J. R. Statist. Soc. B (2002)
64, *Part* 2, *pp.* 281–294

Powerful goodness-of-fit tests based on the likelihood ratio

Jin Zhang

University of Manitoba, Winnipeg, Canada

[Received April 2000. Final revision December 2001]

Summary. A new approach of parameterization is proposed to construct a general goodness-of-fit test. It can not only generate traditional tests (including the Kolmogorov–Smirnov, Cramér–von Mises and Anderson–Darling tests) but also produce new types of omnibus tests, which are generally much more powerful than the old ones.

Keywords: Anderson–Darling test; Cramér–von Mises test; Goodness of fit; Kolmogorov–Smirnov test; Likelihood ratio; Power comparison

1. Introduction

Let X be a continuous random variable with distribution function $F(x)$, and $X_1, X_2, \ldots, X_n$ be a random sample from X with order statistics $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$. We wish to test the null hypothesis

$$H\colon F(x) = F_0(x), \qquad \text{for all } x \in (-\infty, \infty)$$

against the general alternative

$$\bar{H}\colon F(x) \neq F_0(x), \qquad \text{for some } x \in (-\infty, \infty)$$

where $F_0(x)$ is a hypothesized distribution function to be tested. In this paper, we discuss only the basic situation where $F_0(x)$ is completely known. For other cases, see Section 6. Note that

$$H = \bigcap_{t \in (-\infty,\infty)} H_t$$

and

$$\bar{H} = \bigcup_{t \in (-\infty,\infty)} \bar{H}_t,$$

with $H_t\colon F(t) = F_0(t)$ and $\bar{H}_t\colon F(t) \neq F_0(t)$. Then testing H *versus* $\bar{H}$ is equivalent to testing H_t *versus* $\bar{H}_t$ for every $t \in (-\infty, \infty)$.

To test H_t *versus* $\bar{H}_t$ with t fixed, we have a binary random sample based on the indicator function $X_{it} = I(X_i \leqslant t)$ $(i = 1, 2, \ldots, n)$ satisfying $P(X_{it} = 1) = F(t)$ and $P(X_{it} = 0) = 1 - F(t)$.

Address for correspondence: Jin Zhang, Department of Statistics, University of Manitoba, Winnipeg, Manitoba, R3T 2N2, Canada.
E-mail: jin_zhang@umanitoba.ca

 1369–7412/02/64281

$F(x)$ is an arbitrary unknown distribution function, whereas $F(t)$ with t fixed is just an unknown parameter. Through the introduction of the new binary sample, the nonparametric test for H *versus* $\bar{H}$ is simplified to a family of parametric tests for H_t *versus* $\bar{H}_t$, $t \in (-\infty, \infty)$. The simplification is a process of parameterization, through which parametric approaches can be applied to nonparametric tests. Hopefully, the parameterization loses little information about $F(x)$ in the original sample.

For each fixed $t \in (-\infty, \infty)$ and the corresponding random sample $X_{1t}, X_{2t}, \ldots, X_{nt}$, let Z_t be a statistic for testing H_t *versus* $\bar{H}_t$ such that its large values reject H_t. Then two types of statistic for testing H *versus* $\bar{H}$ can be defined by

$$\begin{aligned} Z &= \int_{-\infty}^{\infty} Z_t \, \mathrm{d}w(t), \\ Z_{\max} &= \sup_{t \in (-\infty,\infty)} \{Z_t \, w(t)\}, \end{aligned} \tag{1.1}$$

where $w(t)$ is some weight function and large values of Z or $Z_{\max}$ reject the null hypothesis H.

The power of Z or $Z_{\max}$ depends on Z_t and $w(t)$. Two natural candidates for Z_t are the Pearson χ^2 test statistic and the likelihood ratio test statistic, which are respectively (after simplification)

$$X_t^2 = \frac{n\{F_n(t) - F_0(t)\}^2}{F_0(t)\{1 - F_0(t)\}} \tag{1.2}$$

and

$$G_t^2 = 2n\left[F_n(t) \, \log\left\{\frac{F_n(t)}{F_0(t)}\right\} + \{1 - F_n(t)\} \, \log\left\{\frac{1 - F_n(t)}{1 - F_0(t)}\right\}\right], \tag{1.3}$$

where $F_n(t)$ is the empirical distribution function of the original sample $X_1, X_2, \ldots, X_n$.

A large family of Z_t which embeds X_t^2 and G_t^2 can be obtained by using the Cressie and Read (1984) family of divergence statistics $2nI^\lambda$ for testing the goodness of fit of a multinomial distribution. For the above binary sample $X_{1t}, X_{2t}, \ldots, X_{nt}$ with t fixed, the Cressie–Read family of divergence statistics for testing H_t *versus* $\bar{H}_t$ is

$$2nI_t^\lambda = \frac{2n}{\lambda(\lambda+1)}\left[F_n(t)\left\{\frac{F_n(t)}{F_0(t)}\right\}^\lambda + \{1 - F_n(t)\}\left\{\frac{1 - F_n(t)}{1 - F_0(t)}\right\}^\lambda - 1\right], \tag{1.4}$$

which includes X_t^2 ($\lambda = 1$) and G_t^2 ($\lambda = 0$), as well as other important statistics (Cressie and Read, 1984; Read and Cressie, 1988).

In this paper, we focus on X_t^2 and G_t^2 because X_t^2 is associated with classical tests whereas G_t^2 produces new powerful goodness-of-fit tests. In Sections 2 and 3, we shall show by choosing appropriate weight functions in equations (1.1) that

(a) using X_t^2 as Z_t generates traditional goodness-of-fit test statistics, including the Kolmogorov–Smirnov, Cramér–von Mises and Anderson–Darling statistics, and
(b) using G_t^2 as Z_t produces new statistics, which are sometimes substantially more powerful than the traditional statistics.

Power comparisons are given in Section 4, and simulated percentage points for the new statistics are tabulated in Section 5. Section 6 gives concluding remarks.

2. Derivation of traditional goodness-of-fit tests

Using X_t^2 as Z_t in equations (1.1) but choosing different weight functions, we can derive traditional goodness-of-fit tests. Below are three examples.

