^{n+1} \int_{X_{i-1}}^{X_i} (H(x) - \frac{i-1}{n})^2 \, dH(x)$$

$$= \frac{n}{3} \left[\sum_{i=1}^{n+1} (H(X_i) - \frac{i-1}{n})^3 - \sum_{i=1}^{n+1} (H(X_{i-1}) - \frac{i-1}{n})^3 \right]$$

$$= \frac{n}{3} \left[\sum_{i=1}^{n} \{ (H(X_i) - \frac{i-1}{n})^3 - (H(X_i) - \frac{i}{n})^3 \} + (H(X_{n+1}) - 1)^3 - (H(X_o))^3 \right] .$$

Since $H(X_{n+1}) = 1$ and $H(X_o) = 0$

$$\omega^2 = \frac{n}{3} \sum_{i=1}^{n} \{ [(H^3(X_i) - 3H^2(X_i)(\frac{i-1}{n}) + 3H(X_i) \frac{(i-1)^2}{n^2} - \frac{(i-1)^3}{n^3}]$$

$$- [H^3(X_i) - 3H^2(X_i)\frac{i}{n} + 3H(X_i)\frac{i^2}{n^2} - \frac{i^3}{n^3}] \}$$

$$= \sum_{i=1}^{n} [(H^2(X_i) - \frac{2i-1}{n} H(X_i) + \frac{i^2 - i + 1/3}{n^2}]$$

$$= \sum_{i=1}^{n} [(H(X_i) - \frac{2i-1}{2n})^2 + \frac{1}{12n^2}]$$

$$= \frac{1}{12n} + \sum_{i=1}^{n} (H(X_i) - \frac{2i-1}{2n})^2 .$$

ω^2 was studied earlier by Cramér (1928) and von Mises (1931). If $H(x) = F(x)$ and is continuous, then W_n^2 is distribution free, as can be seen from the fact that in this case

$$(2.3)\quad W_n^2 = n \int_{-\infty}^{\infty} [\frac{1}{n} \sum_{j=1}^{n} \varepsilon\ (x - X_j) - F(x)]^2\ \Psi\ (F(x))dF(x)$$

$$= n \int_{-\infty}^{\infty} [\frac{1}{n} \sum_{j=1}^{n} \varepsilon\ (F(x) - F(X_j)) - F(x)]^2\ \Psi\ (F(x))dF(x)$$

$$= n \int_{0}^{1} [\frac{1}{n} \sum_{j=1}^{n} \varepsilon\ (t - F(X_j)) - t\ ]^2\ \Psi\ (t)dt.$$

Since $\{F(X_j)\}$ are independent and uniformly distributed on the interval $(0,1)$, the distribution of W_n^2 does not depend on $F(x)$. The test is also consistent, if $\Psi > 0$, and requires no arbitrary grouping of the data (Darling, 1957).

For the case in which we desire to test that $F(x)$ is of the form $F(x,\theta)$ where θ is not known, Darling (1955) has studied the statistic

$$(2.4)\quad C_n^2 = n \int_{-\infty}^{\infty} \left[F_n(x) - H(x;\hat{\theta}_n)\right]^2\ dH(x;\hat{\theta}_n)$$

where $\hat{\theta}_n$ is an estimator of θ. The characteristic function of the limiting distribution of C_n^2 was found. C_n^2 is not distribution-free for all cases. However, if the variance $\hat{\theta}_n$ approaches zero sufficiently rapidly, the limiting distribution of C_n^2 is the same as that of ω_n^2. Kac (1949) gave a general method for finding the limiting distributions of statistics of the form

$$(2.5) \quad \int_{-\infty}^{\infty} V(F_n(x) - F(x))\, dF(x)$$

for quite general functions, V.

To illustrate the use of the ω^2 test the hypothesis that the data in Table 1.1 are a random sample from a normal distribution with mean 30 and variance 100 will be tested again.

