Biometrika (1983), **70**, 3, *pp.* 723–6
Printed in Great Britain

A test for normality based on the empirical characteristic function

By T. W. EPPS
Department of Economics, University of Virginia, Charlottesville, Virginia, U.S.A.

AND LAWRENCE B. PULLEY
Department of Economics, Brandeis University, Waltham, Massachusetts, U.S.A.

Summary

An omnibus test of normality is proposed, which has high power against many alternative hypotheses. The test uses a weighted integral of the squared modulus of the difference between the characteristic functions of the sample and of the normal distribution.

Some key words: Moment generating function; Normal distribution; Test of fit.

1. Introduction

If $X_1, \ldots, X_n$ is a random sample from a distribution $F(x)$, the empirical characteristic function is defined as $\phi_n(t) = n^{-1} \Sigma_j \exp(itX_j)$, where t is an arbitrary real parameter. The almost sure convergence of $\phi_n(t)$ to $\phi(t)$, the characteristic function of the population, together with the one-to-one correspondence between $\phi(t)$ and $F(x)$, suggest utilizing $\phi_n(t)$ in statistical inference. Here we use $\phi_n(t)$ to test the composite hypothesis that $F(x)$ is normal.

Let $\phi_0(t) = \exp(it\mu - \frac{1}{2}t^2\sigma^2)$ be the characteristic function under the null hypothesis, with μ and σ^2 the unspecified mean and variance. The test introduced here is based on a weighted integral over t of the squared modulus of $\phi_n(t) - \hat{\phi}_0(t)$, where $\hat{\phi}_0(t)$ depends on sample estimates of μ and σ^2. We compare the power of the test with that of prominent tests based on order statistics and on sample moments.

Heathcote (1972) and Feigin & Heathcote (1977) have previously considered using either the real or imaginary part alone of $\phi_n(t)$ in tests of simple hypotheses. Their tests relied on the fact that, for given t, either component of $\phi_n(t)$ is asymptotically normal with mean given by its population counterpart. If the alternative hypothesis is also simple, it may be possible to find a value of t that maximizes the power of the test in large samples. Epps, Singleton & Pulley (1982) showed how the sample moment generating function may be employed to test composite hypotheses that the data come from one or the other of two separate families of distributions.

Murota & Takeuchi (1981) have recently proposed a location and scale invariant test based on the statistic $\tilde{a}_n(t) = |\phi_n(t/S)|^2$, where S is the sample standard deviation. Applied to the problem of testing normality, $\tilde{a}_n(t)$ was found to have high power in the vicinity of $t = 1{\cdot}0$ against members of six families of symmetric alternatives, but its power against skew distributions was not demonstrated. In seeking to develop an omnibus test for normality, we have studied extensively a related test involving the statistic

$$M(t) = n^{-1} \sum_j \exp\{t(X_j - \bar{X})/S\},$$

the moment generating function of the standardized sample. We found that, especially for skew distributions, power often varied markedly with t, and that no single t value produced a satisfactory omnibus test.

Several tests based on more than one value of t have been proposed. Exploiting the joint normality of real and imaginary parts of $\phi_n(\,.\,)$ evaluated at each of several t values, Koutrouvelis (1980) and Koutrouvelis & Kellermeier (1981) construct test statistics distributed asymptotically as chi-squared under the null hypothesis. Heathcote (1972) has suggested a test based on the supremum over t of $|\phi_n(t)-\phi_0(t)|$. Although t is a continuous parameter, the supremum would have to be relative to a finite set, $(t_1, \ldots, t_J)$. This procedure has considerable intuitive appeal, but it shares with the chi-squared tests a practical difficulty. There seems to be no theoretical way of choosing the set to produce acceptable power against a wide range of distributions, and to do this by trial and error seems a hopeless task. A more tractable approach follows from the suggestion by Feuerverger & Mureika (1977) that a statistic of the form $\int \{\operatorname{Im} \phi_n(t)\}^2 \, dG(t)$, where the integral is over $(-\infty, \infty)$, be used to test for symmetry of $F(x)$. Here the weight function $G(t)$ is itself a distribution function, with $G'(t)$ symmetric about the origin. Depending on the choice of $G(t)$ this statistic can manifest the behaviour of $\phi_n(t)$ over any desired range of t and may take on a form which is computationally tractable.

2. Derivation of the test statistic

Our proposed test of the composite hypothesis of normality is based on the integral

$$T = \int_{-\infty}^{\infty} |\phi_n(t) - \hat{\phi}_0(t)|^2 \, dG(t), \tag{1}$$

where $\hat{\phi}_0(t) = \exp(it\bar{X} - \tfrac{1}{2}t^2 S^2)$, $\bar{X} = n^{-1} \Sigma X_j$ and $S^2 = n^{-1} \Sigma (X_j - \bar{X})^2$.

