The results obtained above are then used for discussing some hypotheses testing problems, and also establishing the asymptotic efficiency (in the Weiss–Wolfowitz sense) of the maximum probability estimates.

Finally, it is indicated that, under suitable regularity conditions, the general results mentioned above can be extended to the following cases: The r.v.'s involved are independent but not necessarily identically distributed; they are coming from a stationary and ergodic Markov process; they are coming from a fairly general stochastic process.

References

[1] J. L. Doob, *Stochastic processes*, John Wiley and Sons, Inc., New York 1953.

[2] J. Hájek, *Limiting properties of likelihoods and inference*, in: *Foundations of statistical inference*, Editors V.P. Godambe and D. A. Sprott, Holt, Rinehart, Winston of Canada 1971, pp. 142–162.

[3] L. LeCam, *Locally asymptotically normal families of distributions*, Univ. California Publ. Statist. 3 (1960), pp. 37–98.

[4] M. Loève, *Probability theory*, 3rd ed., Van Nostrand Reinhold Co., New York 1963.

[5] A.N. Philippou and G.G. Roussas, *Asymptotic distribution of the likelihood function in the independent not identically distributed case*, Ann. Math. Statist. 1 (1973), pp. 454–471.

[6] G. G. Roussas, *Contiguity of probability measures: Some applications in statistics*, Cambridge Univ. Press. 1972.

[7] —, *Asymptotic properties of maximum probability estimates in the i.i.d. case*, Proceedings of the Institute of Mathematical Statistics Summer Institute, Indiana Univ., Bloomington, Indiana 1975.

[8] —, *Asymptotic efficiency of the maximum probability estimate in the i.i.d. case*, Bull. Soc. Math. Grèce 17 (1976), pp. 28–37.

[9] —, *Asymptotic distribution of the log-likelihood function for stochastic processes*, Z. Wahrscheinlichkeistheorie und verw. Gebiete 47 (1979), pp. 31–46.

[10] G.D. Stamatelos, *Asymptotic distribution of the log-likelihood function for stochastic processes: Some examples*, Bull. Soc. Math. Grèce 17 (1976), pp. 92–116.

[11] L. Weiss and J. Wolfowitz, *Maximum probability estimators and related topics*, Lecture Notes in Mathematics, Springer-Verlag, Berlin–Heidelberg–New York 1974.

Presented to the semester
MATHEMATICAL STATISTICS
September 15–December 18, 1976

MATHEMATICAL STATISTICS
BANACH CENTER PUBLICATIONS, VOLUME 6
PWN—POLISH SCIENTIFIC PUBLISHERS
WARSAW 1980

TESTING FOR NORMALITY

K. SARKADI

Mathematical Institute of the Hungarian Academy of Sciences,
H-1053 Budapest V, Hungary

Let $X_1, \ldots, X_n$ denote identically and independently distributed random variables with the distribution function $F(x)$.

Let H_0 denote the hypothesis

$$H_0\colon\ F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right),$$

where $\Phi(x)$ is the standardized normal distribution function. The constants μ and σ are unspecified (nuisance parameters).

Tests for H_0 are called (one-sample) *normality tests*. We deal with such tests in Section 1.

Sometimes we have more than one sample, with different nuisance parameters, for assessing normality. The corresponding tests are called multisample normality tests; they are discussed in Section 2.

1. One sample case

Any goodness of fit test can be used as a test for normality if the empirical moments are substituted in the theoretical distribution function. This modification, however, changes the distribution of the test statistics. For the χ^2-test, the same tables may be used, only the number of the degrees of freedom is to be diminished by the number of estimated constants (Fisher [11] (1924)). Although this solution is of approximative character (see Chernoff and Lehmann [3] (1954)) the accuracy is sufficient in most of the practical cases provided the sample size is large. For other goodness of fit tests, separate tables have been prepared for the modified case. For the Kolmogorov test, the critical values have been tabulated by Lillefors [13] (1967), for the Cramér–Mises, Anderson–Darling and some other tests by Stephens [30] (1974), see Biometrika Tables, Volume 2, Table 54.

A further group of normality tests are the tests of Shapiro and Wilk [27] (1965), Shapiro and Francia [26] (1972), DeWet and Venter [7] (1972) and d'Agostino

[4] (1971). Each of these tests has the test statistic

$$W = \frac{\left(\sum_{i=1}^{n} c_i X_{(i)}\right)^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$$

but the definition of the constants c_i is different for each test. Here $X_{(1)} \leqslant \ldots \leqslant X_{(n)}$ are the ordered sample elements. The constants and the critical values are tabulated for $n \leqslant 50$ in the case of the Shapiro–Wilk-test, and for $50 \leqslant n \leqslant 99$ in the case of the Shapiro–Francia-test. The constants may easily be calculated in the case of the d'Agostino-test and the deWet–Venter-test and the asymptotic critical values are given.

