On distribution-free goodness-of-fit testing of exponentiality

John Haywood* and Estate Khmaladze
Victoria University of Wellington, New Zealand

31 October 2005

Abstract

Local, or better, contiguous alternatives are the closest alternatives against which it is still possible to have some power. With this in mind we would like to think of goodness-of-fit tests as those which have some power against all, or a huge majority, of local alternatives. Tests of that kind are often based on nonlinear functionals, with a complicated asymptotic null distribution. Therefore a second desirable property of a goodness-of-fit test is that its statistic will be asymptotically distribution free.

Goodness-of-fit testing of exponentiality has a very long history and has produced a large number of papers. However, surprisingly many tests have been suggested based on asymptotically linear functionals from the empirical process; hence they can not be considered as goodness-of-fit tests. Such tests will have very low or no asymptotic power against a great majority of local alternatives, although they may have good power against some focused 'cone' of alternatives.

We suggest potentially a whole class of goodness-of-fit tests with both of the desirable properties mentioned above, by constructing a new version of the empirical process that weakly converges to a standard Brownian motion, under the hypothesis of exponentiality. Any statistic based on this process will asymptotically behave as a statistic from the standard Brownian motion and, hence, will be asymptotically distribution free. The idea is not new, but the form of transformation is especially simple in the case of exponentiality. As we note in the paper, this is not the only asymptotically distribution free version of empirical process for this problem, but it is one with a most convenient limit distribution.

Keywords: Asymptotically linear statistics; Brownian motion; Contiguous alternatives; Distribution free statistics; Goodness of fit statistics; Local alternatives; Rate of convergence to Brownian motion; Transformed empirical process

* *Address for correspondence*: John Haywood, School of Mathematics, Statistics and Computer Science, Victoria University of Wellington, PO Box 600, Wellington, New Zealand.
E-mail: John.Haywood@mcs.vuw.ac.nz

1 Introduction

There is no exact definition of goodness-of-fit tests, or goodness-of-fit test statistics. However, intuitively such tests are supposed to be of omnibus nature and be able to detect 'all sorts of deviations' from the hypothesis of interest. Consider one possible definition in the case of testing a parametric hypothesis, of which testing exponentiality is a special case.

Given a random sample $X_1, X_2, \ldots, X_n$, where each X_i has the same but unknown distribution function P, let P_n denote an empirical distribution function based on this sample. Suppose we are testing the null hypothesis that $P \in \mathcal{P}$ where $\mathcal{P} = \{P_\lambda, \lambda \in \mathbb{R}^d\}$ is a given class of distribution functions, indexed by a finite-dimensional parameter λ. As the class of alternative distributions we consider the class of all *converging contiguous alternatives* (as opposed to *chimeric* contiguous alternatives; see Khmaladze (1981)), which are defined as the class of all sequences of distributions A_n, for which there exists $P_\lambda \in \mathcal{P}$ and the function $h(\cdot, \lambda) \in L_2(P_\lambda)$ such that

$$\left(\frac{dA_n}{dP_\lambda}\right)^{\frac{1}{2}}(x) = 1 + \frac{1}{2\sqrt{n}} h(x,\lambda) + r_n(x), \qquad n \int r_n^2(x) dP_\lambda(x) \to 0. \tag{1}$$

The function h describes the 'direction' from which alternatives A_n converge to a particular sub-hypothesis P_λ.

From a geometric point of view it is more or less clear that one should not be concerned to detect alternatives which approach any particular P_λ from the direction tangential to $\mathcal{P}$ at P_λ. Formally it means that in (1) the function h must be orthogonal to this direction, or:

$$\int h(x,\lambda) \frac{\partial \ln p(x,\lambda)}{\partial \lambda} dP_\lambda(x) = 0,$$

where $\partial \ln p(x,\lambda)/\partial\lambda$ is the score function under the null hypothesis. Then we call $\psi_n(P_n, \mathcal{P})$ a *goodness-of-fit statistic*, and we call the test based on it a *goodness-of-fit test*, if

(a) the asymptotic distribution of $\psi_n(P_n, \mathcal{P})$ under any $P \in \mathcal{P}$ does not depend on $\mathcal{P}$, and

(b) the asymptotic distribution of $\psi_n(P_n, \mathcal{P})$ under *any* sequence of contiguous alternatives A_n to $\mathcal{P}$ is different from its asymptotic distribution under the hypothesis, or, in other words, the test based on $\psi_n(P_n, \mathcal{P})$ has *some* power against any sequence of contiguous alternatives.

