DISTRIBUTION-FREE TESTS OF FIT FOR CONTINUOUS DISTRIBUTION FUNCTIONS[1, 2]

By Z. W. Birnbaum

University of Washington

Summary. A class of statistics, large enough to comprise those used in all the known distribution-free tests of fit for continuous distribution functions, is characterized by a structure called "structure (d)." A number of statistics of this class may be constructed and used for tests of fit. To make a reasonable choice among all these statistics, it appears desirable to introduce in the space of continuous distribution functions a distance which would reflect the type of discrepancy the proposed test is intended to detect. By studying the power of various statistics with regard to this distance one may then be able to choose those with optimal properties.

1. Introduction. The main object of this paper is to discuss techniques for deciding whether a sample $X_1, X_2, \cdots, X_n$ of a one-dimensional random variable X was obtained from a population which has a completely specified continuous cumulative distribution function $H(x)$. More specifically, we shall limit ourselves to the following problem.

Given a continuous cumulative probability function $H(x) = \text{Prob}\,\{X \leqq x\}$ and a class $\mathcal{A}$ of continuous cumulative probability functions different from $H(x)$; required is a procedure which, for every sample $X_1, \cdots, X_n$, will enable us to decide whether to accept the "hypothesis" $H(x)$ or the "alternative" $\mathcal{A}$. This procedure should be distribution-free for every sample-size n, in a sense which will be described in Section 3. Procedures of this kind are referred to as "distribution-free tests of fit." In the course of our presentation it will appear necessary to formulate the concept of distribution-free statistics and to discuss some of its properties.

2. Some known tests. The following is a concise description of some distribution-free tests of fit and the statistics on which they are based. The tests mentioned here are chosen mainly for illustrative purposes and no attempt at completeness has been made.

2.1. *The chi square test*[3]. This test compares the empirical histogram determined by a sample with the "expected" histogram determined by the hypothesis

Received 9/3/52.

[1] Presented as a Special Invited Address at the meeting of the Institute of Mathematical Statistics in Eugene, Ore., June 21, 1952.

[2] Work done under the sponsorship of the Office of Naval Research.

[3] An excellent presentation of the many aspects of the χ^2-test was given recently by W. G. Cochran [1], hence only a few lines are devoted to it here.

$H(x)$, by means of the χ^2 statistic. The limiting distribution of this statistic is known and extensively tabulated, but little is known about the manner in which its exact distribution for finite sample size approaches the limiting distribution when the hypothesis is true, and even less when the alternative is true. The chi-square test is not distribution-free in the strict sense, since for finite sample size χ^2 is not a distribution-free statistic. It is being mentioned here mainly for historical reasons, as the procedure to which the term "test of goodness of fit" was originally applied.

2.2 *Smirnov's ω^2-test.* Let $F_n(x)$ be the "empirical distribution function" defined by

$$F_n(x) = k/n \qquad \text{if } k \text{ sample values are } \leqq x,\ k = 0, 1, \cdots, n.$$

Modifying statistics proposed earlier by Cramér and v. Mises, Smirnov [2] compares $F_n(x)$ with the hypothesis $H(x)$ by means of the statistic

$$\omega^2 = n \int_{-\infty}^{+\infty} [F_n(x) - H(x)]^2 \, dH(x) \tag{2.21}$$

which, after some algebra, may also be written

$$\omega^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left[H(X_i) - \frac{2i-1}{n} \right]^2. \tag{2.22}$$

Smirnov showed that, if $H(x)$ is true, the probability distribution of ω^2 is independent of $H(x)$ for any n, and he obtained an asymptotic expression for this probability distribution for $n \to \infty$.

2.3. *The W_n^2-tests.* In [3] T. W. Anderson and D. A. Darling consider a generalization of ω^2, given by

$$W_n^2 = n \int_{-\infty}^{+\infty} [F_n(x) - H(x)]^2 \psi[H(x)] \, dH(x), \tag{2.31}$$

where $\psi(t) \geqq 0$, for $0 \leqq t \leqq 1$, is a given weight function. This statistic can be rewritten

$$W_n^2 = 2 \sum_{j=1}^{n} \left\{ \Phi_2[H(X_j)] - \frac{2j-1}{2n} \Phi_1[H(X_j)] \right\} + k_{n,\psi}, \tag{2.32}$$

where $\Phi_1(t)$, $\Phi_2(t)$ are functions determined by $\psi(t)$ only. It is easily shown that for any $\psi(t)$ and n the probability distribution of W_n^2, if $H(x)$ is true, is independent of $H(x)$. For $\psi(t) \equiv 1$, (2.31) reduces to (2.21) and thus Smirnov's ω^2 is obtained as a special case. Anderson and Darling present a general method for obtaining the asymptotic distribution of W_n^2 for $n \to \infty$. For ω^2 their method yields an expression different from that given by Smirnov, better adapted for computation. A table obtained by using this expression is published in [3].