The tests of Shapiro–Francia and DeWet–Venter are consistent, the d'Agostino test is not. The consistency of the Shapiro–Wilk test is an open question; its version to test the departure from the exponential distribution [28] is not consistent [23], [24].

Further tests to be mentioned are the tests based on the higher moments, in particular, the test of Bowman and Shenton [2] (1975), and modifications of the χ^2-test: Moore [15] (1971), Nikulin [16] (1973).

Shapiro, Wilk and Chen [29] (1968) have carried out extensive Monte Carlo experiments in which the most important tests of normality were compared. From the general practical point of view, a good test should have reasonably good power against all practically important alternatives. A number of such alternatives have been considered. This work, if considered together with the similar computations of Dyer [10] (1974) and Shapiro and Francia [26] (1972), suggests that the tests of Shapiro–Wilk, Shapiro–Francia and Anderson–Darling are practically the best, these three matching each other. Since the test statistic of DeWet–Venter lies very near to that of Shapiro–Francia, we may think that this fourth test shares the good properties of the former ones.

But power is not the only aspect to be taken into account. This author believes that the advantages of the χ^2-test make it still recommendable if the number of data is large. It is particularly suitable if the observations are grouped and this case frequently occurs in practice. The calculation work can be suitably connected with histogram representation.

2. Multisample case

In many cases it is difficult or impossible to provide a sufficiently large series of observations for assessing normality but it is possible to take observations in short series or such data are available from the past.

Let X_{ij} $(i = 1, \ldots, r;\ j = 1, \ldots, n_i)$ denote independent random variables. The distribution function of X_{ij} is $F[(x-\mu_i)/\sigma_i]$ where μ_i and σ_i are, in general, unknown, but in special cases, they may be connected by some relations which are known.

The null hypothesis is that $F(x)$ is the standardized normal distribution function. The alternative hypothesis is that $F(x)$ is not normal.

We consider below two cases: the case of common variance (homoscedastic case) where $\sigma_1 = \ldots = \sigma_r = \sigma$ and the heteroscedastic case where no relation among the nuisance parameters is supposed.

2.1. *Homoscedastic case.* We suppose $n_i \geqslant 2$ $(i = 1, \ldots, r)$, $\sigma_1 = \ldots = \sigma_r = \sigma$.

Our procedure [21], [22] consists of two steps. The first one is the following transformation:

$$Y_{ij} = X_{ij} - U_i,$$

where

$$U_i = \frac{\sum_{j=1}^{n_i-1} X_{ij}}{n_i + \sqrt{n_i}} + \frac{X_{in_i}}{\sqrt{n_i}} \quad (i = 1, \ldots, r;\ j = 1, \ldots, n_i - 1).$$

Before the application of the above transformation it should be ascertained whether the observations are, in fact, in random order. If there is any doubt, we have to change the order of the elements within each sample, by putting a randomly chosen element at the end (the element X_{in_i} has a special role in the transformation). The coefficients $1/(n+\sqrt{n})$ and $1/\sqrt{n}$ are tabulated, up to the sample size 20, in [20].

The number of transformed elements is $\sum_{i=1}^{r} n_i - r$. If the null hypothesis is true, they are independent and normally distributed random variables with expectation 0 and variance σ^2.

The second step of the procedure is the test of the normality of the last-mentioned variables. For choosing the appropriate test we refer to the aspects mentioned in Section 1.

In this step it is better not to use the information that the expectation is 0, see Dyer [10] (1972), Deutler et al. [6] (1975).

2.2. *Heteroscedastic case.* Now we suppose $n_i \geqslant 3$ $(i = 1, \ldots, r)$. No relation between the nuisance parameters $\mu_1, \ldots, \mu_r, \sigma_1, \ldots, \sigma_r$ is supposed.

Now again the procedure consists of two steps. The first step is the following transformation [22]:

$$Y_{ij} = (X_{ij} - U_i)\frac{S_{Yi}}{S_i},$$

where now

$$U_i = \frac{\sum_{i=1}^{n_i-2} X_{ij}}{n_i + \sqrt{2n_i}} + \frac{X_{in_i-1} + X_{in_i}}{\sqrt{2n_i}},$$

$$S_i = \sqrt{\sum_{j=1}^{n_i-2} (X_{ij} - U_i)^2},$$

$$S_{Yi} = \psi_{n_i}\left(\frac{S_i}{|X_{in_i-1} - X_{in_i}|}\right) \quad (i = 1, \ldots, r;\ j = 1, \ldots, n_i-2).$$

$\psi_n(x)$ is a monotone increasing function and is such that $S_{Yi} \overset{\text{distr}}{=} S_i/\sigma_i$ in the case of the null hypothesis. The function $\psi_n(x)$ is fully determined by these conditions. It can be determined with the help of the tables of Student and χ^2 distribution functions, see [22].

