Ann. Inst. Statist. Math.
Vol. 54, No. 3, 531–542 (2002)

GOODNESS-OF-FIT TESTS BASED ON ESTIMATED EXPECTATIONS OF PROBABILITY INTEGRAL TRANSFORMED ORDER STATISTICS*

JAN W. H. SWANEPOEL AND FRANCOIS C. VAN GRAAN

Department of Statistics, Potchefstroom University for CHE, Potchefstroom 2520, South Africa

(Received February 5, 1999; revised February 13, 2001)

Abstract. New goodness-of-fit tests, based on bootstrap estimated expectations of probability integral transformed order statistics, are derived for the location-scale model. The resulting test statistics are location and scale invariant, and are sensitive to discrepancies at the tails of the hypothesized distribution. The limiting null distributions of the test statistics are derived in terms of functionals of a certain Gaussian process, and the tests are shown to be consistent against a broad family of alternatives. Critical points for all sample sizes are provided for tests of normality. A simulation study shows that the proposed tests are more powerful than established tests such as Shapiro-Wilk, Cramér-von Mises and Anderson-Darling, for a wide range of alternative distributions.

Key words and phrases: Bootstrap, consistency, Gaussian process, Monte Carlo simulation, tests for normality, Shapiro-Wilk test.

1. Introduction

Goodness-of-fit has for a long time been a popular research topic and has received considerable attention in the statistical literature. Goodness-of-fit techniques can be described as methods of examining how well a sample of data agrees with a given distribution as its population.

In the formal framework of hypothesis testing, the problem to be considered is the following. Let $X_1, X_2, \dots, X_n$ be independent observations from a population with unknown cumulative distribution function (c.d.f.) $F(x) = P(X \le x)$.

The problem is then to test the composite hypothesis

$$H_0 : F(x) = F_0((x - \mu)/\sigma), \tag{1.1}$$

where F_0 is a known continuous c.d.f. with μ and σ unknown location and scale parameters respectively, against the general alternative $H_1 : F(x) \neq F_0((x-\mu)/\sigma)$. Usually, the goodness-of-fit techniques applied to test H_0 are based on measuring in some way the conformity of the sample data to the hypothesized distribution, or equivalently, its discrepancy from it. The techniques usually give formal statistical tests and the data-based measures of conformity or discrepancy are referred to as test statistics.

In this paper a class of omnibus tests, based on bootstrap estimated expectations of probability integral transformed order statistics, is derived. The resulting test statistics are location and scale invariant, and are sensitive to discrepancies at the tails of the

*This research was supported by grants from the National Research Foundation of South Africa.

hypothesized distribution. The paper is organized as follows. In Section 2 the tests are proposed, and in Section 3 the limiting null distributions of the test statistics are derived in terms of functionals of a certain Gaussian process. Furthermore, the tests are shown to be consistent against a broad family of alternatives. In Section 4 critical points are provided for all sample sizes and three significance levels in the case of tests for normality , that is if $F_0(x) = \Phi(x)$ the c.d.f. of the standard normal distribution. Also, the power performances of the tests are assessed as tests for normality by means of a Monte Carlo study in Section 5. The simulation study shows that the proposed tests are more powerful than established tests such as Shapiro-Wilk, Cramér-von Mises and Anderson-Darling, for a wide range of alternative distributions. Concluding remarks are presented in Section 6 and the proofs of the theorems are given in Section 7.

2. New test statistics

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (i.i.d.) random variables (rv's) with unknown continuous c.d.f. F. Testing of (1.1) will be based on a characterisation of F, which is given by the following theorem.

THEOREM 2.1. *Let $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ be the order statistics of $X_1, X_2, \ldots, X_n$ and G a continuous c.d.f. Then $F = G$ if and only if for every $n \geq 1$ and every k, $1 \leq k \leq n$,*

$$E(G(X_{(k)})) = k/(n+1). \tag{2.1}$$

PROOF. The proof can easily be derived from Lemma 2 of Pollak (1973) and will therefore be omitted. □

The motivation behind the new test statistics is now to propose an estimator for the left-hand side of (2.1) (with $G(x) = F_0((x-\mu)/\sigma)$), and to define appropriate distance measures between this estimator and $D_{n,k} = k/(n+1)$. Firstly, estimate $E\{F_0((X_{(k)} - \mu)/\sigma)\}$ with a bootstrap estimator (see Efron (1979)). Let $X_1^*, X_2^*, \ldots, X_n^*$ be a bootstrap sample of independent rv's with c.d.f. $F_n(x) = n^{-1}\sum_{i=1}^n I(X_i \leq x)$, where $I(\cdot)$ denotes the indicator function. Also, let $X_{(1)}^* \leq X_{(2)}^* \leq \cdots \leq X_{(n)}^*$ be the order statistics of X_i^*, $i = 1, 2, \ldots, n$. It now follows from Efron (1979) that for $i = 1, 2, \ldots, n$, and $k = 1, 2, \ldots, n$,