2.1. $w(t) = n^{-1} F_0(t)\{1 - F_0(t)\}$

Replacing Z_t of the second of equations (1.1) with X_t^2 generates

$$K_S^2 = \{\sup_{t\in(-\infty,\infty)} |F_n(t) - F_0(t)|\}^2 = \left(\max_{1\leqslant i\leqslant n}\left[\max\left\{\frac{i}{n} - F_0(X_{(i)}),\ F_0(X_{(i)}) - \frac{i-1}{n}\right\}\right]\right)^2,$$

where K_S is the Kolmogorov–Smirnov statistic, the best-known statistic for goodness-of-fit tests (Kolmogorov, 1933; Smirnov, 1939; Massey, 1951; Stephens, 1970, 1974; Pratt and Gibbons, 1981; D'Agostino and Stephens, 1986; Gibbons, 1992; Cabaña, 1996; Conover, 1999).

2.2. $w(t) = F_0(t)$

Replacing Z_t of the first of equations (1.1) with X_t^2 generates the Anderson–Darling statistic

$$\begin{aligned} A^2 &= n\int_{-\infty}^{\infty} \{F_n(t) - F_0(t)\}^2\, F_0(t)^{-1}\{1 - F_0(t)\}^{-1}\, \mathrm{d}F_0(t) \\ &= -\frac{2}{n}\sum_{i=1}^{n}\left[\left(i - \frac{1}{2}\right)\log\{F_0(X_{(i)})\} + \left(n - i + \frac{1}{2}\right)\log\{1 - F_0(X_{(i)})\}\right] - n, \end{aligned}$$

one of the most powerful and important goodness-of-fit tests in the literature (Anderson and Darling, 1952, 1954; Stephens, 1970, 1974; D'Agostino and Stephens, 1986; Sinclair and Spurr, 1988).

2.3. $dw(t) = F_0(t)\{1 - F_0(t)\}\ dF_0(t)$

Replacing Z_t of the first of equations (1.1) with X_t^2 generates the famous Cramér–von Mises statistic (Cramér, 1928; von Mises, 1931; Smirnov, 1936, 1937; Stephens, 1970, 1974; Knott, 1974; D'Agostino and Stephens, 1986; Csörgő and Faraway, 1996; Spinelli and Stephens, 1997; Conover, 1999):

$$W^2 = n\int_{-\infty}^{\infty} \{F_n(t) - F_0(t)\}^2\, \mathrm{d}F_0(t) = \sum_{i=1}^{n}\left\{F_0(X_{(i)}) - \frac{i - \frac{1}{2}}{n}\right\}^2 + \frac{1}{12n}.$$

3. New powerful goodness-of-fit tests

It is well known that, when testing the goodness of fit for a multinomial distribution, the Pearson χ^2-statistic is asymptotically equivalent to the likelihood ratio statistic. Therefore, under the null hypothesis H_t in Section 1, the χ^2-statistic X_t^2 and the likelihood ratio statistic G_t^2 are equivalent in large sample situations, but they behave differently under the alternative $\bar{H}_t$. We have seen in Section 2 that traditional tests can be generated by using X_t^2 as Z_t in equations (1.1). In this section we shall use G_t^2 to produce new tests by choosing appropriate weight functions.

Let $U_i = F_0(X_i)$ $(i = 1, 2, \ldots, n)$ so that $U_{(i)} = F_0(X_{(i)})$. Note that $X_1, X_2, \ldots, X_n$ are independent and identically distributed (IID) from F_0 if and only if $U_1, U_2, \ldots, U_n$ are IID from $U(0, 1)$, the standard uniform distribution. To test H *versus* $\bar{H}$, consider a statistic with the form

$T = T(U_1, U_2, \ldots, U_n)$, where $T(\cdot)$ is a given function independent of F_0. Since U_i and $1 - U_i$ are identically distributed under hypothesis H, a reasonable T should satisfy

$$T(U_1, U_2, \ldots, U_n) = T(1 - U_1, 1 - U_2, \ldots, 1 - U_n).$$

In such a case, we say that T is *distribution symmetric* about the median.

It is obvious that the traditional statistics (2.1)–(2.3) are functions of $U_1, U_2, \ldots, U_n$, and they are distribution symmetric. To generate new distribution symmetric tests, we must choose appropriate weight functions. Moreover, we sometimes need to modify $F_n(t)$ at its discontinuity points $X_{(i)}$ $(i = 1, 2, \ldots, n)$ by defining $F_n(X_{(i)}) = (i - c)/(n + 1 - 2c)$, where c is a constant between 0 and 1. The natural and intuitive choice of c is $\frac{1}{2}$ so that $F_n(X_{(i)}) = (i - \frac{1}{2})/n$ or $\{F_n(X_{(i)} - 0) + F_n(X_{(i)} + 0)\}/2$, which is a common 'continuity correction' to the empirical distribution function. In fact, we can imagine that at point $x = X_{(i)}$ there are $i - \frac{1}{2}$ or $n - i + \frac{1}{2}$ observations among $X_1, X_2, \ldots, X_n$ which are less or greater than the x. Finally, the traditional test statistics in Section 2 also suggest that $F_n(X_{(i)})$ should be $(i - \frac{1}{2})/n$ instead of i/n.

When necessary, we always define $F_n(X_{(i)}) = (i - \frac{1}{2})/n$. Then new distribution symmetric tests can be generated by choosing weight functions as follows.

3.1. $w(t) = 1$

Let $X_{(0)} = -\infty$ and $X_{(n+1)} = \infty$. Replacing Z_t of the second of equations (1.1) with G_t^2 produces

$$\sup_{t \in (-\infty, \infty)} (G_t^2) = \max_{0 \leqslant i \leqslant n} \left\{ \sup_{X_{(i)} \leqslant t < X_{(i+1)}} (G_t^2) \right\} = \max_{1 \leqslant i \leqslant n} (G_{X_{(i)}}^2),$$

which is equivalent to

$$Z_K = \max_{1 \leqslant i \leqslant n} \left(\left(i - \frac{1}{2}\right) \log\left\{ \frac{i - \frac{1}{2}}{n\, F_0(X_{(i)})} \right\} + \left(n - i + \frac{1}{2}\right) \log\left[\frac{n - i + \frac{1}{2}}{n\{1 - F_0(X_{(i)})\}} \right] \right). \tag{3.1}$$

3.2. $dw(t) = F_n(t)^{-1}\{1 - F_n(t)\}^{-1}\, dF_n(t)$

Replacing Z_t of the first of equations (1.1) with G_t^2 produces

$$2 \sum_{i=1}^{n} \left(\frac{n}{n - i + \frac{1}{2}} \log\left\{ \frac{i - \frac{1}{2}}{n\, F_0(X_{(i)})} \right\} + \frac{n}{i - \frac{1}{2}} \log\left[\frac{n - i + \frac{1}{2}}{n\{1 - F_0(X_{(i)})\}} \right] \right),$$

which is equivalent to

$$Z_A = -\sum_{i=1}^{n} \left[\frac{\log\{F_0(X_{(i)})\}}{n - i + \frac{1}{2}} + \frac{\log\{1 - F_0(X_{(i)})\}}{i - \frac{1}{2}} \right]. \tag{3.2}$$

3.3. $dw(t) = F_0(t)^{-1}\{1 - F_0(t)\}^{-1}\, dF_0(t)$

Replacing Z_t of the first of equations (1.1) with G_t^2 produces

$$\sum_{i=1}^{n} [\log\{F_0(X_{(i)})^{-1} - 1\} - b_{i-1} + b_i]^2 + C_n,$$

where C_n is a constant and $b_i = i \log(i/n) + (n - i) \log(1 - i/n)$.