Tabulation of the function $\psi_n(x)$ and of the coefficients $1/(n+\sqrt{2n})$ and $1/\sqrt{2n}$ is under way [24].

If there is any doubt regarding randomness, at least the last two elements in each sample (X_{in_i-1} and X_{in_i}) are to be randomly selected.

If the null hypothesis is true, the resulting $\sum_{i=1}^{r} n_i - 2r$ variables Y_{ij} ($i = 1, \ldots, r$; $j = 1, \ldots, n_i-2$) will be independently distributed with standard normal distribution. The second step will test whether they are normal. The aspects for the choice among the available tests is the same as before.

2.3. We may add the following remarks:

Some simplification can be made for the case where n_i takes on its smallest possible value ($n_i = 2$ in Section 2.1 and $n_i = 3$ in Section 2.2).

The 1960 paper of the present author [21] contains formula (2.1) but the transformation given for the heteroscedastic case differs from (2.2). Formula (2.2) (see [22], formula (7.1)) is a slight modification of the transformation proposed by Störmer [31] (1964).

Formerly, Petrov [19] (1951) proposed a multisample test of normality, see also Dunin–Barkovsky and Smirnov [8] (1955), pp. 354–360. The transformed values of Petrov, however, were not independent; therefore, the use of only one transformed value per sample was suggested. [21] was an improvement of the results of Petrov, permitting higher efficiency in the multisample case and appropriate for the one-sample case. It was the first exact test of normality in the one-sample case.

Further tests of normality which use transformation of sample elements are those of Durbin [9] (1961), pp. 49–54, O'Reilly and Quesenberry [17] (1973), Csörgő et al. [5] (1973), Major (Major and Tusnády [14] (1974), p. 277). Durbin uses random numbers. The transformations of Störmer, Durbin and the present author preserve the shape of the empirical distribution function. An optimality property of the transformation here proposed has been proved in [22].

Tusnády [25], Section 6 generalized the method of Störmer for the multidimensional case.

A multisample test of normality of another kind is that of Wilk and Shapiro [29] (1968). This method consists of two steps. In the first step, however, only one statistic per sample is provided, which does not preserve the information regarding the shape of the distribution. This test is appropriate if the alternative hypothesis is complete non-homogeneity (the parent distributions are non-homogeneous in their shape).

2.4. *Characterization.* Transformation (2.1) characterizes the normal distribution in the sense that the normality of the transformed variables implies that of the initial variables. This easily follows from the well-known theorem of Cramér.

Below we give a characterization theorem which is applicable to heteroscedastic transformations.

In the following let $X_1, \ldots, X_6$ denote independent, identically distributed random variables, and U and V arbitrary random variables.

THEOREM. *If*

$$(X_i, X_j, X_k, X_l, V) \overset{\text{distr}}{=} (X_1, X_2, X_3, X_4, V),$$

where $1 \leqslant i, j, k, l \leqslant 6$, i, j, k, l *are all different, and the distribution of*

$$\left(\frac{X_1 - U}{V}, \ldots, \frac{X_6 - U}{V}\right)$$

is normal, then $X_1, \ldots, X_6$ *are normally distributed.*

Proof. Let $W_i = X_{2i-1} - X_{2i}$, $Z_i = W_i/V$ ($i = 1, 2, 3$). It follows that the distributions of Z_1, Z_2 and Z_3 are identical and normal and, moreover, that $(Z_i, Z_j) \overset{\text{distr}}{=} (-Z_i, Z_j)$, $i = 1, 2$; $j = i+1, \ldots, 3$, which implies that Z_1, Z_2, Z_3 are independent and have the expectation 0. Since

$$\left(\frac{Z_1}{|Z_3|}, \frac{Z_2}{|Z_3|}\right) = \left(\frac{W_1}{|W_3|}, \frac{W_2}{|W_3|}\right)$$

and W_1, W_2, W_3 are independent and identically distributed, we conclude from Theorem 13.5.2 of Kagan, Linnik and Rao [12] (1973) that the distribution of W_1, W_2, W_3 is normal. The normality of $X_1, \ldots, X_6$ then follows from the theorem of Cramér.

References

[1] T. W. Anderson and D. A. Darling, *A test of goodness of fit*, J. Amer. Statist. Assoc. 49 (1954), pp. 765–769.

[2] K. O. Bowman and L. R. Shenton, *Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2*, Biometrika 62 (1975), pp. 243–250.

[3] H. Chernoff and E. L. Lehmann, *The use of maximum likelihood estimates in χ^2 tests for goodness of fit*, Ann. Math. Statist. 25 (1954), pp. 579–586.

