



A Test of Goodness of Fit

T. W. Anderson; D. A. Darling

Journal of the American Statistical Association, Vol. 49, No. 268. (Dec., 1954), pp. 765-769.

Stable URL:

<http://links.jstor.org/sici?sici=0162-1459%28195412%2949%3A268%3C765%3AATOGOF%3E2.0.CO%3B2-L>

Journal of the American Statistical Association is currently published by American Statistical Association.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/astata.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact support@jstor.org.

A TEST OF GOODNESS OF FIT

T. W. ANDERSON AND D. A. DARLING*

Columbia University and University of Michigan

Some (large sample) significance points are tabulated for a distribution-free test of goodness of fit which was introduced earlier by the authors. The test, which uses the actual observations without grouping, is sensitive to discrepancies at the tails of the distribution rather than near the median. An illustration is given, using a numerical example used previously by Birnbaum in illustrating the Kolmogorov test.

1. THE PROCEDURE

THE problem of statistical inference considered here is to test the hypothesis that a sample has been drawn from a population with a specified continuous cumulative distribution function $F(x)$. For example, the population may be specified by the hypothesis to be normal with mean 1 and variance $\frac{1}{6}$; the corresponding cumulative distribution function is

$$(1) \quad F(x) = \sqrt{\frac{3}{\pi}} \int_{-\infty}^x e^{-3(y-1)^2} dy.$$

In practice the procedure really tests the hypothesis that the sample has been drawn from a population with a completely specified density function, since the cumulative distribution function is simply the integral of the density.

The test procedure we have proposed earlier [1] is the following: Let $x_1 \leq x_2 \leq \dots \leq x_n$ be the n observations in the sample in order, and let $u_i = F(x_i)$. Then compute

$$(2) \quad W_n^2 = -n - \frac{1}{n} \sum_{j=1}^n (2j-1) [\log u_j + \log (1 - u_{n-j+1})]$$

where the logarithms are the natural logarithms. If this number is too large, the hypothesis is to be rejected.

This procedure may be used if one wishes to reject the hypothesis whenever the true distribution differs materially from the hypothetical and especially when it differs in the tails.

Significance points for W_n^2 are not available for small sample sizes. The asymptotic significance points are given below:

* Work sponsored by Office of Scientific Research, U. S. Air Force, Contract AF18(600)-442, Project No. R-345-20-7.

The authors wish to acknowledge the assistance of Vernon Johns in the computations.

ASYMPTOTIC SIGNIFICANCE POINTS

Significance Level	Significance Point
.10	1.933
.05	2.492
.01	3.857

2. A NUMERICAL ILLUSTRATION

Birnbaum [2] has considered a sample of 40 observations and applied the Kolmogorov statistic to test the hypothesis that the population from which the data came was normal with mean 1 and standard deviation $1/\sqrt{6}$. By this test he found the data were consistent with the hypothesis. We have analyzed the same data using (2), obtaining $W_n^2=1.158$ which is well below the 10 per cent significance point, and we do not reject the hypothesis.

The computation sheet for this calculation had the following columns: x_j , $\sqrt{6}(x_j-1)$, $u_j=F(x_j)$, $1-u_{n-j+1}$, $\log u_j$, $\log(1-u_{n-j+1})$ and $-\log u_j + \log(1-u_{n-j+1})$. The operation $u_j=F(x_j)$ is simply finding the probability to the left of $\sqrt{6}(x_j-1)$ according to the standard normal distribution.

Another test procedure uses the Cramer-von Mises ω^2 criterion given by

$$(3) \quad n\omega^2 = \frac{1}{12n} + \sum_{j=1}^n \left(u_j - \frac{2j-1}{2n} \right)^2.$$

The asymptotic distribution of this statistic is given in [1]. For Birnbaum's data we obtain $n\omega^2=.1789$, which is also well below the 10 per cent asymptotic significance point of .3473

In these two examples we have used the asymptotic percentage points instead of the actual ones based on finite sample size. Empirical study suggests that the asymptotic value is reached very rapidly, and it appears safe to use the asymptotic value for a sample size as large as 40.

Application to the same data of the usual χ^2 criterion of K. Pearson, using 8 categories each with expected frequency 5, shows that $\chi^2=6.4$ which with 7 degrees of freedom is not significant at the 10 per cent level.

3. DERIVATION OF THE CRITERION

Several test procedures are based on comparing the specified cumulative distribution function $F(x)$ with its sample analogue, the empirical cumulative distribution function

$$F_n(x) = \frac{\text{no. of } x_i \leq x}{n}.$$

The present writers suggested [1] the use of the criterion

$$(4) \quad W_n^2 = n \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 \psi(F(x)) dF(x)$$

where $\psi(u)$ is some nonnegative weight function chosen by the experimenter to accentuate the values of $F_n(x) - F(x)$ where the test is desired to have sensitivity. The hypothesis is to be rejected if W_n^2 is sufficiently large. When $\psi(u) \equiv 1$ this criterion is n times the ω^2 criterion.