Since $b_{i-1} - b_i \approx \log\{(n-\frac{1}{2})/(i-\frac{3}{4})-1\}$, the above test statistic is approximately equivalent to

$$Z_C = \sum_{i=1}^{n} \left[\log\left\{ \frac{F_0(X_{(i)})^{-1} - 1}{(n-\frac{1}{2})/(i-\frac{3}{4})-1} \right\} \right]^2 . \tag{3.3}$$

The new statistics Z_K, Z_A and Z_C are distribution symmetric. They appear similar to the Kolmogorov–Smirnov K_S, Anderson–Darling A^2 and Cramér–von Mises W^2 respectively, but they are generally much more powerful (see Section 4). Unfortunately, it seems difficult to use the analogue Z_K of K_S to construct simultaneous confidence bands for the true distribution $F(x)$, which is a useful application of K_S.

4. Power comparison by simulation

In this section we shall use the Monte Carlo approach to examine the powers of the new statistics Z_A, Z_C and Z_K and the traditional Kolmogorov–Smirnov statistic K_S, Cramér–von Mises statistic W^2, Anderson–Darling statistic A^2 and Pearson's χ^2-statistic X^2. For the χ^2-test, the sample observations need to be grouped. Here we use the associated function in S-PLUS with default values (see, for example, MathSoft (2000)).

The simulation size is 10000, and the significance level or the probability of type I error for testing the goodness of fit is $\alpha = 0.05$, at which level the critical values of the tests are simulated independently with 1 million replicates except for the χ^2-test built into S-PLUS involving asymptotic approximation. The actual size of the χ^2-test is close to 0.05 even for small samples according to our simulation. For various null hypotheses H and the alternatives $\bar{H}$, all simulated powers for the seven statistics are illustrated with graphs, where the powers are plotted against the sample size n for selected values $n = 10, 20, 30, 50, 70, 100, 150, 200, 300$.

4.1. Example 1: H: $X_1,\ldots, X_n \sim^{\mathrm{IID}} U(0, 1)$ versus $\bar{H}$: $X_1,\ldots, X_n \sim^{\mathrm{IID}}$ beta(p, q)

Without loss of generality, we can assume that the underlying distribution F is the standard uniform $U(0, 1)$ distribution under the null hypothesis H. Then a natural candidate for F under the alternative $\bar{H}$ is the beta distribution beta(p, q) with parameters p and q, which includes the uniform $U(0, 1)$ distribution, or beta$(1, 1)$ distribution. So, this example is actually a parametric test for H: $(p, q) = (1, 1)$ *versus* $\bar{H}$: $(p, q) \neq (1, 1)$.

For $(p, q) = (0.6, 0.8), (0.6, 0.6), (0.8, 0.8), (1.3, 1.3), (1.6, 1.6), (1.3, 1.6)$, the powers of Z_A, Z_C, Z_K, K_S, W^2, A^2 and X^2 under the alternative hypothesis $\bar{H}$ are plotted in Fig. 1. We see that Z_A and Z_C have the highest power in the cases where $p, q > 1$ or $p, q < 1$, and they dominate all the others. Although Z_K is not as powerful as Z_A and Z_C, it is still more powerful than its analogue K_S.

We also consider the power of the entropy-based test of uniformity proposed by Dudewicz and van der Meulen (1981). Their method involves choosing the best integer m which depends on the sample size n, but the power results in their Table 3 are obtained from choosing the best m not only for different n but also for different alternative distributions. Of course, different tests fit different models. However, when performing a nonparametric test, we have no idea about the alternative distribution. Therefore, if a fixed m is used for the same n but different alternative distributions, the power of such a test is generally lower than that of the Anderson–Darling test A^2.