Goodness-of-fit statistics are typically based on the following version of empirical process, often called the parametric or estimated empirical process: choose $\hat{P} = P_{\hat{\lambda}}$ from $\mathcal{P}$ which fits the sample 'best' and study the normed difference,

$$v_n(x, \mathcal{P}) = \hat{v}_n(x) = \sqrt{n}\{P_n(x) - \hat{P}(x)\}.$$

We also denote the empirical process, $v_n(x) = \sqrt{n}\{P_n(x) - P(x)\}$. The 'usual' property of $\hat{v}_n$ is that under any P from the hypothetical $\mathcal{P}$, it converges in distribution to a zero-mean Gaussian process, while under any sequence of contiguous alternatives it converges to the same Gaussian process, but with a non-zero shift. From this stems the requirement (b), which says that if $\psi_n(P_n, \mathcal{P})$ is some functional from $\hat{v}_n$, which typically it is, then it should 'react' to the presence of any non-zero shift. One can take the Kolmogorov-Smirnov statistic,

$$\sup_x |\hat{v}_n(x)| = \sqrt{n} \sup_x |P_n(x) - \hat{P}(x)|,$$

as an example of such a statistic. However, the requirement (a) may still be a problem.

A widely used counterpart of goodness-of-fit statistics is formed by so called *asymptotically linear statistics*. These are statistics of the form

$$\psi_n(P_n, \mathcal{P}) = \int g(x, \hat{\lambda}) \hat{v}_n(dx) + o_p(1), \tag{2}$$

for some deterministic function g, usually with certain square-integrability requirements. While converging to a zero-mean Gaussian random variable under the hypothesis (i.e. under any $P_\lambda \in \mathcal{P}$), these statistics converge to the same Gaussian random variable plus the shift,

$$\int h(x, \lambda) g(x, \lambda) dP_\lambda(x) \tag{3}$$

under any sequence of contiguous alternatives A_n satisfying (1); see, e.g., Khmaladze (1979), Janssen (2000). We see that, although for $h(x, \lambda) = const \cdot g(x, \lambda)$ the shift (3) can be quite large, for $h(\cdot, \lambda) \perp g(\cdot, \lambda)$ this shift is simply zero and hence any asymptotically linear statistic under all these alternatives will have the same limiting distribution as under the hypothesis. So for this very wide class of alternatives, a test based on any of these asymptotically linear statistics will have asymptotically no power at all.

The term 'goodness-of-fit tests' has been taken to cover a wider class of tests than just of the types already described above. For example, many tests, each based on several asymptotically linear statistics, say, $\int g_j(x) \hat{v}_n(dx), \quad j = 1, \ldots, k$, have also been considered and are perceived as goodness-of-fit tests. The chi-square test is probably the most prominent such example. Tests of this type do not have the property (b) formally. However, for such tests to have no power asymptotically, the function h would need to be orthogonal to each $g_j, \quad j = 1, \ldots, k$, which can rarely occur in practice.

We comment on a selection of asymptotically linear statistics in the next section, including in each case the explicit demonstration of their linear form. Of considerable interest is the fact that different well known statistics are all effectively the same, asymptotically. Hence if any of the linear statistics considered are such that the shift (3) is very small or zero, tests based on all of these similar statistics will have asymptotically very small or no power. We give an example that illustrates this situation.

Section 3 contains the key elements of this paper. Namely, we present a particularly simple transformation of the parametric empirical process for the case of exponentiality, from which a variety of test statistics can be calculated. All such statistics are distribution free, and asymptotically behave, under the hypothesis, as statistics from a standard Brownian motion. This makes asymptotic theory of such statistics convenient and relatively simple. We demonstrate good finite sample convergence properties of representative test statistics calculated from our transformed empirical process in section 4, while in section 5 we give some brief concluding remarks.