TABLE 2.1

Calculations involved in the use of the ω^2 test.

i	x_i	$H(x_i)$	$\frac{2i-1}{2n}$	$(H(x_i) - \frac{2i-1}{2n})$
1	16.7	0.092	0.025	0.067
2	17.4	0.104	0.075	0.029
3	18.1	0.117	0.125	-0.008
4	18.2	0.119	0.175	-0.056
5	18.8	0.131	0.225	-0.094
6	19.3	0.142	0.275	-0.133
7	22.4	0.224	0.325	-0.101
8	22.5	0.227	0.375	-0.148
9	24.0	0.274	0.425	-0.151
10	24.7	0.298	0.475	-0.177
11	25.9	0.341	0.525	-0.184
12	27.0	0.382	0.575	-0.193
13	35.1	0.695	0.625	0.070
14	35.8	0.719	0.675	0.044
15	36.5	0.742	0.725	0.017
16	37.6	0.776	0.775	0.001
17	39.8	0.386	0.825	0.011
18	42.1	0.887	0.875	0.012
19	43.2	0.907	0.925	-0.018
20	46.2	0.947	0.975	-0.028

$$\omega^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left(H(X_i) - \frac{2i-1}{2n}\right)^2$$

$$= \frac{1}{240} + 0.2009$$

$$= 0.205$$

From the table of the asymptotic distribution of ω^2 (Anderson and Darling, 1952) the probability of a value of ω^2 this large or larger is approximately 0.25. This does not justify concluding that these numbers are not a random sample from a normal distribution with mean 30 and variance 100.

3. KOLMOGOROV'S K_n TEST

Anderson and Darling (1952) proposed a test based on the statistic

$$(3.1) \quad K_n = \sup_{-\infty < x < +\infty} (\sqrt{n}\ | F_n(x) - H(x) | \Psi (H(x)))$$

where $\Psi \geq 0$ is a preassigned weight function. A general method for calculating the limiting distribution of statistics of this form was given. Explicit limiting distributions were given for certain weight functions including

$$(3.2) \quad \Psi(t) = \sqrt{\frac{1}{t(1-t)}}$$

and

$$(3.3) \quad \Psi(t) = 1$$

If $\Psi(t) = 1$, K_n^* is the statistic given earlier by Kolmogorov (1933) as reported by Anderson and Darling (1952),

$$(3.4) \quad K_n = \sup_{-\infty < x < +\infty} (\sqrt{n}\ | F_n(x) - H(x) |)$$

If $H(x)$ is continuous and the sample observations are ordered so that

$X_1 \leq X_2 \leq \cdots \leq X_n$ then $K_n = \sqrt{n}\,(\max_{i=1,2,\ldots,n} \max[(H(X_i) - \frac{i-1}{n}),(\frac{i}{n} - H(X_i))])$.

Extensive tabulation of the percentage points of the distribution of K_n are available in Miller (1956). Darling (1957) reported that Smirnov (1939), (1944) gave the limiting distribution of the statisic

$$(3.5) \quad K_n^+ = \sup_{-\infty < x < +\infty} \sqrt{n}(F_n(x) - H(x))$$

as n increases without bound. Also, Darling (1957) reported that the effect of grouping on the K_n and K_n^+ tests have been considered by Illyašenko (1952), Gihman (1952), and Gnedenko (1952).

The use of the K_n test will be illustrated by again testing that the data in Table 1.1 are a random sample from a normal distribution with mean 30 and variance 100.

TABLE 3.1

Calculations involved in the use of the K_n test.

i	x_i	$\frac{i}{n}$	$H(x_i)$	i	x_i	$\frac{i}{n}$	$H(x_i)$
1	16.7	0.05	0.092	11	25.9	0.55	0.341
2	17.4	0.10	0.104	12	27.0	0.60	0.382
3	18.1	0.15	0.117	13	35.1	0.65	0.695
4	18.2	0.20	0.119	14	35.8	0.70	0.719
5	18.8	0.25	0.131	15	36.5	0.75	0.742
6	19.3	0.30	0.142	16	37.6	0.80	0.776
7	22.4	0.35	0.224	17	39.8	0.85	0.836
8	22.5	0.40	0.227	18	42.1	0.90	0.887
9	24.0	0.45	0.274	19	43.2	0.95	0.907
10	24.7	0.50	0.298	20	46.2	1.00	0.947

$$\frac{K_n}{\sqrt{n}} = \max_{i=1, \ldots, n} \left(\max\left[\left(H(X_i) - \frac{i-1}{n}\right), \left(\frac{i}{n} - H(X_i)\right)\right]\right)$$

$$= 0.600 - 0.382$$

$$= 0.218.$$

From tables of the distribution of $\frac{K_n}{\sqrt{n}}$ (Birnbaum, 1952) the probability of the value of K_n this large or larger is about 0.30. This does not justify rejection of the hypothesis.