$$P^*(X_{(k)}^* = X_{(i)}) = \sum_{l=0}^{k-1} \binom{n}{l} ((i-1)/n)^l (1 - ((i-1)/n))^{n-l} - \sum_{l=0}^{k-1} \binom{n}{l} (i/n)^l (1 - i/n)^{n-l},$$

where P^* denotes the conditional probability distribution of $(X_1^*, \ldots, X_n^*)$ given $\boldsymbol{X}_n = (X_1, X_2, \ldots, X_n)$. In order to take into account the possibility of ties in the X_i's, we suggest modifying the above probability to

$$P^*(X_{(k)}^* = \tilde{X}_{(i)}) = \sum_{l=0}^{k-1} \binom{n}{l} (S_{i-1}/n)^l (1 - S_{i-1}/n)^{n-l} \tag{2.2}$$

$$-\sum_{l=0}^{k-1}\binom{n}{l}(S_i/n)^l(1-S_i/n)^{n-l}$$
$$=: w_{n,i}(k),$$

with $S_i = \sum_{j=1}^{n} I(X_j \leq \tilde{X}_{(i)})$, $i = 1, \ldots, m(n)$, where $m(n)$ is the number of different $X_{(i)}$'s with corresponding order statistics denoted by $\tilde{X}_{(1)} < \cdots < \tilde{X}_{(m(n))}$. Hence, a bootstrap estimator of $E\{F_0((X_{(k)} - \mu)/\sigma)\}$ is

$$T_{n,k}(\boldsymbol{X}_n) = E_*\{F_0((X^*_{(k)} - \hat{\mu}_n)/\hat{\sigma}_n)\} = \sum_{i=1}^{m(n)} F_0((\tilde{X}_{(i)} - \hat{\mu}_n)/\hat{\sigma}_n)w_{n,i}(k), \tag{2.3}$$

where $\hat{\mu}_n$ and $\hat{\sigma}_n$ are appropriate estimators of μ and σ respectively, and E_* denotes expectation with respect to P^*.

The test statstics will now be based on appropriate distance measures between $T_{n,k}(\boldsymbol{X}_n)$ and $D_{n,k}$. We propose test statistics of the following types:

(a) Weighted Cramér-von Mises:

$$W_n(g) = \sum_{k=1}^{n} g^2(D_{n,k})\{T_{n,k}(\boldsymbol{X}_n) - D_{n,k}\}^2. \tag{2.4}$$

(b) Weighted Kolmogorov-Smirnov:

$$K_n(g) = \max_{1\leq k\leq n} \sqrt{n}|g(D_{n,k})\{T_{n,k}(\boldsymbol{X}_n) - D_{n,k}\}|, \tag{2.5}$$

where $g(\cdot)$ is some weight function such that $\inf_{0\leq t\leq 1} g(t) \geq \tau$ for some constant $\tau > 0$.

The incorporation of a weight function $g(t)$, $0 \leq t \leq 1$, allows more flexibility in the tests. The function g is to be chosen by the statistician so as to weigh the deviations $|T_{n,k}(\boldsymbol{X}_n) - D_{n,k}|$ according to the importance attached to various subsets of $\{k : k = 1, \ldots, n\}$. Note that $W_n(g)$ and $K_n(g)$ are location and scale invariant if $\hat{\mu}_n$ is equivariant under location and scale and $\hat{\sigma}_n$ is invariant under location but equivariant under scale. For a prescribed significance level α the null hypothesis (1.1) will be rejected if $W_n(g) \geq C_{n,1}(\alpha)(K_n(g) \geq C_{n,2}(\alpha))$, where

$$P_{H_0}(W_n(g) \geq C_{n,1}(\alpha)) = P_{H_0}(K_n(g) \geq C_{n,2}(\alpha)) = \alpha.$$

3. Limiting null distributions and consistency

Let $F_\theta(x) := F_0((x - \mu)/\sigma)$, where $\theta \equiv (\theta_1, \theta_2) \equiv (\mu, \sigma^2)$. In order to derive the asymptotic null distributions of the new test statistics, and to prove consistency, we require the following conditions:

(A) F_0 has an absolutely continuous density f_0 that is uniformly bounded on $(-\infty, \infty)$.

(B) For any θ' in a neighborhood of θ, the following first-order Taylor-series approximation holds

$$\sup_{-\infty<x<\infty} |(F_{\theta'}(x) - F_\theta(x)) - (\theta_1' - \theta_1)G_1(x) - (\theta_2' - \theta_2)G_2(x)|$$
$$= O((\theta_1' - \theta_1)^2 + (\theta_2' - \theta_2)^2),$$

with $G_1(x) := \frac{\partial}{\partial\theta_1}F_\theta(x)$ and $G_2(x) := \frac{\partial}{\partial\theta_2}F_\theta(x)$ being uniformly bounded in $x \in (-\infty, \infty)$.