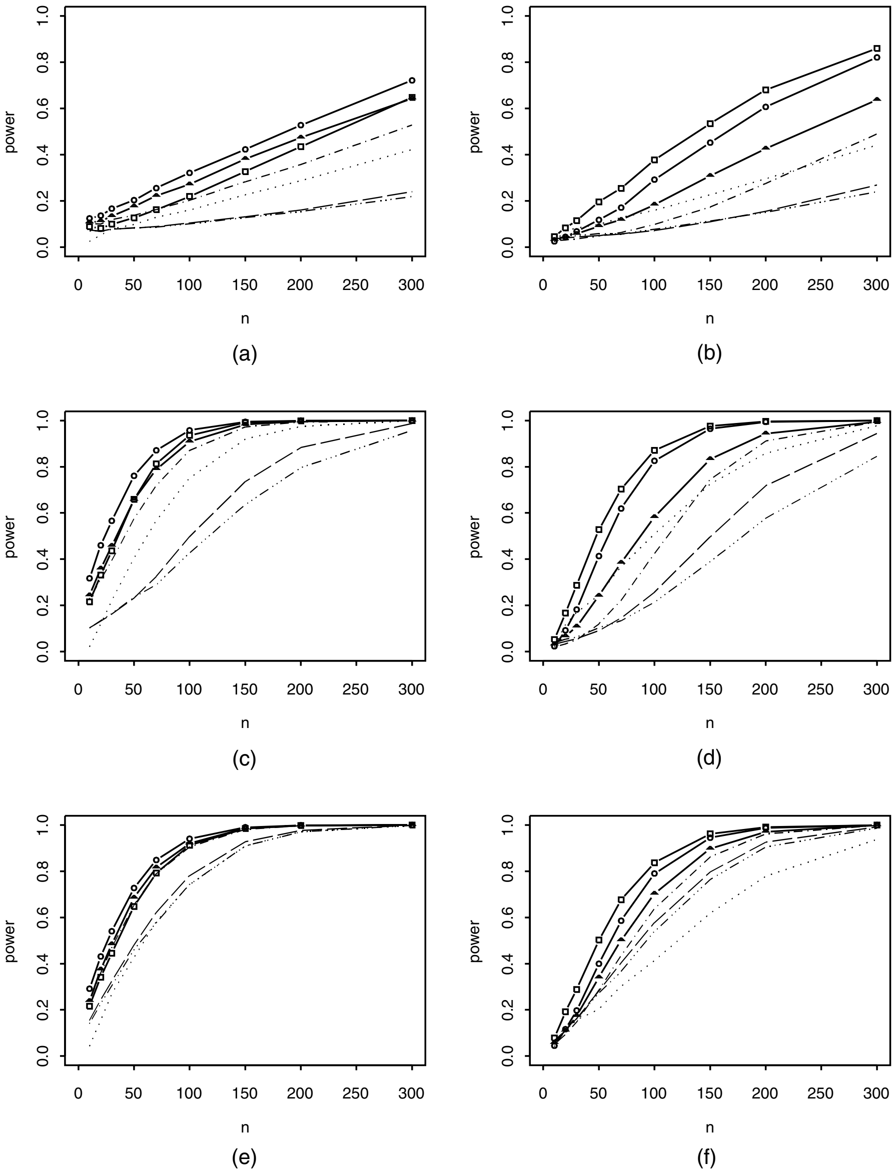

Fig. 1. Comparison of powers when testing $H: F = U(0, 1)$ *versus* $\bar{H}: F = \text{beta}(p, q)$ at level $\alpha = 0.05$ (□, Z_A; ○, Z_C; ▲, Z_K; ⋯⋯, X^2; · — · — ·, A^2; ———, W^2; ⋯—⋯, K_S): (a) uniform *versus* beta(0.8, 0.8); (b) uniform *versus* beta(1.3, 1.3); (c) uniform *versus* beta(0.6, 0.6); (d) uniform *versus* beta(1.6, 1.6); (e) uniform *versus* beta(0.6, 0.8); (f) uniform *versus* beta(1.3, 1.6)

4.2. *Example 2:* $H: X_1, \ldots, X_n \sim^{\text{IID}} N(\mu, \sigma^2)$ *versus* $\bar{H}: X_1, \ldots, X_n \sim^{\text{IID}} t(k)$

Because of the importance of the normal distribution, F is assumed to be a normal distribution $N(\mu, \sigma^2)$ under the null hypothesis H. It is interesting to consider that

(a) F also has a symmetric distribution under the alternative $\bar{H}$, say $t(k)$, the t-distribution with k degrees of freedom, and

(b) both distributions have the same mean and variance, i.e. $\mu = 0$ and $\sigma^2 = k/(k-2)$ $(k \geqslant 3)$.

Since $N(0, 1) = t(\infty)$, testing H: $F = N(\mu, \sigma^2)$ *versus* $\bar{H}$: $F = t(k)$ is equivalent to testing H: $k = \infty$ *versus* $\bar{H}$: $k \neq \infty$. Fig. 2 compares the powers of the seven statistics Z_A, Z_C, Z_K, K_S,

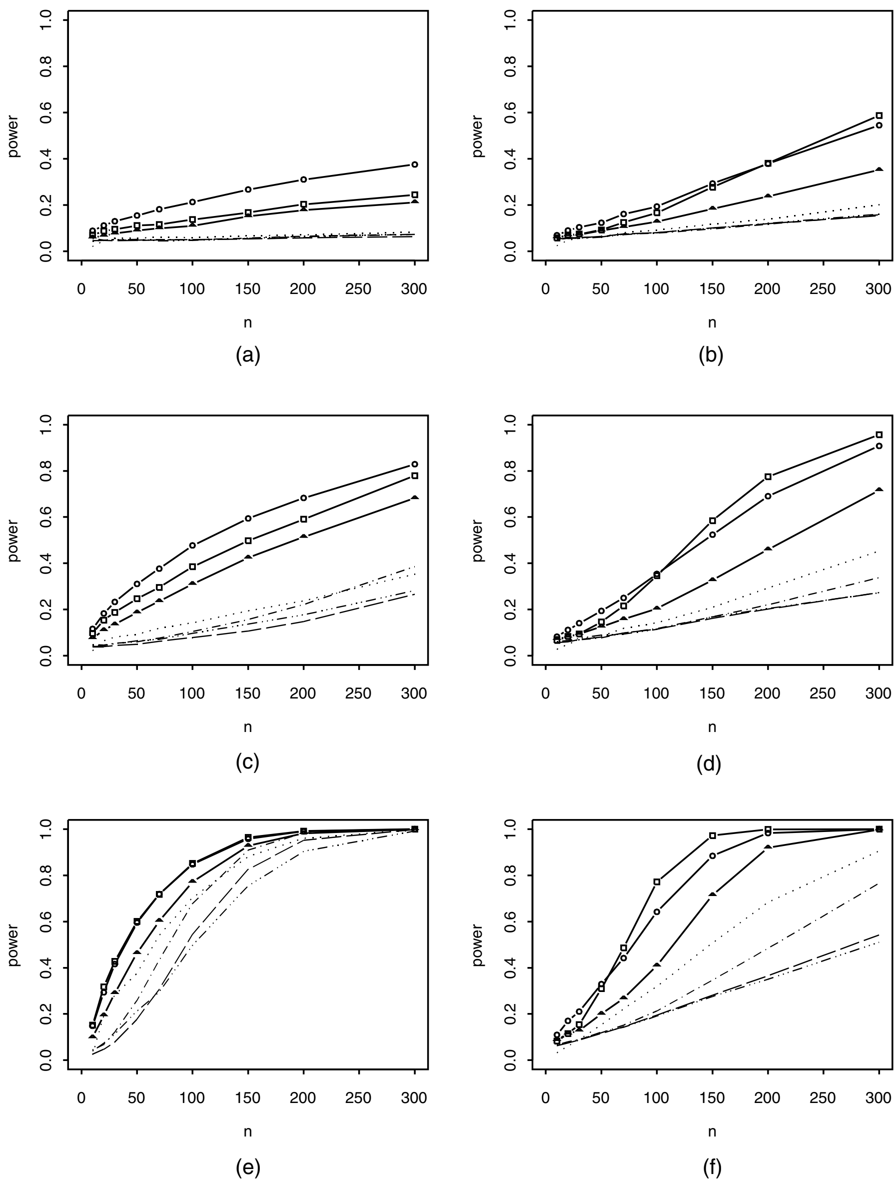

Fig. 2. Comparison of powers when testing either H: $F = N(\mu, \sigma^2)$ *versus* $\bar{H}$: $F = t(k)$ or H: $F = N(\mu, \sigma^2)$ *versus* $\bar{H}$: F = gamma(r, 1) at level $\alpha = 0.05$ (□, Z_A; ○, Z_C; ▲, Z_K; · · · · · · ·, X^2; · – · – ·, A^2; ———, W^2; · · ·–· · ·, K_S): (a) normal *versus* $t(10)$; (b) normal *versus* gamma(20, 1); (c) normal *versus* $t(5)$; (d) normal *versus* gamma(10, 1); (e) normal *versus* $t(3)$; (f) normal *versus* gamma(5, 1)

W^2, A^2 and X^2 for $k = 3, 5, 10$. Clearly Z_C is the best and Z_A, Z_C and Z_K dominate the others (sometimes they are much more powerful).