Rosenblatt (1962a) pointed out that in most cases it is well known that the population from which a sample is drawn is not distributed exactly according to any theoretical distribution proposed for it. As a result, the only reason the hypothesized distribution is not rejected by any test of goodness of fit is that the sample size is not large enough. Instead of trying to show that a population has a given distribution, he proposed that tests be performed which could show if the true distribution is "close" to some hypothesized distribution. A test based on Kolmogorov's statistics was developed which has the desirable property mentioned above. In 1962(b) Rosenblatt extended this idea to the case of composite hypotheses.

4. TESTS BASED ON SPACING OF SAMPLE VALUES

Sherman (1950) discussed statistics based on the spacing of sample values. Kimball (1947) studied the statistic

$$(4.1)\quad \alpha = \sum_{i=1}^{n+1} \left(H(X_i) - H(X_{i-1}) - \frac{1}{n+1}\right)^2$$

where $X_o = -\infty$, $X_{n+1} = +\infty$, and conjectured that $\alpha^{1/2}$ is asymptotically normally distributed. Also, Moran (1947) studied the statistic

$$(4.2)\quad \beta = \sum_{i=1}^{n+1} \left(H(X_i) - H(X_{i-1})\right)^2$$

and proved that β is asymptotically normally distributed. It was noted by Sherman (1950) that the Cramér-von Mises ω^2 criterion is related to deviations from expected spacing of sample values, since

$$(4.3)\quad \omega^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left(H(X_i) - \frac{2i-1}{2n}\right)^2$$

where $\frac{2i-1}{2n}$ is the mid point of the interval $(\frac{i-1}{n}, \frac{i}{n})$, Thus, if the interval (0,1) is partitioned into n equal sub intervals then ω^2 measures the deviation of the values $y_i = H(x_i)$ from the mid point of these intervals. Sherman (1950) introduced the statistic

$$(4.4)\quad \omega_n = 1/2 \sum_{i=1}^{n+1} \left| H(X_i) - H(X_{i-1}) - \frac{1}{n+1} \right|$$

where H is the continuous distribution function of X. He showed that the distribution function of ω_n is

$$(4.5)\quad F(\omega) = 1 + \sum_{k=1}^{n} (b_k\, \omega^k) \qquad 0 \leq \omega \leq \frac{n}{n+1}$$

$$= 0 \qquad \omega < 0$$

$$= 1 \qquad \frac{n}{n+1} < \omega$$

where $b_i = \sum_{q=0}^{r} (-1)^{q+k+1} \binom{n+1}{q+1} \binom{q+k}{q} \binom{n}{k} \binom{n-q}{n+1}^{n-k}$

and r is the non-negative integer determined by

$$\frac{r}{n+1} \leq \omega < \frac{r-1}{n+1} \quad .$$

In 1957 Sherman gave a table of the 99th, 95th, and 90th percentiles of the distribution of ω_n for $n = 1,2, \ldots , 20$. Also, in this paper it was shown that

$$(4.6)\quad \omega_n^{(2)} = \frac{e^2 n}{2e-5} \left(\omega_n - \frac{1}{e}\right)$$

is distributed in the limit as n increases without bound as a standard normal variable. He stated, however, that it appears that n would have to be greater than 100 before the 99, 95, and 90th percentiles would be within one per cent of the limiting normal values.

The use of Sherman's ω_n test will be illustrated by again testing that the data in Table 1.1 are a random sample from a normal distribution with mean 30 and variance 100. Calculations are shown in Table 4.1. From the tables of the ω_n statistic given by Sherman (1957) the probability of a value of ω_n this large or larger is greater than 0.10. On the basis of this test, there is no justification for concluding that the data are not a sample from a normal distribution with mean 30 and variance 100.