Three considerations influence the choice of $G(t)$. The first is that it assign high weight where $|\phi_1(t) - \phi_0(t)|$ is large, $\phi_1(t)$ pertaining to the alternative hypothesis. If put in standard form, many continuous distributions may be seen to have large values of $|\phi_1(t) - \phi_0(t)|$ in the interval $0 < t < 3$. The second is that $G(t)$ give high weight where the statistic $\phi_n(t)$ is a relatively precise estimator of $\phi(t)$. It may be shown that

$$E\{|\phi_n(t) - \phi(t)|^2\} = n^{-1}\{1 - |\phi(t)|^2\}.$$

Since $|\phi(0)| = 1$ and, for any continuous distribution, $\lim |\phi(t)| = 0$ as $t \to \infty$, it is apparent that the precision of the sample characteristic function is greatest near the origin and ultimately declines as one moves away from the origin. Thus, it appears that $G(t)$ should assign high weight in some interval around the origin. It should be clear from the definitions of $\phi_n(t)$ and $\phi(t)$ that the length of such an interval will depend inversely on the scale of the data. A final practical consideration is that $G(t)$ be such that the integral (1) have a closed form.

With $dG(t) = g(t)\,dt$, consider the function

$$g(t) = \{\alpha S/\sqrt{(2\pi)}\} \exp(-\tfrac{1}{2}\alpha^2 S^2 t^2), \tag{2}$$

which is a normal density with zero mean and variance $(\alpha S)^{-2}$. Here $\alpha > 0$, and S estimates the scale of the distribution being sampled. The first two requirements of the weighting function are met, given a judicious choice of α; moreover, it yields the closed form for (1):

$$T(\alpha) = n^{-2} \sum_{j=1}^{n} \sum_{k=1}^{n} \exp\{-\tfrac{1}{2}(X_j - X_k)^2/(\alpha^2 S^2)\}$$
$$- 2n^{-1}(1+\alpha^{-2})^{-\frac{1}{2}} \sum_{j=1}^{n} \exp[-\tfrac{1}{2}(X_j - \bar{X})^2/\{S^2(1+\alpha^2)\}] + (1+2\alpha^{-2})^{-\frac{1}{2}}. \tag{3}$$

Also $T(\alpha)$ is invariant under changes in the location and scale of the data. For reasons given below, it is convenient to take as the test statistic $T^*(\alpha) = -\log\{nT(\alpha)\}$.

Monte Carlo methods were used to determine the fractiles of $T^*(\alpha)$ under H_0. Initially, for a few sample sizes we computed percentage points for each of eight values of α ranging from 0·1 to 2·5. We then made a preliminary study of the powers of these eight tests against a group of distributions with diverse levels of skewness and kurtosis. The results led us to select the values $\alpha = 0{\cdot}5$, 0·7, 1·0 and 1·3 for further study. For each of these choices of α we calculated 10,000 values of $T^*(\alpha)$ from normal samples of size $n = 5\,(2)\,20\,(5)\,50$; 6000 values from samples of size 60 (10) 100; and 3000 values from samples of size 120 (20) 200. Selected fractiles of the simulated null distributions of $T^*(0{\cdot}7)$ and $T^*(1{\cdot}0)$ are given in Table 1. The transformation to T^* and the retention of the final constant term in (3) facilitate tabulation.

Table 1. *Selected empirical fractiles of* $T^*(\alpha)$, *for* $\alpha = 0{\cdot}7$, 1·0

Sample size	$\alpha = 0{\cdot}7$ 0·025	0·050	0·950	0·975	$\alpha = 1{\cdot}0$ 0·025	0·050	0·950	0·975
4	2·90	2·76	0·42	0·33	4·22	3·91	1·23	1·15
6	3·07	2·79	0·39	0·24	4·34	4·00	1·13	0·90
8	3·16	2·91	0·38	0·22	4·39	4·03	1·09	0·86
10	3·24	2·96	0·37	0·19	4·43	4·08	1·06	0·84
12	3·29	2·98	0·36	0·17	4·44	4·09	1·03	0·82
>12	3·30	3·00	0·35	0·17	4·45	4·10	1·00	0·79

3. Power of the test

Table 2 presents estimates of power of $T^*(\alpha)$ against ten alternative distributions, for which the usual dimensionless measures of skewness and kurtosis, $\sqrt{\beta_1}$ and β_2, are shown. Definitions of the systems of distributions and a key to symbols are in Tables 5 and 6 of Pearson, D'Agostino & Bowman (1977).