The Cauchy and logistic distributions are also typical examples of symmetric distributions which can be considered as the underlying distribution under hypothesis $\bar{H}$. The power comparison for the logistic distribution is like that for $t(9)$, whereas the Cauchy distribution is $t(1)$.

4.3. Example 3: $H: X_1, \ldots, X_n \sim^{\mathrm{IID}} N(\mu, \sigma^2)$ versus $\bar{H}: X_1, \ldots, X_n \sim^{\mathrm{IID}}$ gamma$(r, 1)$

In this example F is also assumed to be $N(\mu, \sigma^2)$ under hypothesis H, but it has an asymmetric distribution under $\bar{H}$, such as gamma$(r, 1)$, the gamma distribution with shape parameter r and scale parameter 1, which includes exponential and χ^2-distributions. We also assume that both distributions have the same mean and variance, i.e. $\mu = r$ and $\sigma^2 = r$.

Similarly, since the asymptotic distribution of gamma$(r, 1)$ is normal when $r \to \infty$, testing H: $F = N(\mu, \sigma^2)$ *versus* $\bar{H}$: $F =$ gamma$(r, 1)$ is equivalent to testing H: $r = \infty$ *versus* $\bar{H}$: $r \neq \infty$.

Simulated powers of the seven statistics for $r = 5, 10, 20$ are also plotted in Fig. 2, which shows that

(a) Z_A, Z_C and Z_K dominate the others and
(b) the powers of Z_A and Z_C are sometimes substantially higher than those of the traditional statistics.

Other asymmetric distributions, such as the log-normal, Weibull, F- and beta distributions, were also considered as alternative distributions against the normal distribution. These situations are similar to that of the gamma distribution.

4.4. Example 4: $H: X_1, \ldots, X_n \sim^{\mathrm{IID}} N(0, 1)$ versus $\bar{H}: X_1, \ldots, X_n \sim^{\mathrm{IID}} N(\mu, \sigma^2)$

In the last example, F is assumed to be normal under both hypotheses H and $\bar{H}$. Without loss of generality, we need to consider only the test for H: $F = N(0, 1)$ *versus* $\bar{H}$: $F = N(\mu, \sigma^2)$, or equivalently H: $(\mu, \sigma^2) = (0, 1)$ *versus* $\bar{H}$: $(\mu, \sigma^2) \neq (0, 1)$.

Six cases are considered with alternatives

(a) $N(0.1, 1)$,
(b) $N(0.4, 1)$,
(c) $N(0, 1.5)$,
(d) $N(0, 2)$,
(e) $N(0.1, 2)$ and
(f) $N(0.4, 1.5)$.

In cases (a) and (b), the two distributions have the same variance but different means, and in cases (c) and (d) they have the same mean but different variances. In cases (e) and (f), the means and variances are both different.

For normal distribution models, the distributions differ in the mean and variance only. There is no shape difference in terms of skewness and kurtosis. For each case Fig. 3 compares the powers of the seven tests, as well as the optimal parametric t-test and the χ^2-test for a normal mean and variance.

It is clear that for cases (a), (b) and (f) where the major difference between the two distributions arises from their means rather than variances, there is no significant difference in power between the new tests and their analogues A^2, W^2 and K_S. Conversely, for the other three cases, the