2 Comments on linear statistics

It is remarkable that many papers explicitly or implicitly devoted to goodness-of-fit testing of exponentiality are actually based on a single asymptotically linear statistic. Before we consider examples of such usage we need to make one observation. Suppose $g_\perp(\cdot, \lambda)$ is the part of $g(\cdot, \lambda)$ orthogonal to the hypothetical score function $\partial \ln p(x, \lambda)/\partial\lambda$. Since all functions $h(\cdot, \lambda)$ are orthogonal to it we have

$$\int h(x, \lambda) g(x, \lambda) dP_\lambda(x) = \int h(x, \lambda) g_\perp(x, \lambda) dP_\lambda(x).$$

Therefore it will be inefficient to use the (asymptotically) linear statistics with $g(\cdot, \lambda) \neq g_\perp(\cdot, \lambda)$: the part of it equal to $\int \{g(x, \hat{\lambda}) - g_\perp(x, \hat{\lambda})\} \hat{v}_n(dx)$ will contribute to its asymptotic variance but not to its asymptotic shift. Therefore, we will choose below kernels g such that $g(\cdot, \lambda) = g_\perp(\cdot, \lambda)$. Further, it will be convenient to subtract $\int g(x, \lambda) dP_\lambda(x)$ from the kernel, which we will do in each case without explicitly mentioning it. Since $\int const \cdot \hat{v}_n(dx) = 0$ this does not change the linear statistic.

Given an average $n^{-1} \sum_{i=1}^n g(X_i, \hat{\lambda}) = \int g(x, \hat{\lambda}) dP_n(x)$, to determine its asymptotic behaviour under any sub-hypothesis P_λ it would be quite possible to centre it with $\int g(x,\hat\lambda)dP_\lambda(x)$. In this way we have

$$\sqrt{n}\left\{\int g(x,\hat\lambda)dP_n(x)-\int g(x,\hat\lambda)dP_\lambda(x)\right\}=\int g(x,\hat\lambda)v_n(dx)=\int g(x,\lambda)v_n(dx)+o_p(1)$$

under mild regularity assumptions on $g(\cdot,\lambda)$ with respect to λ. However, the simple and practical way to represent this average as an asymptotically linear functional with the kernel $g_\perp(\cdot,\hat\lambda)$ is to centre it with $\int g(x,\hat\lambda)dP_{\hat\lambda}(x)$. Indeed, in this way we obtain:

$$\begin{aligned}&\sqrt{n}\left\{\int g(x,\hat\lambda)dP_n(x)-\int g(x,\hat\lambda)dP_{\hat\lambda}(x)\right\}=\int g(x,\hat\lambda)\hat v_n(dx)\\&=\int g(x,\lambda)v_n(dx)-\int g(x,\lambda)\frac{\partial\ln p(x,\lambda)}{\partial\lambda}dP_\lambda(x)\sqrt{n}(\hat\lambda-\lambda)+o_p(1).\end{aligned}$$

If for the estimator $\hat\lambda$ the asymptotic representation

$$\sqrt{n}(\hat\lambda-\lambda)=\left\{\int l(x,\lambda)\frac{\partial\ln p(x,\lambda)}{\partial\lambda}dP_\lambda(x)\right\}^{-1}\int l(x,\lambda)dv_n(x)+o_p(1)$$

is true, then eventually we obtain

$$\int g(x,\hat\lambda)\hat v_n(dx)=\int g_\perp(x,\lambda)v_n(dx)+o_p(1),$$

where

$$g_\perp(x,\lambda)=g(x,\lambda)-\int g(x,\lambda)\frac{\partial\ln p(x,\lambda)}{\partial\lambda}dP_\lambda(x)\left\{\int l(x,\lambda)\frac{\partial\ln p(x,\lambda)}{\partial\lambda}dP_\lambda(x)\right\}^{-1}l(x,\lambda)$$

is the part of $g(\cdot,\lambda)$ orthogonal to $\partial\ln p(\cdot,\lambda)/\partial\lambda$ as required.

Consider now several well known statistics. In what follows we denote by P the exponential distribution function P_λ, $P_\lambda(x)=1-e^{-\lambda x}$, with $\lambda=1$.