(C) $G_1(\cdot)$ and $G_2(\cdot)$ are uniformly continuous on $(-\infty, \infty)$.

(D) Any natural estimator $\hat\theta_n \equiv (\hat\mu_n, \hat\sigma_n^2)$ of $\theta \equiv (\mu, \sigma^2)$ satisfies

$$\sqrt{n}(\hat\mu_n - \mu) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_1(\xi_{ni}) + o_P(1),$$
$$\sqrt{n}(\hat\sigma_n^2 - \sigma^2) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_2(\xi_{ni}) + o_P(1),$$

for some real-valued functions ψ_1 and ψ_2, where $\xi_{n1}, \ldots, \xi_{nn}$ is some triangular array of row-independent random variables with $E_\theta(\psi_1(\xi_{ni})) = E_\theta(\psi_2(\xi_{ni})) = 0$, $E_\theta(\psi_1^2(\xi_{ni})) < \infty$, $E_\theta(\psi_2^2(\xi_{ni})) < \infty$, $i = 1, \ldots, n$.

Before we can formulate our theorems, we need some additional notation. Thus, define the inverse of a c.d.f. in general by $H^{-1}(t) := \inf\{x : H(x) \geq t\}$, and set

$$(3.1) \qquad h_1(t) := \left(-\frac{1}{\sigma}\right) f_0(F_0^{-1}(t)), \qquad h_2(t) := \left(-\frac{1}{2\sigma^2}\right) F_0^{-1}(t) f_0(F_0^{-1}(t)).$$

Let $\Lambda = \|\lambda_{ij}\|$ denote the Fisher information matrix, i.e.,

$$(3.2) \ \lambda_{ij} := E\left(\frac{\partial}{\partial\theta_i}\log\frac{1}{\sqrt{\theta_2}}f_0\left(\frac{X_1-\theta_1}{\sqrt{\theta_2}}\right)\cdot\frac{\partial}{\partial\theta_j}\log\frac{1}{\sqrt{\theta_2}}f_0\left(\frac{X_1-\theta_1}{\sqrt{\theta_2}}\right)\right) \qquad i, j = 1, 2,$$

where $\theta \equiv (\theta_1, \theta_2) \equiv (\mu, \sigma^2)$. Further, denote the elements of the inverse of Λ by

$$(3.3) \qquad \Lambda^{-1} = \|\sigma_{ij}\|, \qquad \text{for} \quad i, j = 1, 2,$$

and define

$$(3.4) \qquad \phi_1(t) := (\sigma_{11} - \sigma_{12}^2/\sigma_{22})^{1/2} h_1(t),$$
$$(3.5) \qquad \phi_2(t) := (\sigma_{12}/\sqrt{\sigma_{22}}) h_1(t) + \sqrt{\sigma_{22}} h_2(t).$$

In the following theorem the limiting distributions of $W_n(g)$ and $K_n(g)$ under H_0 are derived and the proofs are deferred to Section 7.

THEOREM 3.1. *If assumptions* (A)–(D) *hold, and if the weight function g is continuous on* $[0, 1]$, *then under* H_0

(a) $W_n(g) \to^d \int_0^1 g^2(t) B^2(t) dt$ *as* $n \to \infty$,

(b) $K_n(g) \to^d \sup_{0 \leq t \leq 1} |g(t)B(t)|$ *as* $n \to \infty$,

where $\{B(t), 0 \leq t \leq 1\}$ *is a Gaussian process with* $B(0) = B(1) = 0$, $E\{B(t)\} = 0$ *and covariance function given by*

$$\mathrm{Cov}\{B(s), B(t)\} = \min(s, t) - st - a(s, t),$$

where (*see equations* (3.1)–(3.5))

$$(3.6) \qquad a(s, t) := \phi_1(s)\phi_1(t) + \phi_2(s)\phi_2(t).$$

The next theorem shows that the goodness-of-fit tests based on $W_n(g)$ and $K_n(g)$ are consistent against a broad family of alternatives.

THEOREM 3.2. *Suppose F is continuous and monotone increasing for all $x \in (-\infty, \infty)$, and that the conditions of Theorem 3.1 hold. Then for any significance level $\alpha \in (0,1)$, the tests based on $W_n(g)$ and $K_n(g)$ are consistent.*

Remarks. (i) If $\hat{\theta}_n$ is efficient (see, e.g., Durbin (1973), p. 287) , then typically condition (D) is satisfied with (ψ_1, ψ_2) being the usual product of score vector times the inverse of Fisher's information matrix.

(ii) Suppose we choose $F_0(x) = \Phi(x)$, the standard normal c.d.f., in Theorem 3.1. Then, in this special case, $\sigma_{11} = \sigma^2$, $\sigma_{22} = 2\sigma^4$, $\sigma_{12} = \sigma_{21} = 0$, so that by (3.1)–(3.6),

$$a(s,t) = (2\pi)^{-1}\{1 + J(s)J(t)/2\} \exp\{-(J^2(s) + J^2(t))/2\},$$

with $J(y) = \Phi^{-1}(y)$.