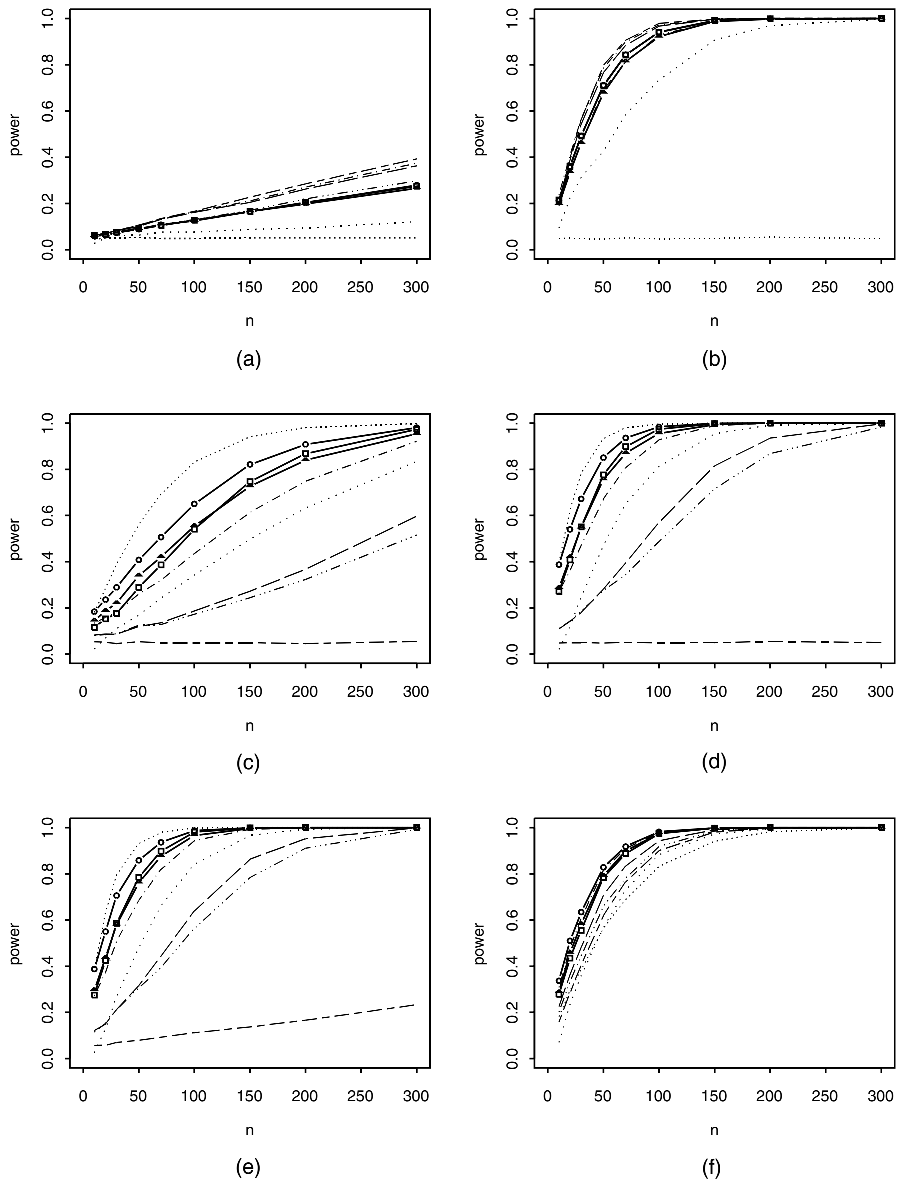

Fig. 3. Comparison of powers when testing H: $F = N(0, 1)$ *versus* $\bar{H}$: $F = N(\mu, \sigma^2)$ at level $\alpha = 0.05$ ($\square$, Z_A; $\circ$, Z_C; $\blacktriangle$, Z_K; ·······, X^2; · – · – ·, A^2; — — —, W^2; ···–···, K_S; - – - –, t; ·········, χ^2): (a) $N(0, 1)$ *versus* $N(0.1, 1)$; (b) $N(0, 1)$ *versus* $N(0.4, 1)$; (c) $N(0, 1)$ *versus* $N(0, 1.5)$; (d) $N(0, 1)$ *versus* $N(0, 2)$; (e) $N(0, 1)$ *versus* $N(0.1, 2)$; (f) $N(0, 1)$ *versus* $N(0.4, 1.5)$

advantage of the new tests is obvious. When the difference in distribution arises from their means only, such as cases (a) and (b), the six tests are almost as powerful as the optimal t-test. In cases (c) and (d) where the only difference comes from the variances, the power lost by using the new tests over the χ^2-test is much less than that by using their analogues. Of A^2, W^2 and K_S, A^2 is the best in almost all cases, which is also true for other examples.

Table 1. Percentage points for $10Z_A - 32$

n	Percentage points for the following levels α:														
	0.001	0.01	0.05	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	0.95	0.99	0.999
5	−0.1639	0.140	0.666	1.094	1.824	2.549	3.334	4.235	5.317	6.71	8.70	12.24	15.98	25.43	39.77
6	0.0718	0.358	0.836	1.224	1.875	2.514	3.200	3.989	4.929	6.14	7.86	10.88	14.04	22.10	34.55
7	0.2386	0.506	0.942	1.292	1.879	2.452	3.064	3.763	4.599	5.67	7.18	9.83	12.62	19.68	30.39
8	0.3532	0.613	1.015	1.335	1.872	2.392	2.946	3.574	4.322	5.28	6.63	9.01	11.49	17.69	27.46
9	0.4454	0.683	1.061	1.359	1.854	2.327	2.834	3.405	4.084	4.95	6.18	8.32	10.54	16.17	24.74
10	0.5140	0.742	1.095	1.374	1.831	2.270	2.736	3.261	3.887	4.69	5.80	7.75	9.79	14.89	22.74
12	0.6158	0.823	1.137	1.381	1.782	2.164	2.566	3.017	3.554	4.23	5.19	6.86	8.60	12.89	19.64
14	0.6861	0.872	1.156	1.375	1.732	2.071	2.427	2.824	3.294	3.89	4.73	6.19	7.70	11.46	17.32
16	0.7350	0.906	1.164	1.364	1.686	1.989	2.309	2.665	3.087	3.62	4.37	5.67	7.01	10.35	15.47
18	0.7659	0.929	1.168	1.351	1.645	1.923	2.214	2.536	2.917	3.40	4.07	5.24	6.46	9.47	14.11
20	0.7964	0.945	1.167	1.336	1.607	1.862	2.127	2.421	2.769	3.21	3.82	4.89	5.99	8.69	12.84
25	0.8427	0.972	1.159	1.301	1.528	1.741	1.961	2.204	2.490	2.85	3.35	4.23	5.13	7.32	10.71
30	0.8668	0.982	1.147	1.271	1.467	1.650	1.838	2.047	2.291	2.60	3.03	3.76	4.53	6.39	9.22
40	0.8980	0.990	1.122	1.220	1.377	1.521	1.668	1.831	2.022	2.26	2.59	3.16	3.75	5.17	7.37
50	0.9120	0.990	1.102	1.184	1.314	1.433	1.555	1.689	1.845	2.04	2.31	2.77	3.25	4.41	6.16
70	0.9233	0.983	1.070	1.132	1.230	1.319	1.410	1.510	1.626	1.77	1.97	2.31	2.65	3.49	4.75
100	0.9270	0.974	1.038	1.085	1.157	1.222	1.289	1.361	1.445	1.55	1.69	1.93	2.18	2.78	3.70
150	0.9272	0.960	1.007	1.040	1.092	1.138	1.184	1.234	1.292	1.36	1.46	1.63	1.80	2.20	2.81
200	0.9246	0.951	0.988	1.014	1.054	1.089	1.125	1.164	1.209	1.26	1.34	1.47	1.59	1.90	2.35
300	0.9207	0.940	0.966	0.985	1.013	1.037	1.062	1.089	1.120	1.16	1.21	1.30	1.38	1.59	1.90
500	0.9156	0.928	0.945	0.957	0.975	0.991	1.006	1.023	1.042	1.07	1.10	1.15	1.20	1.33	1.52
1000	0.9097	0.917	0.926	0.933	0.942	0.951	0.959	0.968	0.978	0.99	1.01	1.03	1.06	1.13	1.22