Table 2. *Percentage powers of level* 0·05 *of* $T^*(\alpha)$, *Shapiro–Wilk,* W *and* R *tests*

H_1	$\sqrt{\beta_1}$	β_2	$n = 20$ $T^*(0{\cdot}7)$	$T^*(1{\cdot}0)$	W	R	$n = 50$ $T^*(0{\cdot}7)$	$T^*(1{\cdot}0)$	W	R
Beta (2, 1)	−0·57	2·4	16	15	29	19	62	63	87	37
$S_B\,(1, 1)$	0·73	2·9	21	21	30	17	60	65	84	47
$S_U\,(-1, 2)$	0·87	5·6	15	15	21	21	25	33	35	39
$\chi^2\,(4)$	1·41	6·0	29	39	55	39	83	87	96	87
LN	6·18	113·9	84	89	93	94	100	100	100	100
Beta (1, 1)	0·00	1·8	8	4	16	20	45	34	86	80
$t\,(10)$	0·00	4·0	7	7	11	9	13	15	13	19
Laplace	0·00	6·0	21	23	34	35	42	43	51	59
$S_U\,(0, 1)$	0·00	36·2	34	38	44	42	69	72	67	76
Cauchy	—	—	83	82	88	85	100	100	100	100

S_B and S_U denote Johnson's S_B and S_U systems; LN denotes standard log normal. Complete descriptions of the distributions and conventions in the use of parameters are given by Pearson *et al.* (1977).

Besides 0·05 level powers of $T^*(\alpha)$ for two values of α, Table 2 gives estimates of power of the Shapiro–Wilk (1965) W test and the R test. The R test is the simpler of two omnibus tests based on the sample skewness, $\sqrt{b_1}$, and kurtosis, b_2, proposed by Pearson *et al.* (1977) and rejects H_0 when rejected by either $\sqrt{b_1}$ or b_2 alone, critical points of these statistics being adjusted to give the correct level.

Table 2 summarizes a small part of the work. For skew alternatives the power of $T^*(\alpha)$ against platykurtic distributions usually declines as α increases. For leptokurtic distributions the power increases slightly with α. At $\alpha = 0{\cdot}7$ or $\alpha = 1{\cdot}0$ the powers of $T^*(\alpha)$ and R are in most cases comparable and usually somewhat lower than those of W. Also $T^*(\alpha)$ has higher power than R against short-tailed, moderately skew distributions. For symmetric alternatives, $T^*(\alpha)$ has less power than either W or R against platykurtic distributions, but power improves markedly when α is reduced below 1·0. Against mesokurtic and leptokurtic distributions the power of $T^*(\alpha)$ increases with α, and at $\alpha = 0{\cdot}7$ or 1·0 it is comparable to those of the other tests.

With $\alpha = 0{\cdot}7$ or 1·0 the test based on $T^*(\alpha)$ appears to merit use as an omnibus test of normality. It offers a slight computational advantage over the Shapiro–Wilk test in that the data need not be ordered and in that tabulated constants are not required; and, unlike the R test, only two critical points are needed for each sample size. Moreover, the test allows one to make use of prior information about the population: setting $\alpha = 0{\cdot}5$ tends to raise power against platykurtic distributions, whether symmetric or not; while $\alpha = 1{\cdot}3$ is appropriate for the leptokurtic class. The test is not recommended, however, if high power is required against distributions which are short-tailed and symmetric.

A longer version of this paper is available containing additional fractiles of $T^*(\alpha)$ for all four values of α, a FORTRAN IV subroutine to calculate $T^*(\alpha)$, extensions of the power study to 76 distributions, and details of the Monte Carlo work.

References

EPPS, T. W., SINGLETON, K. J. & PULLEY, L. B. (1982). A test of separate families of distributions based on the empirical moment generating function. *Biometrika* **69**, 391–99.

FEIGIN, P. D. & HEATHCOTE, C. R. (1976). The empirical characteristic function and the Cramer–von Mises statistic. *Sankhyā* A **38**, 309–25.

FEUERVERGER, A. & MUREIKA, R. A. (1977). The empirical characteristic function and its applications. *Ann. Statist.* **5**, 88–97.

HEATHCOTE, C. R. (1972). A test of goodness of fit for symmetric random variables. *Aust. J. Statist.* **14**, 172–81.

KOUTROUVELIS, I. A. (1980). A goodness-of-fit test of simple hypotheses based on the empirical characteristic function. *Biometrika* **67**, 238–40.

KOUTROUVELIS, I. A. & KELLERMEIER, J. (1981). A goodness-of-fit test based on the empirical characteristic function when parameters must be estimated. *J. R. Statist. Soc.* B **43**, 173–6.

MUROTA, K. & TAKEUCHI, K. (1981). The studentized empirical characteristic function and its application to test for the shape of distribution. *Biometrika* **68**, 55–65.

PEARSON, E. S., D'AGOSTINO, R. B. & BOWMAN, K. O. (1977). Tests for departure from normality: Comparison of powers. *Biometrika* **64**, 231–46.

SHAPIRO, S. S. & WILK, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika* **52**, 591–611.

[*Received May* 1982. *Revised March* 1983]