2.4. *The W_n^2-test with $\psi(t) = 1/t(1-t)$.* Since for x small the empirical distribution function $F_n(x)$ and the hypothetical $H(x)$ are both close to 0, and for x large both are close to 1, the ω^2 test is likely not to detect a discrepancy in the tails of the distribution. For cases where such discrepancies are of importance, Anderson and Darling [3] consider the weight function $\psi(t) = 1/t(1 - t)$. They derive an asymptotic expression for the distribution of the resulting W_n^2 statistic, but a tabulation of this distribution is not available.

2.5. *Kolmogorov's test.* Kolmogorov [4] introduced the statistic

$$D_n = \sup_{-\infty < x < +\infty} | F_n(x) - H(x) |, \tag{2.51)}$$

showed that its probability distribution, if $H(x)$ is true, is independent of $H(x)$, and derived the asymptotic probability distribution of D_n for $n \to \infty$. A tabulation of this asymptotic distribution was given by Smirnov [5]. In [4] Kolmogorov also obtained recursion formulae which make it possible to compute the probability distribution of D_n for finite n. This distribution has been tabulated by Massey [6], [7], and Birnbaum [8]. Wald and Wolfowitz [9] considered a more general class of distribution-free statistics and showed how their probability-distributions can be computed for finite n.

2.6. *The K_n-tests.* In generalization of Kolmogorov's D_n, Anderson and Darling [3] consider the statistic

$$K_n = \sup_{-\infty < x < +\infty} \sqrt{n} \, | F_n(x) - H(x) | \, \sqrt{\psi[H(x)]}, \tag{2.61}$$

where $\psi(t) \geqq 0$ is a given weight function. This statistic may also be written in the form

$$K_n = \frac{1}{\sqrt{n}} \operatorname*{Max}_{j=1,\cdots,n} \{\sqrt{\psi[H(X_j)]} \operatorname{Max} [nH(X_j) - (j - 1), j - nH(X_j)]\}. \tag{2.62}$$

The probability distribution of K_n, if $H(x)$ is true, does not depend on $H(x)$, and a method for obtaining the asymptotic distribution of K_n for $n \to \infty$ is developed in [3]. Kolmogorov's D_n is equivalent with the special case obtained from (2.61) by setting $\psi(t) = 1$. Anderson and Darling consider also the important special case $\psi(t) = 1/t(1 - t)$ which yields a statistic suitable for detecting discrepancies in the tails of the distribution. They have not succeeded, however, in obtaining an expression for the asymptotic distribution which would lend itself for practical use.

2.7. *Tests related to spacing of sample values.* If $X_1 \leqq X_2 \leqq \cdots \leqq X_n$ is an ordered sample of a random variable with probability distribution $H(x)$, then the expectation of $H(X_{i+1}) - H(X_i)$ is $1/(n + 1)$ for $i = 0, 1, \cdots, n$, with the notations $H(X_0) = 0$, $H(X_{n+1}) = 1$. Any statistic which evaluates the discrepancy between this expected and the actual spacing of the values $H(X_i)$ may be used to test the hypothesis $H(x)$. Thus Kimball [10] considers the statistic

$$\sum_{i=1}^{n+1} \left[H(X_i) - H(X_{i-1}) - \frac{1}{n + 1} \right]^2 \tag{2.71}$$

without deriving its asymptotic distribution. Moran [11] uses the statistic

$$\sum_{i=1}^{n+1} [H(X_i) - H(X_{i-1})]^2 \tag{2.72}$$

which differs from (2.71) by a constant, and shows that it is asymptotically normally distributed. Sherman [12] recently proposed the statistic

$$\frac{1}{2} \sum_{i=1}^{n+1} | H(X_i) - H(X_{i-1}) |, \tag{2.73}$$

derived expressions for its distribution for finite n, and showed that it is asymptotically normal.