4. Critical points for testing normality

In this section we consider the case where $F_0(x) = \Phi(x)$ in (1.1), and in (2.3) we choose $\hat{\mu}_n = \bar{X}_n = n^{-1}\sum_{i=1}^n X_i$, and $\hat{\sigma}_n^2 = s_n^2 = (n-1)^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$. An extensive Monte Carlo study was performed to calculate critical points empirically for the test statistics $W_n(g)$ and $K_n(g)$ for sample sizes $5 \le n \le 1000$ and significance levels $\alpha = 0.01$, 0.05 and 0.10. The calculations were based on 100000 independent trials, and we chose the weight function in (2.4) and (2.5) as $g(t) = ((t+c)(1-t+c))^{-1/2}$, where c is an arbitrary small positive constant, e.g., $c = 10^{-6}$. The introduction of such a c is to ensure that g is continuous on [0,1] and hence bounded on [0,1], so that the assumption placed on g in Theorems 3.1 and 3.2 is satisfied. This choice of $g(t)$ makes the tests based on $W_n(g)$ and $K_n(g)$ sensitive to discrepancies at the tails of the hypothesized distribution.

Table 1(a) displays Monte Carlo estimates $\hat{C}_{n,1}(\alpha)$ and $\hat{C}_{n,2}(\alpha)$ of $C_{n,1}(\alpha)$ and $C_{n,2}(\alpha)$ respectively, for $5 \le n \le 20$. For each α, smooth curves of the form $\beta_1/(1 + \beta_2 n^{-1/2} + \beta_3 n^{-1})$ were fitted to both $\hat{C}_{n,1}(\alpha)$ and $\hat{C}_{n,2}(\alpha)$ for $n \ge 21$. This range of values of n was chosen in order to obtain very accurate fits. In fact, in all cases the residuals in absolute value were found to be at most 0.001. Hence, to test H_0 when $n \ge 21$, we suggest rejecting H_0 if the appropriate test statistic exceeds $\hat{\beta}_1/(1 + \hat{\beta}_2 n^{-1/2} + \hat{\beta}_3 n^{-1})$. The values of $\hat{\beta}_1, \hat{\beta}_2$, and $\hat{\beta}_3$ corresponding to $W_n(g)$ and $K_n(g)$ are listed in Table 1(b).

It is worth noting the close correspondence between the estimated asymptotic critical points $\hat{\beta}_1$ of $W_n(g)$ in Table 1(b) and that of the Anderson- Darling test given in Table 4.7 of D'Agostino and Stephens ((1986), p. 123), namely 0.631, 0.752 and 1.035 for $\alpha = 0.10, 0.05$ and 0.01 respectively, which comes as no surprise since the Anderson-Darling test statistic also has (for the choice of g above) the limiting null distribution (see, e.g., Stephens (1976)) given in Theorem 3.1 (a). Although the Anderson-Darling test statistic and $W_n(g)$ have the same asymptotic null distribution, their small and moderate sample behaviour differ as can be seen from the critical points given in Tables 1(a)–1(b) and Table 4.7 of D'Agostino and Stephens (1986). This aspect will be discussed further in Section 5. The calculation of the asymptotic critical points for $K_n(g)$ from its limiting null distribution given in Theorem 3.1 is analytically very difficult, and therefore these points were approximated by Monte Carlo simulations for a sample size $n = 10000$ based

Table 1(a). Estimated critical points based on 100000 trials.

	Significance level α					
	$\alpha = 10\%$		$\alpha = 5\%$		$\alpha = 1\%$	
n	$W_n(g)$	$K_n(g)$	$W_n(g)$	$K_n(g)$	$W_n(g)$	$K_n(g)$
5	0.110	0.447	0.156	0.530	0.251	0.676
6	0.133	0.534	0.187	0.633	0.323	0.826
7	0.157	0.615	0.217	0.730	0.377	0.959
8	0.174	0.683	0.242	0.808	0.421	1.070
9	0.189	0.741	0.263	0.876	0.450	1.161
10	0.206	0.801	0.282	0.944	0.478	1.247
11	0.219	0.851	0.298	1.004	0.508	1.330
12	0.232	0.898	0.313	1.056	0.524	1.398
13	0.242	0.937	0.325	1.105	0.542	1.461
14	0.251	0.977	0.336	1.148	0.557	1.525
15	0.262	1.011	0.350	1.191	0.568	1.576
16	0.269	1.044	0.357	1.227	0.588	1.629
17	0.278	1.077	0.370	1.266	0.590	1.683
18	0.285	1.101	0.375	1.303	0.603	1.725
19	0.292	1.129	0.381	1.331	0.620	1.771
20	0.301	1.158	0.394	1.363	0.632	1.820

Table 1(b). Fitted coefficients $\hat{\beta}_i$, $i = 1, 2, 3$.