The papers of Deshpande (1983) and Bandyopadhyay and Basu (1989) are based on testing whether $1-P(bx)=\{1-P(x)\}^b$, and the test statistic can be written as the integral

$$\frac{1}{n(n-1)}\sum_{i\neq j}\mathbb{I}\{X_j>bX_i\}=\int_0^\infty\{1-P_n(bx+)\}P_n(dx),\tag{4}$$

where $\mathbb{I}\{\cdot\}$ denotes the indicator function. This statistic is scale invariant and hence is independent of the parameter of the exponential distribution. For any fixed b, statistic (4) is, in fact, asymptotically linear:

$$\begin{aligned}&\sqrt{n}\left\{\int_0^\infty\{1-P_n(bx+)\}P_n(dx)-\int_0^\infty\{1-P(bx)\}dP(x)\right\}\\&=\int_0^\infty\{\exp(-bx)-\exp(-x/b)\}v_n(dx)+o_p(1)=\int_0^\infty g_{\perp,D}(x)v_n(dx)+o_p(1)\end{aligned}\tag{5}$$

where

$$g_{\perp,D}(x) = \exp(-bx) - \exp(-x/b) - (1-b)/(1+b).$$

Deshpande (1983) and Bandyopadhyay and Basu (1989) each suggest different choices of b, but it is one fixed value of b in both cases. It may be natural to study the right hand side of (5) as a process in b. Based on the combination of the Laplace transform of the empirical process $\hat{v}_n$, this process in b will converge to a zero-mean Gaussian process under the hypothesis, and will have non-zero shift under all contiguous alternatives A_n. Basically such an approach was taken and extended in Baringhaus and Henze (1991), Henze (1993) and Baringhaus and Henze (2000), where quadratic functionals from not only the Laplace transform but also from the Fourier transform of $\hat{v}_n$ were studied. Epps and Pulley (1986) previously investigated 'smooth' functionals from the empirical characteristic function.

An interesting statistic known as the Gini index,

$$G_n = \frac{\sum_{i \neq j} |X_i - X_j|}{2n(n-1)\bar{X}},$$

received a thorough treatment in Gail and Gastwirth (1978). The Gini index is also scale invariant, and can be rewritten as

$$G_n = \frac{n}{(n-1)} \frac{1}{2\bar{X}} \int \int |x-y| dP_n(x) dP_n(y).$$

After normalisation, it becomes evident that G_n is also an asymptotically linear statistic:

$$\begin{aligned}
& \frac{n}{(n-1)\bar{X}} \frac{\sqrt{n}}{2} \int \int |x-y| \{dP_n(x)dP_n(y) - dP(x)dP(y)\} \\
& + \frac{\sqrt{n}}{2} \left(\frac{1}{\bar{X}} - 1 \right) \int \int |x-y| dP(x) dP(y) \\
= & \int \left\{ \int |x-y| P(dy) \right\} v_n(dx) - \sqrt{n} \left(\frac{\bar{X}-1}{2} \right) + o_p(1) \qquad (6) \\
= & \int \{x - 1 + 2\exp(-x)\} v_n(dx) - \int \frac{x-1}{2} v_n(dx) + o_p(1) \\
= & \int g_{\perp,G}(x) v_n(dx) + o_p(1)
\end{aligned}$$

where

$$g_{\perp,G}(x) = \frac{x-1}{2} - 1 + 2\exp(-x).$$

Here we also set the parameter of the exponential distribution to 1, so $\bar{X} \to 1$, and we made use of $\int |x-y| \exp(-y) dy = x - 1 + 2\exp(-x)$ and $\int \int |x-y| \exp(-x) \exp(-y) dx dy = 1$. Therefore it is obvious that G_n is asymptotically normal, but Gail and Gastwirth (1978) contains much more: they show the exact distribution of G_n and demonstrate that the convergence to the normal distribution is quick.

The so called Moran statistic, introduced in Moran (1951), has the form

$$M_n = \frac{1}{n}\sum_{i=1}^{n} \ln \frac{X_i}{\bar{X}}.$$

This is the statistic of the locally most powerful test against Gamma distributions with the density $a(x) = \lambda^{-\alpha}\Gamma^{-1}(\alpha)x^{\alpha-1}\exp(-x/\lambda)$ at $\alpha = 1$. Indeed, the score function in this case is $\partial \ln a(x)/\partial\alpha = \ln(x/\lambda) - \Gamma'(1)/\Gamma(1)$. Recently Tchirina (2002) studied the Bahadur efficiency of a test based on the Moran statistic and some of its local properties. The statistic is scale invariant and we can assume therefore that $\lambda = 1$. Following the method described above, for the normalised form of this statistic we obtain