The criterion W_n^2 is an average of the squared discrepancy $[F_n(x) - F(x)]^2$, weighted by $\psi[F(x)]$ and the increase in $F(x)$ (and the normalization n). If one wishes the test to have good power against alternatives in which $H(x)$, the true distribution, and $F(x)$ disagree near the tails of $F(x)$, and to this end is willing to sacrifice power against alternatives in which $H(x)$ and $F(x)$ disagree near the median of $F(x)$, it seems that one ought to choose $\psi(u)$ to be large for u near 0 and 1, and small near $u = \frac{1}{2}$. Even if the alternative hypotheses are closely delineated, however, it appears difficult to find an "optimum" weight function $\psi(u)$. For a discussion of the general nature of power of distribution-free tests, see, for example, Birnbaum [3] and Lehmann [4].

For a given value of x , $F_n(x)$ is a binomial variable; it is distributed in the same way as the proportion of successes in n trials, where the probability of success is $H(x)$. Thus, $E[F_n(x)] = H(x)$ and

$$(5) \quad nE[F_n(x) - F(x)]^2 = nE[F_n(x) - H(x)]^2 + n[F(x) - H(x)]^2 \\ = H(x)[1 - H(x)] + n[F(x) - H(x)]^2.$$

Under the null hypothesis ($H(x) = F(x)$), the variance is $F(x)[1 - F(x)]$. In a sense, we would equalize the sampling error over the entire range of x by weighting the deviation by the reciprocal of the standard deviation under the null hypothesis, that is, by using

$$(6) \quad \psi(u) = \frac{1}{u(1 - u)}$$

as a weight function. This function has the effect of weighting the tails heavily since this function is large near $u = 0$ and $u = 1$. It is this weight function (6) which we treat in the present note.

Formula (2) is obtained by writing (4) as

$$\begin{aligned} \frac{1}{n} W_n^2 &= \int_{-\infty}^{\infty} \frac{[F_n(x) - F(x)]^2}{F(x)[1 - F(x)]} dF(x) \\ &= \int_{-\infty}^{x_1} \frac{F^2(x)}{F(x)[1 - F(x)]} dF(x) + \int_{x_1}^{x_2} \frac{[F_n(x) - F(x)]^2}{F_n(x)[1 - F(x)]} dF(x) \\ &\quad + \cdots + \int_{x_n}^{\infty} \frac{[1 - F(x)]^2}{F(x)[1 - F(x)]} dF(x), \end{aligned}$$

and letting $F(x) = u (dF(x) = du)$. Straightforward integration and collection of terms gives (2). The formula (2.5) in [1] cannot be used directly here, for that formula requires that $\int_0^1 \psi(u) du < \infty$, which is not true of (6).

4. COMPUTATION OF THE ASYMPTOTIC SIGNIFICANCE POINTS

It was proved in [1] that the limiting characteristic function of W_n^2 defined in either (2) or (4) is

$$(7) \quad \phi(t) = \lim_{n \rightarrow \infty} E(e^{itW_n^2}) = \sqrt{\frac{-2\pi it}{\cos\left(\frac{\pi}{2}\sqrt{1+8it}\right)}}$$

and that the inversion of this characteristic function gave for the limiting cumulative distribution of W_n^2 the expression

$$(8) \quad \frac{\sqrt{2}}{z} \sum_{j=0}^{\infty} \frac{(-1)^j \Gamma(j + \frac{1}{2})(4j + 1)}{j!} e^{-[(4j+1)^2 \pi^2] / (8z)} \cdot \int_0^{\infty} e^{z/[8(w^2+1)] - [(4j+1)^2 \pi^2 w^2] / (8z)} \cdot dw.$$

The terms of this series alternate in sign and the $(j+1)$ st term is less than the j th term, $j \geq 1$; thus the error involved in using only j terms of this series is less than the $(j+1)$ st term for $j \geq 0$. By using the fact that $e^{z/[8(w^2+1)]} \leq e^{z/8}$, one can easily verify that to compute the probabilities correctly to four decimal places, one needs only the 0-th term for the first two significance points and the 0-th and 1-st terms for the third significance point. The laborious part of the computation is the evaluation of the integral. Let $[(4j+1)\pi/2\sqrt{z}]w = y$; then the integrand is $f(y)e^{-\frac{1}{2}y^2}$. The y -axis was divided into intervals according to the integral $e^{-\frac{1}{2}y^2}$ and numerical integration was performed.

The moments of the asymptotic distribution are fairly easy to obtain from formulas given in [1]. The first two are

$$\lim_{n \rightarrow \infty} E(W_n^2) = E(W_n^2) = \sum_{j=1}^{\infty} \frac{1}{j(j+1)} = 1,$$
$$\lim_{n \rightarrow \infty} \text{Var}(W_n^2) = 2 \sum_{j=1}^{\infty} \frac{1}{j^2(j+1)^2} = \frac{2}{3} (\pi^2 - 9) \sim .57974.$$

The asymptotic significance points are computed to assure the probabilities (significance levels) to be correct to four decimal places.

REFERENCES

- [1] Anderson, T. W., and Darling, D. A., "Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes," *Annals of Mathematical Statistics*, 23 (1952), 193-212.
- [2] Birnbaum, Z. W., "Numerical tabulation of the distribution of Kolmogorov's statistic for finite sample size," *Journal of the American Statistical Association*, 47 (1952), 425-41.
- [3] Birnbaum, Z. W., "Distribution-free tests of fit for continuous distribution functions," *Annals of Mathematical Statistics*, 24 (1953), 1-8.
- [4] Lehmann, E. L., "The power of rank tests," *Annals of Mathematical Statistics*, 24 (1953), 23-43.