	$W_n(g)$			$K_n(g)$		
α	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
0.10	0.6164	1.8521	12.5808	2.9591	4.4581	10.9645
0.05	0.7360	1.5024	10.6601	3.6375	4.6091	12.5754
0.01	1.0390	1.4042	6.6071	5.1877	4.6412	16.5924

on 500000 independent trials. These estimated asymptotic critical points were found to differ from the values of $\hat{\beta}_1$ given in Table 1(b) by at most 0.015 for each α.

5. Power comparisons

In this section we present the results of a Monte Carlo study that compares the power performance of the proposed tests for normality with some well known competitors. The following alternative density functions, which represent a variety of different distributional shapes, were considered: contaminated normal ($CN(\varepsilon, \sigma)$), i.e. $(1-\varepsilon)\phi(x) + (\varepsilon/\sigma)\phi(x/\sigma)$ with $\phi(\cdot)$ the standard normal density function, Student's t with ν degrees of freedom ($t(\nu)$), Tukey with scale 1 and shape $\lambda(T(\lambda))$, standard double exponential (DE) , standard logistic (L), Weibull with scale 1 and shape $\beta(W(\beta))$, chi-square with k degrees of freedom ($\chi^2(k)$), gamma with scale 1 and shape $\gamma(G(\gamma))$, lognormal ($LN(\mu, \sigma)$) , standard extreme value (EV), F with m and n degrees of freedom ($F(m, n)$), binomial ($B(n, p)$), Poisson ($P(\lambda)$) . The power comparisons were made

Table 2. Estimated power functions based on 3000 trials.

	$n = 20$			$n = 50$		
Distribution	SW	$W_n(g)$	$K_n(g)$	SW	$W_n(g)$	$K_n(g)$
$CN(0.50, 0.20)$	0.45	0.55	0.33	0.75	0.93	0.59
$CN(0.50, 0.30)$	0.28	0.34	0.19	0.45	0.69	0.25
$CN(0.50, 0.40)$	0.17	0.20	0.12	0.23	0.39	0.11
$CN(0.50, 3.00)$	0.27	0.32	0.18	0.37	0.58	0.20
$CN(0.50, 4.00)$	0.38	0.46	0.27	0.60	0.83	0.41
$CN(0.50, 5.00)$	0.48	0.57	0.35	0.76	0.93	0.59
$CN(0.25, 0.10)$	0.20	0.26	0.12	0.28	0.55	0.16
$CN(0.25, 0.15)$	0.17	0.22	0.11	0.21	0.46	0.12
$CN(0.25, 0.20)$	0.14	0.18	0.09	0.17	0.38	0.08
$CN(0.25, 2.50)$	0.27	0.28	0.21	0.37	0.49	0.21
$CN(0.25, 3.00)$	0.38	0.40	0.29	0.60	0.71	0.39
$CN(0.25, 3.50)$	0.48	0.52	0.38	0.75	0.84	0.56
$t(1)$	0.86	0.89	0.83	0.99	1.00	0.98
$t(2)$	0.53	0.56	0.45	0.80	0.86	0.68
$t(3)$	0.35	0.36	0.28	0.52	0.61	0.38
$t(4)$	0.23	0.25	0.19	0.37	0.45	0.23
$T(6.0)$	0.20	0.26	0.12	0.38	0.61	0.16
$T(6.5)$	0.28	0.37	0.20	0.55	0.79	0.31
$T(7.0)$	0.37	0.48	0.27	0.70	0.90	0.48
$T(7.5)$	0.47	0.58	0.35	0.82	0.96	0.64
DE	0.26	0.30	0.19	0.40	0.57	0.21
L	0.12	0.13	0.09	0.14	0.18	0.08
$W(1.00)$	0.82	0.81	0.91	1.00	1.00	1.00
$W(1.25)$	0.61	0.60	0.70	0.98	0.98	1.00
$W(1.50)$	0.39	0.38	0.46	0.87	0.86	0.95
$W(1.75)$	0.24	0.23	0.28	0.64	0.63	0.79
$\chi^2(4)$	0.52	0.51	0.59	0.94	0.93	0.98
$\chi^2(5)$	0.45	0.44	0.50	0.89	0.88	0.94
$\chi^2(6)$	0.37	0.36	0.42	0.80	0.79	0.88
$\chi^2(7)$	0.33	0.32	0.36	0.75	0.74	0.82
$G(1)$	0.82	0.81	0.91	1.00	1.00	1.00
$G(2)$	0.53	0.52	0.60	0.95	0.94	0.99
$G(3)$	0.38	0.37	0.42	0.82	0.81	0.88
$G(4)$	0.30	0.29	0.32	0.68	0.67	0.75

for sample sizes $n = 20$ and $n = 50$ at level $\alpha = 5\%$. The number of Monte Carlo repetitions was 3000.