$$\sqrt{n}\left\{\int \ln \frac{x}{\bar{X}} dP_n(x) - \int \ln x \, dP(x)\right\} = \int g_{\perp,M}(x) v_n(dx) + o_p(1) \tag{7}$$

where

$$g_{\perp,M}(x) = \ln x - \gamma - (x-1)$$

and $\gamma = \int \ln x \exp(-x)dx$, so that $-\gamma$ is Euler's constant. This statistic is suggested as an "omnibus" goodness-of-fit test by Stephens (1986), although he adds the caveat that its use can be "risky", due to problems if the data includes ties or zeros. For an omnibus test, Stephens in fact restricts attention to alternatives with monotone failure rate.

One more statistic used for the same purpose and which is also asymptotically linear and scale invariant is the statistic suggested in Cox and Oakes (1984),

$$T_n = n^{-1}\sum_{i=1}^{n}(1 - \frac{X_i}{\bar{X}}) \ln \frac{X_i}{\bar{X}}. \tag{8}$$

Its representation as an asymptotically linear statistic, taking $\lambda = 1$, is

$$\int g_{\perp,CO}(x) v_n(dx) + o_p(1) \quad \text{where} \quad g_{\perp,CO}(x) = (1-x)(\ln x + 1 - \gamma). \tag{9}$$

Statistic (8) was introduced explicitly as a locally most powerful test against Weibull alternatives $A(x) = 1 - \exp(-x/\lambda)^{\alpha}$ at $\alpha = 1$. Indeed, the function $(1 - x/\lambda)\ln(x/\lambda) + 1$ is the score function of the Weibull density $a(x) = \alpha/\lambda(x/\lambda)^{\alpha-1}\exp(-x/\lambda)$ with respect to parameter α at $\alpha = 1$. Later however, see for example, Ascher (1990) and Henze and Meintanis (2005), the Cox and Oakes statistic has been studied (and recommended) for goodness-of-fit testing.