TABLE 4.1

Calculations involved in the use of Sherman's ω_n test.

i	x_i	$H(x_i)$	$(H(x_i) - H(x_{i-1}) - \frac{1}{21})$
0	$-\infty$	0.000	
1	16.7	0.092	+0.044
2	17.4	0.104	-0.036
3	18.1	0.117	-0.035
4	18.2	0.119	-0.042
5	18.8	0.131	-0.036
6	19.3	0.142	-0.037
7	22.4	0.224	+0.034
8	22.5	0.227	-0.045
9	24.0	0.274	-0.001
10	24.7	0.298	-0.024
11	25.9	0.341	-0.005
12	27.0	0.382	-0.004
13	35.1	0.695	+0.265
14	35.8	0.719	-0.024
15	36.5	0.742	-0.005
16	37.6	0.776	-0.012
17	39.8	0.836	+0.012
18	42.1	0.887	+0.003
19	43.2	0.907	-0.028
20	46.2	0.947	-0.008
21	$+\infty$	1.000	+0.005

$$\sum_{i=1}^{n+1} \left| H(X_i) - H(X_{i-1}) - \frac{1}{n+1} \right| = 0.705$$

$$\omega_n = \frac{1}{2}(0.705)$$

$$= 0.353 \ .$$

5. SMOOTH TEST FOR GOODNESS OF FIT

This test which was presented by Neyman (1937) is designed to test hypotheses of specified continuous distribution functions against "smooth" alternatives, as defined later. If $F(x) = H(x)$ then the distribution of $Y = H(X)$ is uniform on the interval $(0,1)$. If $F(x) \neq H(x)$ then the range of Y is the interval $(0,1)$ but the distribution is not uniform. If $F(x)$ is some "smooth" alternative to $H(x)$ then the graph of the density function of Y will be a smooth curve on the interval $(0,1)$. This smooth curve can be represented in the following manner: Let $\pi_0, \pi_1, \ldots, \pi_k$ be a system of polynomials in y, orthogonal and normal in the interval $(0,1)$, π_i being of ith order. That is,

$$(5.1) \quad \pi_i (y) = a_{io} + a_{i1}y + \ldots + a_{ii}y^i \qquad i = 0,1,2, \ldots , k$$

where $\{a_{ij}\}$ $i = 0,1, \ldots , k$; $j = 0,1,2, \ldots , i$ are constants and $a_{ii} \neq 0$. By the statement that the polynomials are orthogonal on the the interval $(0,1)$ it is meant that

$$(5.2) \quad \int_0^1 \pi_i (y) \pi_j (y) \, dy = 0 \qquad i \neq j \quad .$$

By the statement that the polynomials are normal on the interval $(0,1)$ it is meant that

$$(5.3)\quad \int_0^1 \pi_i^2\,(y)dy = 1\ .$$

The first five of these polynomials are:

$$(5.4)\quad \pi_0(y) = 1$$

$$(5.5)\quad \pi_1(y) = \sqrt{12}\ (y-1/2)$$

$$(5.6)\quad \pi_2(y) = \sqrt{5}\ (6(y-1/2)^2 - 1/2)$$

$$(5.7)\quad \pi_3(y) = \sqrt{7}\ (20(y-1/2)^3 - 3(y-1/2))$$

$$(5.8)\quad \pi_4(y) = 210\ (y-1/2)^4 - 45(y-1/2)^2 + \frac{9}{8}\ .$$

Other polynomials of this set may be found by the recurrence formula (David, 1939)

$$(5.9)\quad \frac{(n+1)}{\sqrt{2n+3}}\,\pi_{n+1}(y) - 2(y-1/2)\ \sqrt{2n-1}\ \ \pi_n(y) + \frac{n}{\sqrt{2n-1}}\ \pi_{n-1}(y) = 0.$$

Now the set Ω_k of smooth alternative density functions is that set which can be represented by formulas of the form

$$(5.10)\quad f(y) = c\ \exp\{\sum_{i=1}^{k} \theta_i\ \pi_i\ (y)\}\qquad 0 \le y \le 1$$

$$= 0 \qquad\qquad \text{otherwise}$$

where $\{\theta_i\}$ is a set of parameters and c is the function of $\{\theta_i\}$ such that

$$(5.11)\quad \int_0^1 f(y)dy = 1\ .$$

The density of y under the null hypothesis is given in the case $\theta_1 = \theta_2 = \dots = \theta_k = 0$. k will be called the order of the test. The higher the order of the test which is used the more general will be the set of alternatives against which the test is sensitive. A test of order k+1 is sensitive to all alternatives to which a test of order k is sensitive. However, a test of order k+1 is sensitive to some alternatives to which a test of order k is completely insensitive. When applying this test the order should be chosen large enough to include all alternatives to which it is desired to have the test most sensitive. A test of higher order than necessary should not be used, because as the order increases the power of the test against alternatives to which a low order test is sensitive decreases. Neyman (1937) conjectured that in most practical cases, there is no need to use a test of order higher than 4. It has been shown by David (1939) that Neyman's Ψ_k^2 criterion, defined below, is distributed in the limit as n increases without bound as a χ^2 variable with k degrees of freedom.