The competing tests which we considered in the power comparisons were the Cramér-von Mises, Anderson-Darling and Shapiro-Wilk (see, e.g., D'Agostino and Stephens (1986)). Overall, the Shapiro-Wilk test appeared slightly better than the Cramér-von

Table 2. (continued)

	$n = 20$			$n = 50$		
Distribution	SW	$W_n(g)$	$K_n(g)$	SW	$W_n(g)$	$K_n(g)$
$LN(0, 0.30)$	0.24	0.23	0.24	0.53	0.53	0.56
$LN(0, 0.50)$	0.51	0.52	0.55	0.91	0.91	0.94
EV	0.32	0.32	0.34	0.67	0.67	0.67
$F(10, 30)$	0.44	0.42	0.46	0.84	0.83	0.88
$F(30, 10)$	0.68	0.67	0.72	0.98	0.99	0.99
$B(20, 0.10)$	0.57	0.95	0.94	1.00	1.00	1.00
$B(20, 0.50)$	0.12	0.78	0.63	0.26	1.00	1.00
$P(4)$	0.25	0.86	0.78	0.64	1.00	1.00
$P(10)$	0.10	0.64	0.43	0.20	1.00	1.00

Mises and Anderson-Darling tests. This is in agreement with previous power studies done in the literature, see for example Stephens (1974), and D'Agostino and Stephens ((1986), p. 406) who concluded that the Shapiro-Wilk test is probably overall the most powerful existing omnibus test for normality. Consequently, only power comparisons between $W_n(g)$, $K_n(g)$, and SW(Shapiro-Wilk) are presented in Table 2. The standard errors of the estimated probabilities in this table are no greater than $\sqrt{0.25/3000} \cong 0.01$. The SW test was implemented as follows: (a) Calculate $Y = \sum_{i=1}^{r} a_{n+1-i}(X_{(n+1-i)} - X_{(i)})$, where $r = (n-1)/2$ if n is odd, and $r = n/2$ if n is even. (b) Define $SW = Y^2/((n-1)s_n^2)$. (c) If $SW \leq C_n(\alpha)$, reject H_0 at level α . The values of $\{a_{n+1-i}\}$ and $C_n(\alpha)$ can be obtained from Tables 5.4 and 5.5 in D'Agostino and Stephens ((1986), pp. 209–213).

Inspection of Table 2 reveals that $W_n(g)$ has superior power performance for symmetric alternatives (especially for $n = 50$). On the other hand, $K_n(g)$ is the test with the highest power in the case of skewed alternative distributions, whereas $W_n(g)$ and SW have almost identical power performance in these cases. When the alternative distribution is discrete, both $W_n(g)$ and $K_n(g)$ outperform SW in all cases. This can be ascribed, among others, to the fact that $W_n(g)$ and $K_n(g)$ are designed to explicitly take into account the possibility of ties, whereas SW appears to be sensitive to ties in the observations.

Similar results regarding power comparisons were also obtained for other alternative distributions, but are not reported here. In view of its overall power performance and computational simplicity, it is our conviction that $W_n(g)$ can be recommended as a powerful *omnibus* test for normality.

6. Concluding remarks

This paper proposed a class of test statistics based on estimated expectations of probability integral transformed order statistics. Our approach is extremely flexible in that other distance measures than those used in (2.4) and (2.5) can be applied, for example Hellinger and Kullback-Leibler measures, to yield alternative test statistics. Weight functions other than $g(t) = ((t + c)(1 - t + c))^{-1/2}$ can be used, as well as different estimators of μ and σ such as efficient robust estimators if outliers are present in the data.

Finally, the techniques developed in this paper can be extended and applied to goodness-of-fit testing in regression and time series analyses. This will be the content of another paper.

Computations were performed using FORTRAN programs and IMSL(Version 2.1) routines on an IBM RS6000 43P PowerPC. Fortran code that carries out the new tests is available from the authors (by e-mail from *sttfcvg@puknet.puk.ac.za*).

7. Proofs

A formal proof of Theorem 3.1 is tedious and therefore only a sketch of the proof is provided.

PROOF OF THEOREM 3.1. Only the proof of (a) is given ; the proof of (b) follows similarly. Thus, from the definition of $T_{n,k}(\boldsymbol{X}_n)$ given in (2.2) and (2.3) (note that since F_0 is assumed to be continuous, with probability one, $\tilde{X}_{(i)} = X_{(i)}$, $S_i = i$, $m(n) = n$, for $i = 1, \dots, n$) it readily follows (by also applying condition (D)) that

$$\max_{1 \le k \le n} |T_{n,k}(\boldsymbol{X}_n) - F_0((X_{(k)} - \hat{\mu}_n)/\hat{\sigma}_n)| = o_P(n^{-1/2}). \tag{7.1}$$