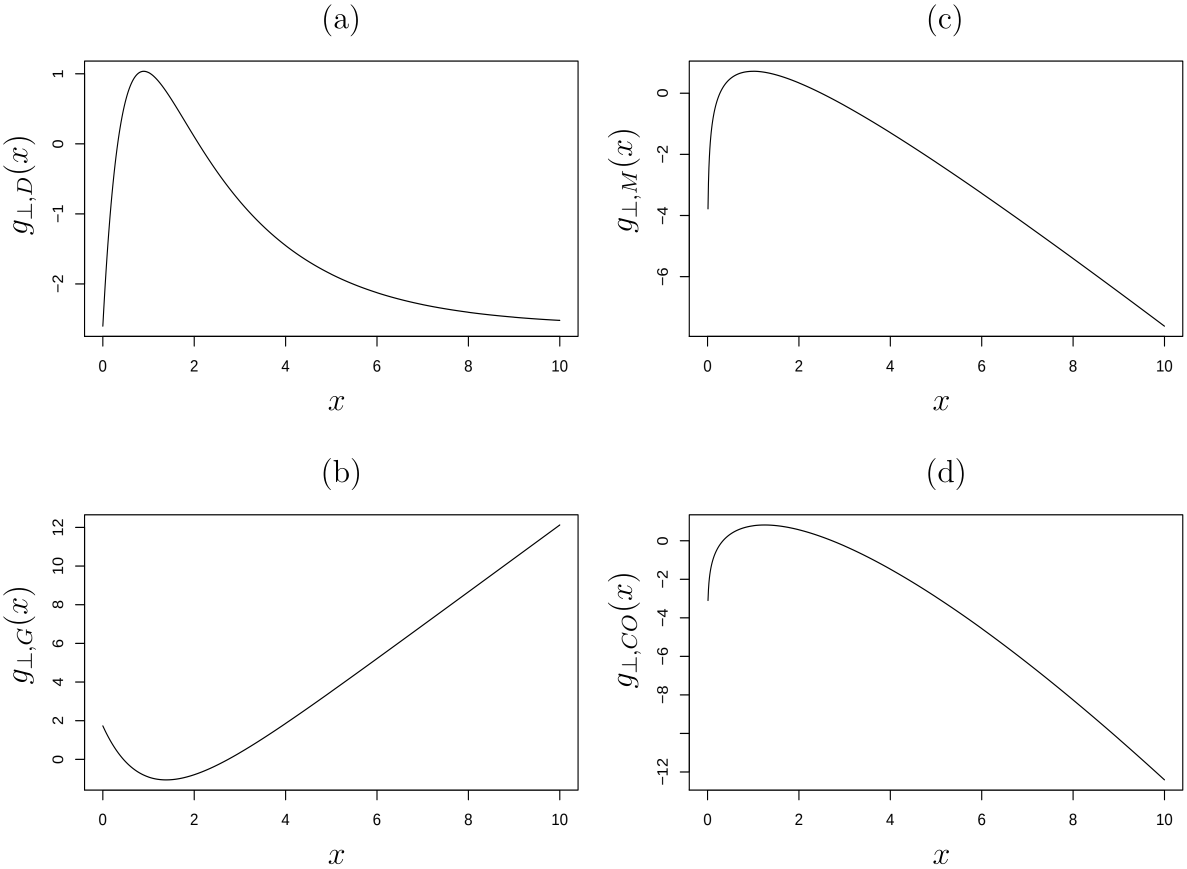

Figure 1: The kernel functions of the four asymptotically linear statistics discussed in section 2: (a) Deshpande; (b) Gini; (c) Moran; (d) Cox-Oakes.

Considering the four asymptotically linear statistics above, it is possible to get an impression that we have many different statistics, each capable of detecting deviations from exponentiality in different directions. Thus it might appear important and interesting to compare their asymptotic behaviour against various alternatives.

However, as far as local alternatives are concerned, all four statistics are extremely similar. Though formally the corresponding kernel functions $g_\perp$ are different they all have very similar graphs, as shown in Figure 1, and all four statistics are very highly correlated. In particular, the correlations between statistic (9) and (5), (6), (7) are $0.909, -0.936, 0.971$; between statistic (7) and (5), (6) are $0.909, -0.833$; and between statistic (6) and (5) -0.865. Note that, following the recommendation of Bandyopadhyay and Basu (1989), we use a value of $b = 0.44$ in Deshpande's (1983) statistic (4).

Therefore in reality all four statistics lead, asymptotically, to more or less the same test and it is of little importance to study their relative power, at least for local alternatives.

Furthermore, as noted above, no asymptotically linear statistic can provide a goodness-of-fit test with property (b) or similar to it. To illustrate this we now provide an explicit example of a local alternative for which all four above mentioned statistics have indeed very low power, while the power of an omnibus test (calculated on the transformed empirical process introduced in §3) is essentially higher. Perhaps somewhat unexpectedly, for such an example we use the uniform distribution on $[0,1]$, against which the Cox-Oakes statistic, for example, was reported to have high power (see Ascher (1990) and Henze and Meintanis (2005)). Let us convert this uniform distribution into a local alternative by using the mixture

$$a_n(x) = \left(1 - \frac{c}{\sqrt{n}}\right) p(x) + \frac{c}{\sqrt{n}} q(x) = p(x) + \frac{c}{\sqrt{n}}\{q(x) - p(x)\}$$

where c is a scalar, $p(x)$ is the exponential density with parameter $\lambda = 1$ and $q(x)$ is the uniform density on $[0,1]$. Denote $\|g_{\perp,CO}\| = \left(\int g_{\perp,CO}^2(x)e^{-x}dx\right)^{1/2}$. One can check that the value of the integral

$$\frac{1}{\|g_{\perp,CO}\|}\int g_{\perp,CO}(x)\{q(x) - p(x)\}dx = \frac{1}{\|g_{\perp,CO}\|}\int_0^1 g_{\perp,CO}(x)dx = 0.0138$$

is quite small. Hence the asymptotic shift of the standardized Cox-Oakes statistic under alternatives a_n, being $0.0138c$, is also small (for moderate c), which means that its power against a_n will be very low.

Indeed, Figure 2 compares the distributions of the Cox-Oakes statistic and the Kolmogorov-Smirnov statistic d_n (see §4) calculated from the transformed empirical process w_n given in (11) below. The statistics are compared under the hypothesis and under the alternative a_n above, using 20,000 replications of size $n = 900$ with $c = 5$. (That is, under the alternative, 150 of the 900 observations in each replicate sample come from the uniform distribution.) It is very clear that the change in distribution of the Kolmogorov-Smirnov statistic is essentially sharper, and the power of the Kolmogorov-Smirnov test based on the transformed version of the empirical process is larger.