Table 2. Percentage points for Z_C

n	*Percentage points for the following levels* α:														
	0.001	*0.01*	*0.05*	*0.10*	*0.20*	*0.30*	*0.40*	*0.50*	*0.60*	*0.70*	*0.80*	*0.90*	*0.95*	*0.99*	*0.999*
5	0.319	0.801	1.63	2.26	3.24	4.13	5.04	6.03	7.18	8.61	10.6	14.2	18.3	30.6	54.6
6	0.437	0.977	1.86	2.51	3.54	4.46	5.41	6.44	7.63	9.11	11.2	14.8	18.9	31.1	54.9
7	0.536	1.118	2.04	2.72	3.78	4.74	5.71	6.76	7.98	9.50	11.6	15.3	19.4	31.5	54.4
8	0.621	1.241	2.20	2.91	3.99	4.97	5.97	7.05	8.29	9.84	12.0	15.7	19.9	31.8	54.8
9	0.703	1.352	2.34	3.06	4.18	5.18	6.19	7.29	8.56	10.14	12.3	16.1	20.2	32.1	55.6
10	0.762	1.444	2.46	3.20	4.34	5.36	6.40	7.52	8.81	10.42	12.6	16.5	20.6	32.4	55.1
12	0.892	1.625	2.68	3.45	4.63	5.68	6.75	7.90	9.22	10.85	13.1	17.0	21.2	32.9	56.2
14	1.002	1.763	2.86	3.65	4.87	5.95	7.04	8.21	9.56	11.24	13.5	17.5	21.7	33.4	56.2
16	1.087	1.893	3.02	3.83	5.07	6.18	7.29	8.48	9.86	11.57	13.9	17.9	22.2	33.8	56.3
18	1.168	1.999	3.15	3.99	5.26	6.38	7.51	8.73	10.13	11.85	14.2	18.3	22.6	34.3	56.8
20	1.232	2.087	3.27	4.12	5.41	6.55	7.69	8.93	10.35	12.10	14.5	18.6	22.9	34.5	57.1
25	1.408	2.286	3.53	4.41	5.75	6.93	8.11	9.38	10.82	12.62	15.1	19.3	23.6	35.4	57.6
30	1.520	2.472	3.74	4.65	6.02	7.24	8.44	9.74	11.22	13.05	15.5	19.8	24.2	35.8	57.4
40	1.720	2.714	4.07	5.03	6.46	7.72	8.97	10.32	11.86	13.75	16.3	20.7	25.2	36.9	59.1
50	1.907	2.928	4.33	5.32	6.79	8.09	9.39	10.76	12.34	14.26	16.9	21.3	25.9	37.5	59.4
70	2.130	3.244	4.72	5.76	7.31	8.66	9.99	11.42	13.05	15.03	17.7	22.3	26.9	38.6	60.2
100	2.390	3.583	5.15	6.23	7.84	9.25	10.64	12.12	13.79	15.84	18.6	23.3	28.0	39.8	61.4
150	2.701	3.981	5.63	6.78	8.47	9.95	11.38	12.93	14.67	16.79	19.6	24.4	29.2	41.1	62.2
200	2.939	4.256	5.98	7.16	8.90	10.41	11.90	13.48	15.26	17.43	20.3	25.2	30.1	42.2	63.5
300	3.269	4.662	6.48	7.71	9.54	11.11	12.66	14.29	16.13	18.36	21.4	26.3	31.3	43.5	64.7
500	3.630	5.178	7.11	8.41	10.33	11.97	13.60	15.30	17.21	19.52	22.6	27.7	32.8	45.1	66.1
1000	4.217	5.867	7.96	9.37	11.42	13.17	14.86	16.65	18.66	21.06	24.3	29.6	34.8	47.3	68.5

Table 3. Percentage points for Z_K

n	Percentage points for the following levels α:														
	0.001	*0.01*	*0.05*	*0.10*	*0.20*	*0.30*	*0.40*	*0.50*	*0.60*	*0.70*	*0.80*	*0.90*	*0.95*	*0.99*	*0.999*
5	0.0385	0.0952	0.190	0.268	0.402	0.532	0.672	0.829	1.02	1.25	1.59	2.16	2.74	4.10	6.01
6	0.0572	0.1225	0.229	0.316	0.461	0.601	0.749	0.916	1.11	1.36	1.71	2.30	2.90	4.30	6.29
7	0.0740	0.1462	0.263	0.356	0.510	0.658	0.812	0.985	1.19	1.45	1.81	2.42	3.03	4.46	6.45
8	0.0870	0.1670	0.293	0.392	0.554	0.708	0.868	1.048	1.26	1.52	1.89	2.51	3.13	4.57	6.59
9	0.1014	0.1865	0.319	0.423	0.592	0.751	0.916	1.100	1.32	1.59	1.96	2.59	3.22	4.69	6.70
10	0.1130	0.2041	0.344	0.451	0.626	0.790	0.960	1.148	1.37	1.64	2.02	2.66	3.30	4.77	6.83
12	0.1380	0.2370	0.387	0.502	0.686	0.857	1.034	1.228	1.46	1.74	2.13	2.78	3.43	4.94	7.08
14	0.1575	0.2634	0.422	0.542	0.734	0.911	1.095	1.295	1.53	1.82	2.22	2.88	3.54	5.06	7.20
16	0.1740	0.2873	0.453	0.578	0.775	0.957	1.145	1.350	1.59	1.88	2.29	2.96	3.63	5.16	7.31
18	0.1900	0.3094	0.480	0.610	0.814	1.000	1.191	1.400	1.64	1.94	2.35	3.04	3.71	5.25	7.41
20	0.2039	0.3268	0.504	0.637	0.844	1.034	1.229	1.441	1.69	1.99	2.41	3.10	3.78	5.32	7.48
25	0.2367	0.3683	0.555	0.694	0.910	1.108	1.309	1.527	1.78	2.09	2.51	3.21	3.90	5.47	7.67
30	0.2589	0.3986	0.596	0.741	0.964	1.167	1.373	1.597	1.85	2.17	2.60	3.31	4.01	5.58	7.78
40	0.3001	0.4492	0.660	0.811	1.047	1.259	1.473	1.703	1.97	2.29	2.73	3.46	4.17	5.78	7.97
50	0.3298	0.4879	0.707	0.866	1.108	1.326	1.544	1.781	2.05	2.38	2.82	3.55	4.27	5.89	8.13
70	0.3756	0.5460	0.778	0.943	1.195	1.420	1.645	1.887	2.16	2.50	2.96	3.70	4.42	6.05	8.31
100	0.4208	0.6040	0.850	1.022	1.285	1.517	1.750	1.998	2.28	2.62	3.09	3.84	4.57	6.23	8.52
150	0.4728	0.6665	0.926	1.108	1.381	1.623	1.861	2.117	2.41	2.76	3.22	3.99	4.73	6.40	8.67
200	0.5075	0.7091	0.977	1.164	1.444	1.689	1.933	2.191	2.48	2.84	3.31	4.08	4.83	6.51	8.84
300	0.5564	0.7702	1.051	1.243	1.531	1.782	2.031	2.296	2.59	2.95	3.43	4.21	4.96	6.65	8.97
500	0.6174	0.8414	1.134	1.335	1.633	1.891	2.145	2.415	2.72	3.08	3.56	4.35	5.10	6.80	9.16
1000	0.6859	0.9332	1.239	1.449	1.757	2.024	2.284	2.557	2.87	3.23	3.72	4.52	5.28	6.99	9.35

5. The distributions of Z_A, Z_C and Z_K

Like the Anderson–Darling A^2, Cramér–von Mises W^2 and Kolmogorov–Smirnov K_S, the new statistics Z_A, Z_C and Z_K are distribution free. Our simulation for skewness and kurtosis shows that the sampling distributions of Z_A, Z_C and Z_K converge very slowly. Therefore, it is of limited practical value to study their asymptotic distributions. Just as for A^2, W^2 and K_S, it is difficult to find their exact null distributions for finite sample cases except for small sample sizes.