The following steps should be used in the application of the "smooth test" for goodness of fit:

1. Obtain for each sample value x_j the value y_j by means of the formula $y_j = H(x_j)$.

2. Choose the order, k, of test to be used.

3. Calculate for each y_j the values of $\pi_i\ (y_j)$ for $i = 1,2, \dots , k$.

4. Calculate the quantities

$$u_i = \sum_{j=1}^{n} \pi_i\ (y_j) \qquad \text{for } i = 1,2, \dots , k.$$

5. Calculate

$$\Psi_k^2 = \sum_{i=1}^{k} u_i^2 .$$

6. Reject the hypothesized distribution if

$$\Psi_k^2 \geq \Psi_\varepsilon^2 (k)$$

where $$\frac{1}{2^{(\frac{k}{2})}\Gamma(\frac{k}{2})} \int_{\Psi_\varepsilon^2(k)}^{\infty} u^{(\frac{k-2}{2})} e^{-(\frac{u}{2})} du = \varepsilon$$

and ε is the probability of rejecting the hypothesized distribution when it is in fact true. Calculations can be somewhat simplified by letting $z_j = y_j - 1/2$. If $\sum_{j=1}^{n} z_j{}^a$ is denoted by $[z^a]$ then

$$(5.12) \quad u_1^2 = 12 \; [z]^2 \; n^{-1}$$

$$(5.13) \quad u_2^2 = 180([z^2] - \frac{n}{12}) \; n^{-1}$$

$$(5.14) \quad u_3^2 = 7(20 \; [z^3] - 3 \; [z])^2 \; n^{-1}$$

$$(5.15) \quad u_4^2 = (210 \; [z^4] - 45 \; [z^2] + \frac{9n}{8})^2 \; n^{-1} .$$

The following properties of the smooth test are given:

1. If the hypothesized distribution is correct, the probability of an unjust rejection of this hypothesis is the chosen level of significance ε.

2. If the hypothesis tested is incorrect, for small values of $\lambda = \sum_{i=1}^{k} \theta_i^2$ the power of the test is the same against all alternatives in the set Ω_k such that λ is the same constant.

3. If the hypothesis tested is false and one of the alternatives is true, for which the resulting value of λ is small, the chance of detecting the falsehood of the hypothesis tested is greater than that corresponding to any other similar test having the properties 1 and 2.

The limiting power function is found and the statements made here concerning the power of the test are illustrated in the paper by Neyman (1937). David (1939) examined the distributions of Ψ_1^2 and Ψ_2^2 and concluded that no great error will be made if it is assumed for samples of size 20 and over that if $H(x)$ is correct, Ψ_1^2 and Ψ_2^2 are χ^2 variables with one and two degrees of freedom, respectively. Tables of exact 0.05 and 0.01 percentage points of the distribution of Ψ_1^2 were given by Barton (1953). In this paper, Barton (1953) also discussed the power of the test and gave methods of determining the sample size necessary to give a specified power against alternatives which deviate from the hypothesized distribution by a given amount when a particular level of significance is desired. A form of the smooth test which is applicable to grouped or discrete data was developed and investigated by Barton (1955). The χ^2 goodness of fit test was shown to be a special case of this test. The loss of power caused by grouping was shown to be small and to be minimized by grouping so that all classes have equal expected frequencies. In the case where some parameters must be estimated from the sample, Barton (1956) found the large sample distributions of Ψ_k^2 in both the original form and the extended form for grouped data. In this case Ψ_k^2 is not distribution-free.

The smooth test was used by Neyman (1937) to test two hypotheses:

1. That the first 100 figures given in Mahalanobis' table of random normal deviates (Mahalanobis et al., 1933) were a random sample from the standard normal distribution.