2.8. *One-sided tests of fit.* Wald and Wolfowitz [9] studied the statistic

$$D_n^+ = \sup_{-\infty < x < +\infty} [F_n(x) - H(x)] = \operatorname*{Max}_{j=1,\cdots,n} \left[\frac{j}{n} - H(X_j)\right] \tag{2.81}$$

which, if $H(x)$ is true, has a probability distribution independent of $H(x)$, and obtained expressions for this distribution for finite n. Birnbaum and Tingey [13] obtained an alternative expression for the distribution of D_n^+ for finite n and tabulated it. Smirnov [14] obtained the asymptotic distribution in form of an elementary function.

3. On the concept of a distribution-free statistic.

3.1. *Statistics of structure* (d). The tests described in the preceding sections are all based on statistics which can be written in the form $\Phi[H(X_1), \cdots, H(X_n)]$, where $\Phi(U_1, \cdots, U_n)$ is a measurable symmetric function of $U_1, \cdots, U_n$. We will refer to such statistics as *statistics of structure* (d) and investigate how this particular structure of a statistic is related to its distribution-free character.

3.2. *Distribution-free and strongly distribution-free statistics.* Let Ω be a family of cumulative probability functions. We shall say that $S[X_1, \cdots, X_n, G]$ is a *distribution-free statistic in* Ω if

(i) for every $G \in \Omega$ it is a measurable function of $X_1, \cdots, X_n$, and

(ii) whenever $X_1, \cdots, X_n$ is a sample of a random variable X with the cumulative probability function G, the probability distribution of $S[X_1, \cdots, X_n, G]$ is independent of G, that is,

$$\text{Prob}\,\{S[X_1, \cdots, X_n, G] \leqq s; G\} = \varphi(s)$$

where $\varphi(s)$ depends only on s.

It is easily verified that if a statistic has structure (d) then it is distribution-free in the family Ω_2 of all continuous cumulative distribution functions. It can, however, be shown by counter-example [15] that not every distribution-free symmetric statistic in Ω_2 is of structure (d).

For every continuous cumulative probability function $G(x)$ an inverse function $G^{-1}(u)$ may be uniquely defined in $0 \leqq u \leqq 1$ by setting

$G^{-1}(0) = -\infty$

$G^{-1}(u) =$ greatest lower bound of x such that $G(x) = u$, for $0 < u \leqq 1$.

If $G(x) < 1$ for all real x, this definition shall mean $G^{-1}(1) = +\infty$.

Let $F(x)$ and $G(x)$ be two continuous cumulative probability functions. The function $\tau(u) = F\,G^{-1}(u)$, $0 \leqq u \leqq 1$, constitutes a monotone mapping of the unit-interval into itself. If that mapping is the identity, i.e. if $\tau(u) = u$, then and only then $F(x) = G(x)$, and therefore the function $\tau(u) - u$ may be interpreted as a very detailed description of the discrepancy between $G(x)$ and $F(x)$.

Let Ω and Ω' be two families of continuous cumulative probability functions. We shall say that $S[X_1, \cdots, X_n, G]$ is *strongly distribution-free in Ω with respect to Ω'* if

(i) for every $G \in \Omega$ it is a measurable function of $X_1, \cdots, X_n$, and

(ii) the probability distribution of $S[X_1, \cdots, X_n, G]$, where $X_1, \cdots, X_n$ are a sample of a random variable X with a cumulative probability function $F \in \Omega'$ and G is any element of Ω, depends only on $F\,G^{-1}$, that is,

$$\text{Prob}\ \{S[X_1, \cdots, X_n, G] < s;\ F\} = \psi(s, FG^{-1}).$$

It is obvious that if $S[X_1, \cdots, X_n, G]$ is strongly distribution-free in Ω with respect to Ω' and if $\Omega \subset \Omega'$, then $S[X_1, \cdots, X_n, G]$ is distribution-free in Ω.

The following theorem can be proven [15]. If Ω is the class of all strictly monotone continuous cumulative probability functions, $\Omega' = \Omega$, and $S[X_1, \cdots, X_n, G]$ is symmetric in $X_1, \cdots, X_n$ and strongly distribution-free, then it has structure (d).