[4] R. B. D'Agostino, *An omnibus test for normality of moderate and large size samples*. Biometrika 58 (1971), pp. 341–348.

[5] M. Csörgő, V. Seshadri, and M. Yalovsky, *Some exact tests for normality in the presence of unknown parameters*, J. Roy. Statist. Soc. Ser. B 35 (1973), pp. 507–522.

[6] T. Deutler, H. Griesenbrock, and H. Schwensfeier, *Der Kolmogorov–Smirnov–Einstichprobentest auf Normalität*, Allgemein. Statist. Arch. 59 (1975), pp. 228–250.

[7] T. DeWet and J. H. Venter, *Asymptotic distributions of certain test criteria of normality*, South Afr. Statist. J. 6 (1972), pp. 135–49.

[8] I. V. Dunin-Barkovsky and N. V. Smirnov, *Teoriya veroyatnostey i matematicheskaya statistika v tehnike*, Gostechizdat, Moscow 1955.

[9] J. Durbin, *Some methods of constructing exact tests*, Biometrika 48 (1961), pp. 41–55.

[10] A. R. Dyer, *Comparison of tests of normality with a cautionary note*, ibid. 61 (1974), pp. 185–189.

[11] R. A. Fisher, *The conditions under which χ^2 measures the discrepancy between observation and hypothesis*, J. Roy. Statist. Soc. 87 (1924), p. 442.

[12] A. M. Kagan, Yu. V. Linnik, and C. R. Rao, *Characterization problems in mathematical statistics*, Wiley, New York 1973.

[13] H. W. Lilliefors, *On the Kolmogorov–Smirnov test for normality with mean and variance unknown*, J. Amer. Statist. Assoc. 62 (1967), pp. 393–402.

[14] P. Major and G. Tusnady, *Testing for normality*, Magyar Tud. Akad. III. Oszt. Közleményei 22 (1974), pp. 257–281.

[15] D. S. Moore, *A chi-square statistic with random cell boundaries*, Ann. Math. Statist. 42 (1971), pp. 147–156.

[16] M. S. Nikulin, *Chi-square test for continuous distributions with location and scale parameters*, Teor. Verojatnost. 18 (1973), pp. 583–592.

[17] F. J. O'Reilly and C. P. Quesenberry, *The conditional probability integral transformation and applications to obtain composite chi-square goodness-of-fit tests*, Ann. Statist. 1 (1973), pp. 74–83.

[18] E. S. Pearson and H. C. Hartley, *Biometrika tables for statisticians, Vol. II*, Cambridge University Press, Cambridge 1972.

[19] A. A. Petrov, *Proverka hipotezi o normalnosti raspredelenii po malim viborkam*, Dokl. Akad. Nauk 76 (1951), pp. 355–358.

[20] L. Sallay and K. Sarkadi, *Test of normality based on a number of small samples, Proceedings of the 20th Conference of European Organization of Quality Control*, Copenhagen 1976, pp. 21–27.

[21] K. Sarkadi, *On testing for normality*, Publ. Math. Inst. Hung. Acad. Sci. 5 (1960), pp. 269–275.

[22] —, *On testing for normality, Proc. 5th Berkeley Symp. on Math. Statist. and Prob. I*, 1967, pp. 373–387.

[23] —, *The consistency of the Shapiro–Francia test*, Biometrika 62 (1975), pp. 445–450.

[24] —, *On some goodness-of-fit tests*, to be published.

[25] K. Sarkadi and G. Tusnady, *On testing for normality and for the exponential distribution, Proceedings of the 5-th Brasov Conference on Probability Theory*, 1977, pp. 99–118.

[26] S. S. Shapiro and R. S. Francia, *An approximate analysis of variance test for normality*, J. Amer. Statist. Assoc. 67 (1972), pp. 215–216.

[27] S. S. Shapiro and M. B. Wilk, *An analysis of variance test for normality*, Biometrika 42 (1965), pp. 591–611.

[28] —, —, *An analysis of variance test for the exponential distribution*, Technometrics 14 (1972), pp. 355–370.

[29] S. S. Shapiro, M. B. Wilk, and H. J. Chen, *A comparative study of various tests for normality*. J. Amer. Statist. Assoc. 63 (1968), pp. 1343–1372.

[30] M. A. Stephens, *EDF statistics for goodness of fit and some comparisons*, J. Amer. Statist. Assoc. 69 (1974), pp. 730–737.

[31] H. Störmer, *Ein Test zum Erkennen von Normalverteilungen*, Z. Wahrscheinlichkeitstheorie und verw. Gebiete, 2 (1964), pp. 420–428.

[32] M. B. Wilk and S. S. Shapiro, *The joint assessment of normality of several independent samples*, Technometrics 10 (1968), pp. 825–839.

Presented to the semester
MATHEMATICAL STATISTICS
September 15–December 18, 1976