Again Monte Carlo simulation is used to approximate the percentage points of Z_A, Z_C and Z_K for some selected sample sizes. Tables 1, 2 and 3 respectively give their approximate percentage points at different levels, which are based on a simulation of size 1 million. The simulation error can be estimated in terms of percentage levels rather than percentage points. Specifically, the standard error or deviation of the simulation at percentage level α is $\sqrt{\{\alpha(1-\alpha)/N\}}$, where $N = 1000000$ is the number of replicates of the simulation.

6. Concluding remarks

In this paper, we have established powerful goodness-of-fit tests for the basic situation where the hypothetical distribution $F_0(x)$ is completely known. If $F_0(x)$ has some unknown parameters, we need to estimate the parameters first and then to apply the tests. This is a common approach for general goodness-of-fit tests for parametric models. However, the test statistics are then no longer distribution free. In such a case, we are testing the goodness of fit for a family of distributions rather than a specific distribution. As a result, for different families, the sampling distributions of the statistics are different.

For example, if $F_0(x) = \Phi\{(x-\mu)/\sigma\}$, the distribution function of a normal population $N(\mu, \sigma^2)$ with μ and σ^2 unknown, we can use the sample mean and the sample variance to estimate μ and σ^2 respectively. Then Z_A, Z_C and Z_K in equations (3.1)–(3.3) can be applied to test the goodness of fit for normality. In such a case, our simulation results (Zhang, 2001) show that Z_A and Z_C outperform the best tests of normality in the literature, including the Shapiro–Wilk W (Shapiro and Wilk, 1965) and the Anderson–Darling A^2; see D'Agostino and Stephens (1986).

The parameterization approach in this paper has been developed and applied to the general two-sample and even k-sample tests, where the underlying distribution of each population is totally unknown. Parallel results are obtained (Zhang, 2001).

Acknowledgements

This work is mainly a part of the author's doctoral thesis. I would like to thank my supervisor Professor Yuehua Wu at York University for all her help and support. I also thank the Joint Editor, Associate Editor and referees for their helpful comments, suggestions and corrections. The research is partially supported by the Natural Science and Engineering Research Council of Canada.

References

Anderson, T. W. and Darling, D. A. (1952) Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *Ann. Math. Statist.*, **23**, 193–212.

———(1954) A test of goodness of fit. *J. Am. Statist. Ass.*, **49**, 765–769.

Cabaña, A. (1996) Transformations of the empirical measure and Kolmogorov-Smirnov tests. *Ann. Statist.*, **24**, 2020–2035.

Conover, W. J. (1999) *Practical Nonparametric Statistics*, 3rd edn. New York: Wiley.
Cramér, H. (1928) On the composition of elementary errors: II, Statistical applications. *Skand. Akt.*, **11**, 141–180.
Cressie, N. and Read, T. R. C. (1984) Multinomial goodness-of-fit tests. *J. R. Statist. Soc.* B, **46**, 440–464.
Csörgő, S. and Faraway, J. J. (1996) The exact and asymptotic distributions of Cramér–von Mises statistics. *J. R. Statist. Soc.* B, **58**, 221–234.
D'Agostino, R. B. and Stephens, M. A. (1986) *Goodness-of-fit Techniques*. New York: Dekker.
Dudewicz, E. J. and van der Meulen, E. C. (1981) Entropy-based tests of uniformity. *J. Am. Statist. Ass.*, **76**, 967–974.
Gibbons, J. D. (1992) *Nonparametric Statistical Inference*, 3rd edn. New York: Dekker.
Knott, M. (1974) The distribution of the Cramér–von Mises statistic for small sample sizes. *J. R. Statist. Soc.* B, **36**, 430–438.
Kolmogorov, A. (1993) Sulla determinazione empirica di una legge di distribuzione. *G. Inst. Ital. Att.*, **4**, 83–91.
Massey, F. J. (1951) The Kolmogorov-Smirnov test for goodness of fit. *J. Am. Statist. Ass.*, **46**, 68–77.
MathSoft (2000) *S-Plus 2000 Guide to Statistics*, vol. 1, p. 100. Seattle: MathSoft.
von Mises, R. (1931) *Wahrscheinlichkeitsrechnung und ihre Anwendung in der Statistik und Theoretischen Physik*. Leipzig: Deuticke.
Pratt, J. W. and Gibbons, J. D. (1981) *Concepts of Nonparametric Theory*. New York: Springer.
Read, T. R. C. and Cressie, N. (1988) *Goodness-of-fit Statistics for Discrete Multivariate Data*. New York: Springer.
Shapiro, S. S. and Wilk, M. B. (1965) Analysis of variance test for normality (complete samples). *Biometrika*, **52**, 591–611.
Sinclair, C. D. and Spurr, B. D. (1988) Approximations to the distribution function of Anderson-Darling test statistic. *J. Am. Statist. Ass.*, **83**, 1190–1191.
Smirnov, N. V. (1936) Sur la distribution de ω^2 (critérium de M. R. v. Mises). *C. R. Acad. Sci. Paris*, **202**, 449–452.
———(1937) On the distribution of Mises' ω^2-criterion (in Russian). *Mat. Sb. Nov. Ser.*, **2**, 973–993.
———(1939) Estimate of derivation between empirical distribution functions in two independent samples (in Russian). *Bull. Mosk. Univ.*, **2**, 3–16.
Spinelli, J. J. and Stephens, M. A. (1997) Cramér-von Mises tests of fit for the Poisson distribution. *Can. J. Statist.*, **25**, 257–268.
Stephens, M. A. (1970) Use of the Kolmogorov–Smirnov, Cramér–von Mises and related statistics without extensive tables. *J. R. Statist. Soc.* B, **32**, 115–122.
———(1974) EDF statistics for goodness-of-fit and some comparisons. *J. Am. Statist. Ass.*, **69**, 730–737.
Zhang, J. (2001) Powerful goodness-of-fit and multi-sample tests. *PhD Thesis*. York University, Toronto.