4. On choosing a distribution-free test of fit.

4.1. *Need and criteria for making a choice.* A number of distribution-free tests of fit have been described in Section 2, a number of other such tests have been proposed in literature, and still more can be constructed by selecting additional statistics of structure (d) and adapting them for the use in tests of fit. For any such statistic $S[X_1, \cdots, X_n, H]$ the probability distribution, if $H(x)$ is true, does not depend on $H(x)$. It may therefore be assumed that $H(x)$ is the uniform probability distribution, and under this assumption it is usually possible to write the cumulative probability function of the statistic in form of a multiple integral of a constant integrand over a polyhedral region. This integral can sometimes be evaluated explicitly, sometimes it can be reduced to a system of recursion formulae, or it may be possible to tabulate it numerically with the aid of modern computing equipment.

The statistician is, therefore, faced with the problem of deciding in concrete situations on using one of the already known tests for which the necessary tables are available; or, on a more theoretical level, he may have to decide which of the many more possible tests deserve to be studied and developed.

Besides obvious reasons of expediency, such as availability of tables, ease of computation, simplicity in use by untrained personnel, the statistician will have to consider various properties which make some of the tests theoretically more or less advantageous. For example, having to choose between the chi square and the Kolmogorov test, he will have to consider that the former can be adjusted to parametric families where the parameters are to be estimated from the sample, while the latter has the advantage that it requires no arbitrary grouping of observations and that the exact probability distribution for finite sample sizes has been extensively tabulated.

It would appear, however, that the most essential preliminary consideration should be to determine what kind of discrepancy between hypothesis and alternative is materially important in a concrete situation. Then one may attempt to select a test best capable of detecting this kind of discrepancy. For example the chi square test is sensitive for discrepancies in the histograms, while the Kolmogorov test appears more likely to detect vertical discrepancies between the cumulative probability functions.

In order to judge how good a distribution-free test of fit is for a definite problem, one has therefore first to decide on a way to measure discrepancies between distributions by introducing in the appropriate space of probability distributions a distance which may either satisfy the axioms of a Hausdorff metric or some other set of postulates. Once it is defined, it may be possible to study the power of various tests with regard to this distance and to select the test which is optimal with regard to some of the well known properties based on the power.

4.2. *Distances based on* $\tau(u) = F\,G^{-1}(u)$. Since practically all distribution-free statistics used for tests of fit are strongly distribution-free, it seems desirable to use a metric which ascribes the same distance $\delta(F, G)$ to all pairs F, G for which $\tau(u) = F\,G^{-1}(u)$ is the same. Examples of such distances are

$$\sqrt{\int_0^1 [\tau(u) - u]^2\, du} = \sqrt{\int_{-\infty}^{+\infty} [F(x) - G(x)]^2\, dG(x)}, \tag{4.21}$$

$$\int_0^1 |\tau(u) - u|\, du = \int_{-\infty}^{+\infty} |F(x) - G(x)|\, dG(x), \tag{4.22}$$

$$\sup_{0<u<1} |\tau(u) - u| = \sup_{-\infty<x<+\infty} |F(x) - G(x)|, \tag{4.23}$$

$$\sup_{0<u<1} [\tau(u) - u] = \sup_{-\infty<x<+\infty} [F(x) - G(x)]. \tag{4.24}$$

While (4.23) defines a Hausdorff metric, the other three expressions define directed distances $\overrightarrow{FG}$.

On intuitive grounds one would be inclined to use Smirnov's ω^2 statistic (2.21), (2.22) if discrepancies described by metric (4.21) are considered important, Kolmogorov's statistic (2.51) for the metric (4.23), the statistic D_n^+ in (2.81) if (4.24) is the discrepancy that matters, etc. A systematic treatment would possibly require the introduction of a distance which depends on n.

Very few attempts have been made at studies in this direction. Mann and Wald [16] investigated the power of the chi square test with regard to the distance (4.23). An elaboration of their results, together with useful numerical tabulations, was made by Williams [17]. A comparison of the power of the chi square test with that of Kolmogorov's test, based on the metric (4.23), was made by Massey [7] who, as one would expect, found the latter test more powerful.

4.3. *Power of tests using strongly distribution-free statistics.* Let $S(X_1, \cdots, X_n, G)$ be a strongly distribution-free statistic in Ω with respect to Ω' and $\Phi[S(X_1, \cdots, X_n, H)]$ a (randomized) test function for the simple hypothesis $H(x)$, that is, a function such that (1) $0 \leqq \Phi[S(X_1, \cdots, X_n, H)] \leqq 1$ for all $X_1, \cdots, X_n$ in the sample space and every $H \epsilon \Omega$, and (2) $\Phi[S(X_1, \cdots, X_n, H)]$ is the probability of rejecting H. Then, the power of the test depends only on $\tau(u) = FH^{-1}(u)$, that is we have

$$E\{\Phi[S(X_1, \cdots, X_n, H)]; F\} = \psi[\tau(u)]$$

for all $H \epsilon \Omega$, $F \epsilon \Omega'$.