2. That the same 100 figures were a random sample from a distribution with density function,

$$(5.16)\quad H(x) = \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})\ \sqrt{\pi}\ \sqrt{1+x^2}} \qquad \text{with } n = 20.$$

The necessary calculations are given in the tables below:

TABLE 5.1

Powers of z corresponding to each x_i for hypothesis 1.

	x_i	z_1	z_1^2	z_1^3	z_1^4
	-0.54	-0.205	0.0420	-0.0086	0.0018
	-0.21	-0.083	0.0069	-0.0006	0.0000
	.	.	.	.	.
	.	.	.	.	.
	.	.	.	.	.
	+2.02	+0.478	0.2285	+0.1092	0.0522
	-1.00	-0.341	0.1163	-0.0397	0.0135
Total		+0.623	7.867	+0.453	1.133

TABLE 5.2

Powers of z corresponding to each x_i for hypothesis 2.

x_i	z_2	z_2^2	z_2^3	z_2^4
-0.54	-0.49	0.2401	-0.1176	0.0576
-0.21	-0.31	0.0961	-0.0298	0.0092
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
+2.02	+0.50	0.2500	+0.1250	0.0625
-1.00	-0.50	0.2500	-0.1250	0.0625
Total	-2.45	19.903	-0.599	4.599

TABLE 5.3

Values of u_k^2 and v_k^2 calculated for hypotheses 1 and 2.

Order of the test k	Hypothesis 1		Hypothesis 2			
	u_k^2	v_k^2	u_k^2	v_k^2	$v_{0.05}^2(k)$	$v_{0.01}^2(k)$
1	0.047	0.047	0.720	0.720	3.8415	6.6349
2	0.391	0.438	240.957	241.677	5.9915	9.2103
3	3.624	4.062	1.494	243.171	7.8147	11.3449
4	0.129	4.191	333.426	574.597	9.4877	13.2767

Although tests of 4 orders are given here, ordinarily only one order of the test should be used. The order should be chosen on the basis of the types of alternatives against which it is desired that the test be most sensitive. Although, it appears intuitively that this test should be quite powerful against the alternatives it is designed to detect, there are difficulties which impair its usefulness. It is necessary that the hypothesized distribution be completely specified. Calculations of the values, y and z, corresponding to each sample observation appear laborious unless the hypothesized distribution is well tabled, but could be accomplished by the use of digital computing equipment.

ACKNOWLEDGMENTS

The author wishes to thank his Major Professor, Dr. W. J. Conover, Assistant Professor of Statistics for his advice and encouragement throughout the preparation of the manuscript. He also wishes to thank Dr. H. C. Fryer, Professor and Head of Statistics, Dr. K. S. Banerjee, Visiting Professor of Statistics, and Dr. L. E. Fuller, Professor of Mathematics for reviewing the manuscript and serving on the committee.

REFERENCES

Anderson, T. W., and Darling, D. A. (1952). Asymptotic theory of certain "Goodness of fit" criteria based on stochastic processes. Ann. Math. Statist. 23 193-212.

Barton, D. E. (1953). On Neyman's smooth test for goodness of fit and its power with respect to a particular system of alternatives. Skandinavisk Aktuarietidskrift 36 24-63.

Barton, D. E. (1955). A form of Neyman's χ^2 test of goodness of fit applicable to grouped and discrete data. Skandinavisk Aktuarietidskrift 38 1-16.

Barton, D. E. (1956). Neyman's χ^2 test of goodness of fit when the null hypothesis is composite. Skandinavisk Akturietidskrift 39 216-245.

Birnbaum, Z. W. (1952). Numerical tabulation of the distribution of Kolmogorov's statistic for finite sample size. J. Amer. Statist. Assoc. 47 425-441.

Birnbaum, Z. W. (1953). Distribution-free tests of fit for continuous distribution functions. Ann. Math. Statist. 24 1-8.

Chernoff, H., and Lehmann, E. L. (1954). The use of maximum likelihood estimates in χ^2 tests of goodness of fit. Ann. Math. Statist. 25 579-586.

Cochran, W. G. (1952). The χ^2 test of goodness of fit. Ann. Math. Statist. 23 315-345.

Cramér, H. (1928). On the composition of elementary errors. Skandinavisk Aktuarietidskrift 11 13-74, 141-180.

Darling, D. A. (1955). The Cramér-Smirnov test in the parametric case. Ann. Math. Statist. 26 1-20.

Darling, D. A. (1957). The Kolmogorov-Smirnov, Cramér-von Mises tests. Ann. Math. Statist. 28 823-828.

David, F. N. (1939). On Neyman's "Smooth" test for goodness of fit. I. Distribution of the criterion Ψ^2 when the hypothesis is true. Biometrika 31 191-199.