Hence, with $F_n(\cdot)$ denoting the empirical distribution function, we obtain from (7.1) that as $n \to \infty$

$$\begin{aligned} W_n(g) &= \sum_{k=1}^{n} g^2\left(\frac{nF_n(X_{(k)})}{n+1}\right)\left\{F_0\left(\frac{X_{(k)} - \hat{\mu}_n}{\hat{\sigma}_n}\right) - \frac{nF_n(X_{(k)})}{n+1} + o_P(n^{-1/2})\right\}^2 \\ &= n\int_{-\infty}^{\infty} g^2\left(\frac{nF_n(x)}{n+1}\right)\left\{F_0\left(\frac{x - \hat{\mu}_n}{\hat{\sigma}_n}\right) - \frac{nF_n(x)}{n+1} + o_P(n^{-1/2})\right\}^2 dF_n(x) \\ &= n\int_0^1 g^2\left(\frac{nF_n(\theta_n(t))}{n+1}\right)\{F_n(\theta_n(t)) - t + o_P(n^{-1/2})\}^2 dF_n(\theta_n(t)), \end{aligned} \tag{7.2}$$

where $\theta_n(t)$ is defined by

$$\theta_n(t) := \hat{\mu}_n + \hat{\sigma}_n F_0^{-1}(t). \tag{7.3}$$

Note that the $o_P(n^{-1/2})$ term does not depend on t . Hence, writing $dF_n(\theta_n(t))$ as $dt + d(F_n(\theta_n(t)) - t)$, then using partial integration, the boundedness of g, and the Glivenko-Cantelli theorem which states that, with probability one, $F_n(x) \to F(x)$ uniformly in x as $n \to \infty$, we obtain under H_0 that

$$W_n(g) = n\int_0^1 g^2\left(\frac{nF_n(\theta_n(t))}{n+1}\right)\{F_n(\theta_n(t)) - t + o_P(n^{-1/2})\}^2 dt + o_P(1). \tag{7.4}$$

Let $\{B_n(t), 0 \le t \le 1\}$ be a sequence of random elements defined by

$$B_n(t) := \sqrt{n}\{F_n(\theta_n(t)) - t\}.$$

It is well known (see, e.g., Kac *et al.* (1955), p. 192, Sukhatme (1972), Shorack and Wellner (1986), pp. 228–234) that under conditions (A)–(D), and under H_0, as $n \to \infty$

$$B_n \to B \text{ weakly on } D[0,1], \tag{7.5}$$

where $D[0,1]$ is the space of real-valued functions defined on $[0,1]$ that are right-continuous and have left-hand limits.

Next, let $\varepsilon > 0$ be an arbitrary small constant. Since g is uniformly continuous on $[\varepsilon, 1-\varepsilon]$, it readily follows from condition (D) and (7.3) that under H_0, as $n \to \infty$

$$\sup_{\varepsilon \le t \le 1-\varepsilon} \left| g^2\left(\frac{nF_n(\theta_n(t))}{n+1}\right) - g^2(t)\right| = o_P(1). \tag{7.6}$$

Now, note that (7.4) implies

$$\begin{aligned} W_n(g) &\ge \int_\varepsilon^{1-\varepsilon} g^2\left(\frac{nF_n(\theta_n(t))}{n+1}\right)\{B_n(t) + o_P(1)\}^2 dt + o_P(1) \\ &=: L_n(\varepsilon) + o_P(1). \end{aligned} \tag{7.7}$$

For ease of notation, set

$$A_n := \left\{ \sup_{\varepsilon \le t \le 1-\varepsilon} \left| g^2\left(\frac{nF_n(\theta_n(t))}{n+1}\right) - g^2(t)\right| \ge \varepsilon \right\}.$$

Using the fact that g is bounded (by say a constant c), we derive also an upper bound for $W_n(g)$, viz.

$$\begin{aligned} W_n(g) &\le \int_\varepsilon^{1-\varepsilon} (g^2(t) + \varepsilon)\{B_n(t) + o_P(1)\}^2 dt \\ &\quad + c^2 I(A_n) \int_\varepsilon^{1-\varepsilon} \{B_n(t) + o_P(1)\}^2 dt \\ &\quad + c^2 \int_0^\varepsilon \{B_n(t) + o_P(1)\}^2 dt \\ &\quad + c^2 \int_{1-\varepsilon}^1 \{B_n(t) + o_P(1)\}^2 dt + o_P(1) \\ &=: U_n(\varepsilon) + o_P(1). \end{aligned} \tag{7.8}$$

Hence, from (7.5)–(7.8), the boundedness of g, Slutsky's theorem and the continuous mapping theorem we deduce that

$$L_n(\varepsilon) \xrightarrow{d} \int_\varepsilon^{1-\varepsilon} g^2(t)B^2(t)dt, \tag{7.9}$$

and

$$U_n(\varepsilon) \xrightarrow{d} \int_\varepsilon^{1-\varepsilon} (g^2(t) + \varepsilon)B^2(t)dt + c^2 \int_0^\varepsilon B^2(t)dt + c^2 \int_{1-\varepsilon}^1 B^2(t)dt. \tag{7.10}$$