Let $\delta(F, G)$ be a distance depending only on $\tau(u)$, such as the distances described in 4.2. For fixed H one may consider the "sphere" consisting of all $F \epsilon \Omega'$ such that $\delta(F, H) = \delta_0$, and try to determine the greatest lower bound of the power for all these F, which will depend only on the distance δ_0:

$$\inf_{\delta(F,H)=\delta_0} E\{\Phi[S(X_1, \cdots, X_n, H)]; F\} = \mu(\delta_0).$$

A problem of this kind has been treated in [18] where a sharp lower (and upper) bound for the power is obtained for a one-sided test of fit, with regard to the directed metric (4.24).

4.4. *The alternative described in terms of a distance.* The problem of testing the simple hypothesis that the true cumulative probability function is exactly equal to a completely specified $H(x)$ is, in this formulation, somewhat unrealistic. In fact, it leads to the known difficulty that if a consistent test is used, for a sufficiently large sample one will practically always reject the hypothesis. It is well known that this difficulty can be avoided by stating carefully the hypothesis and the alternative. If a distance $\delta(F, G)$ is defined in Ω, this can be done by considering the simple hypothesis $H(x)$ and the composite alternative consisting of all $F \epsilon \Omega$ such that $\delta(F, H) \geqq \delta_1$. With this formulation of the problem, the hypothesis will be rejected only if the test produces empirical evidence that the true distribution differs from H by a distance of at least δ_1.

REFERENCES

[1] W. G. Cochran, "The χ^2 test of goodness of fit," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 315–345.

[2] N. Smirnov, "Sur la distribution de ω^2," *C. R. Acad. Sci. Paris*, Vol. 202 (1936), pp. 449–452.

[3] T. W. Anderson, and D. A. Darling, "Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 193–212.

[4] A. Kolmogorov, "Sulla determinazione empirica di una legge di distribuzione," *Giorn. Ist. Ital. Attuari*, Vol. 4 (1933), pp. 83–91.

[5] N. Smirnov, "Table for estimating the goodness of fit of empirical distributions," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 279–281.

[6] Frank J. Massey, Jr., "A note on the estimation of a distribution function by confidence limits," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 116–119.

[7] Frank J. Massey, Jr., "The Kolmogorov-Smirnov test for goodness of fit," *J. Amer. Stat. Assn.*, Vol. 46 (1951), pp. 68–78.

[8] Z. W. BIRNBAUM, "Numerical tabulation of the distribution of Kolmogorov's statistic for finite sample size," *J. Amer. Stat. Assn.*, Vol. 47 (1952), pp. 425–441.

[9] A. WALD AND J. WOLFOWITZ, "Confidence limits for continuous distribution functions," *Ann. Math. Stat.*, Vol. 10 (1939), pp. 105–118.

[10] B. F. KIMBALL, "Some basic theorems for developing tests of fit for the case of the non-parametric probability distribution function, I," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 540–548.

[11] P. A. P. MORAN, "The random division of an interval," *J. Roy. Stat. Soc., Suppl.*, Vol. 9 (1947), pp. 92–98.

[12] B. SHERMAN, "A random variable related to the spacing of sample values," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 339–361.

[13] Z. W. BIRNBAUM AND F. H. TINGEY, "One-sided confidence contours for probability distribution functions," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 592–596.

[14] N. SMIRNOV, "Sur les ecarts de la courbe de distribution empirique," *Rec. Math.* (*Mat. Sbornik*), N. S. Vol. 6 (1939), pp. 3–26.

[15] Z. W. BIRNBAUM AND H. RUBIN, "On distribution-free statistics," unpublished.

[16] H. B. MANN AND A. WALD, "On the choice of the number of class intervals in the application of the chi square test," *Ann. Math. Stat.*, Vol. 13 (1942), pp. 306–317.

[17] C. A. WILLIAMS, JR.. "On the choice of the number and width of classes for the chi square test of goodness of fit," *J. Amer. Stat. Assn.*, Vol. 45 (1950), pp. 77–86.

[18] Z. W. BIRNBAUM, "On the power of a one-sided test of fit for continuous probability functions," submitted for publication.