Gihman, I. I. (1952). On a criterion of fit for discrete random variables. Dopovidi Akademii Nauk Ukrains'koi Radyans'koi Socialisticnoi Respubliki 7-9.

Gnedenko, B. V. (1952). Some remarks on the papers of O. A. Illyašenko and I. I. Gihman. Dopovidi Akademii Nauk Ukrains'koi Radyans'koi Socialisticnoi Respubliki 10-12.

Illyašenko, O. A. (1952). On the influence of grouping of empirical data on A. N. Kolmogorov's criterion of fit. Dopovidi Akademii Nauk Ukrains'koi Radyans'koi Socialisticnoi Respubliki 3-6.

Kac, M. (1949). On deviations between theoretical and empirical distributions. Proc. Nat. Acad. Sci. U.S.A. 35 252-257.

Kimball, B. F. (1947). Some basic theorems for developing tests of fit for the case of the non-parametric probability distribution function. I. Ann. Math. Statist. 18 540-548.

Kolmogorov, A. (1933). Sulla determinazione emperica di una legge di distribuzione. Giorn. Ist. Ital. Attuari 4 83-91.

Mahalanobis, P. C., Bose, S. S., Ray, P. R., and Banerji, S. K. (1933). Tables of random samples from a normal population. Sankhya 1 289-328.

Mann, H. B., and Wald, A. (1942). On the choice of the number of class intervals in the application of the chi square test. Ann. Math. Statist. 13 306-317.

Miller, L. H. (1956). Table of percentage points of Kolmogorov statistics. J. Amer. Statist. Assoc. 51 111-121.

Mood, A. M., and Graybill, F. A. (1963). Introduction to the Theory of Statistics. McGraw-Hill, New York.

Moran, P. A. P. (1947). The random division of an interval. J. Roy. Statist. Soc. Suppl. 9 92-98.

Neyman, J. (1937). Smooth test for goodness of fit. Skandinavisk Aktuarietidskrift 20 150-199.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philos. Mag. and J. of Sci. Series 5 50 157-172.

Rosenblatt, J. (1962a). Some modified Kolmogorov-Smirnov tests of approximate hypotheses and their properties. Ann. Math. Statist. 33 513-524.

Rosenblatt, J. (1962b). Testing approximate hypotheses in the composite case. Ann. Math. Statist. 33 1256-1364.

Sherman, B. (1950). A random variable related to the spacing of sample values. Ann. Math. Statist. 21 339-361.

Sherman, B. (1957). Percentiles of the ω_n statistic. Ann. Math. Statist. 28 259-261.

Smirnov, N. (1936). Sur la distribution de ω^2. C. R. Acad. Sci. Paris. 202 449-452.

Smirnov, N. (1939). On the estimation of the discrepancy between empirical curves of distribution for two independent samples. Bull. Math. Univ. Moscow **2** No. 2 3-14.

Smirnov, N. (1944). Approximate laws of distribution of random variables from empirical data. Uspehi Matem. Nauk **10** 179-206.

Tate, M. W., and Clelland, R. C. (1957). Nonparametric and Shortcut Statistics. Interstate Printers and Publishers, Inc., Danville Illinois.

von Mises, R. (1931). Wahrscheinlichkeitsrechnung. Leipzig u. wien.

DISTRIBUTION-FREE TESTS OF GOODNESS OF FIT

by

JOHN MARVIN ROGERS

B. S. in Ed., Marion College, Marion, Indiana, 1962

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1966

Frequently there is a need for a test of whether a set of observations is a random sample from a population with a distribution of a given form. Distribution-free tests of goodness of fit are among the most useful of such tests. Their usefulness does not depend on the particular form of the hypothesized distribution.

The χ^2 test of goodness of fit is useful whether the hypothesized distribution is completely specified or of the parametric type. If some parameters of the hypothesized distribution are estimated from the sample, estimation of parameters must be by a specified type of estimators or else the distribution of the X^2 statistic will be altered.

Smirnov's ω^2 and Kolmogorov's K_n tests are generally useful for testing completely specified continuous hypothesized distributions. Generalizations of these tests make possible tests for specialized uses. In some cases these can be extended to the parametric hypothesized distribution.

Neyman's "smooth" test is useful against "smooth" alternatives to a completely specified, hypothesized continuous distribution. However, calculations are quite difficult if the hypothesized distribution is not well tabled.

Tests based on spacing of sample observations were discussed. Examples of the use of the various tests were presented.