Finally, from (7.7)–(7.10) and the dominated convergence theorem it easily follows (letting $\varepsilon \to 0$) that as $n \to \infty$, $W_n(g) \to^d \int_0^1 g^2(t)B^2(t)dt$. □

PROOF OF THEOREM 3.2. We only prove the consistency of the test based on $W_n(g)$; the proof of the consistency of $K_n(g)$ follows similarly. From Theorem 3.1(a) we deduce that for each α, $0 < \alpha < 1$, $C_{n,1}(\alpha) \to C_1(\alpha) < \infty$ the $(1-\alpha)$-th quantile of the distribution of $\int_0^1 g^2(t)B^2(t)dt$. Hence, the consistency of $W_n(g)$ will follow if it can be shown that

(7.11) $$P_{H_1}(W_n(g) \geq M) \to 1 \quad \text{as} \quad n \to \infty,$$

for any constant $M > 0$.

Under H_1 we have that $F(x) \neq F_0((x-\mu)/\sigma)$ for at least one $x \in (-\infty, \infty)$. This implies that $\{F(x_0) - F_0((x_0-\mu)/\sigma)\}^2 > 0$ for some finite x_0. Since F and F_0 are continuous functions for all x, there exists a fixed neighborhood of x_0, say $I_0 = [x_0 - c, x_0 + d]$, for some positive constants c and d, such that

(7.12) $$\inf_{x \in I_0} \{F(x) - F_0((x-\mu)/\sigma)\}^2 \geq \delta$$

for some constant $\delta > 0$.

Since F_0 is uniformly continuous on I_0 , it readily follows from condition (D), the Glivenko-Cantelli theorem and Slutsky's theorem that as $n \to \infty$

(7.13) $$\sup_{x \in I_0} |V_n(x) - \{F_0((x-\mu)/\sigma) - F(x)\}^2| \to 0$$

in probability, where

$$V_n(x) := \left\{F_0\left(\frac{x - \hat{\mu}_n}{\hat{\sigma}_n}\right) - \frac{nF_n(x)}{n+1} + o_P(n^{-1/2})\right\}^2,$$

with $o_P(n^{-1/2})$ (independent of x) being the $o(\cdot)$-term appearing in (7.2). Hence, from (7.12) and (7.13) it easily follows that as $n \to \infty$

(7.14) $$P_{H_1}\left(\inf_{x \in I_0} V_n(x) > \delta - \varepsilon\right) \to 1,$$

for any constant ε, $0 < \varepsilon < \delta$.

Now, denote

$$W_n^0(g) := n \int_{I_0} g^2(nF_n(x)/(n+1))V_n(x)dF_n(x),$$

and recall that $\inf_{0 \leq t \leq 1} g(t) \geq \delta > 0$. Finally, by (7.2) and (7.14) we obtain

$$\begin{aligned} &P_{H_1}(W_n(g) \geq M) \\ &\geq P_{H_1}(W_n^0(g) \geq M) \\ &\geq P_{H_1}\left(\delta^2(\delta-\varepsilon)\sum_{i=1}^n I(X_i \in I_0) \geq M, \inf_{x \in I_0} V_n(x) > \delta - \varepsilon\right) \\ &\geq P_{H_1}\left(n\delta^2(\delta-\varepsilon)n^{-1}\sum_{i=1}^n I(X_i \in I_0) \geq M\right) \\ &\quad -P_{H_1}\left(\inf_{x \in I_0} V_n(x) \leq \delta - \varepsilon\right) \\ &\to 1 \quad \text{as } n \to \infty, \end{aligned}$$

since by the strong law of large numbers, with probability one, as $n \to \infty$

$$n^{-1}\sum_{i=1}^n I(X_i \in I_0) \to F(x_0 + d) - F(x_0 - c) > 0.$$

This completes the proof of (7.11) and therefore the proof of the theorem. □

Acknowledgements

The authors are very grateful to the referees for their valuable comments.

References

D'Agostino, R. B. and Stephens, M. A. (1986). *Goodness-of-fit Techniques*, Marcel Dekker, New York.

Durbin, J. (1973). Weak convergence of the sample distribution function when parameters are estimated, *Ann. Statist.*, **1**, 279–290.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife, *Ann. Statist.*, **7**, 1–26.

Kac, M., Kiefer, J. and Wolfowitz, J. (1955). On tests of normality and other tests of goodness of fit based on distance methods, *Ann. Math. Statist.*, **26**, 189–211.

Pollak, M. (1973). On equal distributions, *Ann. Statist.*, **1**, 180–182.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*, Wiley, New York.

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

Stephens, M. A. (1976). Asymptotic results for goodness-of-fit statistics with unknown parameters, *Ann. Statist.*, **4**, 357–369.

Sukhatme, S. (1972). Fredholm determinant of a positive definite kernel of a special type and its application, *Ann. Math. Statist.*, **43**, 1914–1926.