Another example was found in our recent research on certain historical data (Khmaladze *et al.*, 2005). Figure 3 shows the empirical distribution function of the durations of rule for Chinese Emperors, as given in Encyclopaedia Britannica (2002), along with its exponential approximation. The approximation looks very good and the value of the standardized Cox-Oakes statistic is 0.08, which is obviously close to the median of the standard normal distribution. However, the value of the Kolmogorov-Smirnov statistic from the transformed

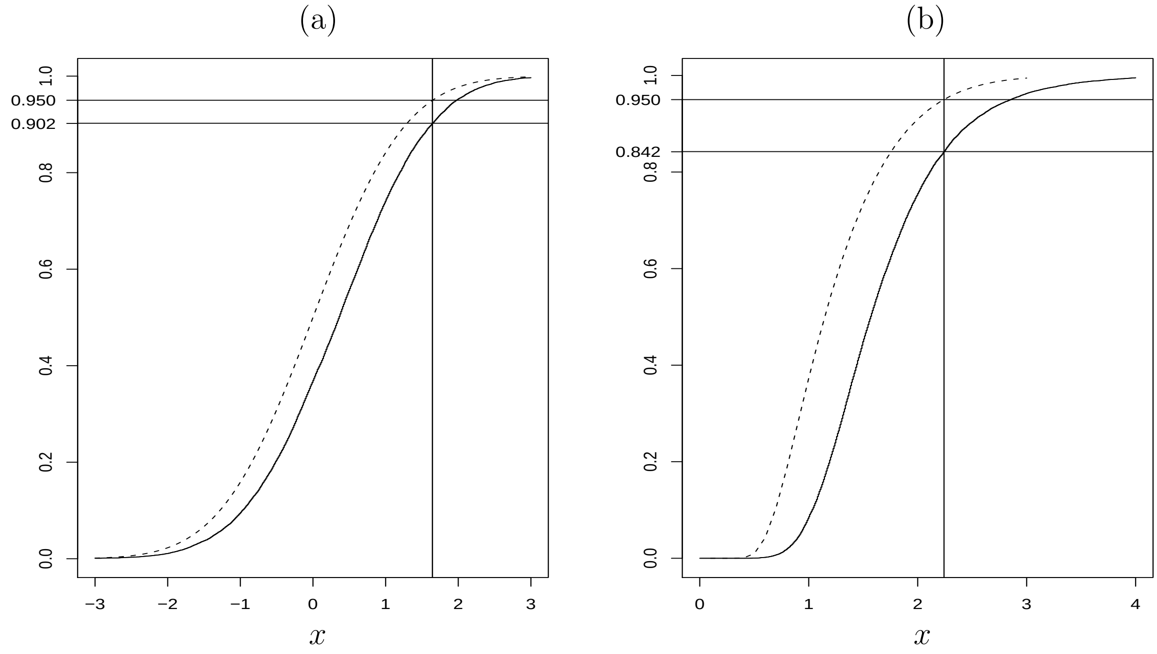

Figure 2: The change in distribution of (a) the Cox-Oakes statistic and (b) the Kolmogorov-Smirnov statistic d_n from the transformed empirical process w_n, calculated from 20,000 samples of size 900. Distributions are shown under the hypothesis of exponentiality (dashed line) and under the local alternative mixture discussed in the text (solid line).

empirical process is equal to $d_n = 2.46$, which corresponds to a p-value of 0.03. Hence the hypothesis of exponentiality is rejected using the statistic based on the transformed empirical process, as we know it should be from more detailed analysis reported in Khmaladze *et al.* (2005).

3 Tests with a transformed version of empirical process

Although we mentioned the Kolmogorov-Smirnov test explicitly in the previous sections, in this paper we are not suggesting or advocating any particular test. What we are suggesting is a version of parametric empirical process w_n, different from $\hat{v}_n$. This process w_n, after the time transformation

$$t = P_{\hat{\lambda}}(x), \tag{10}$$

converges in distribution under the hypothesis to standard Brownian motion and converges to standard Brownian motion plus non-zero shift under any sequence of contiguous alternatives (1). Therefore any statistic that is invariant under the time transformation (10)

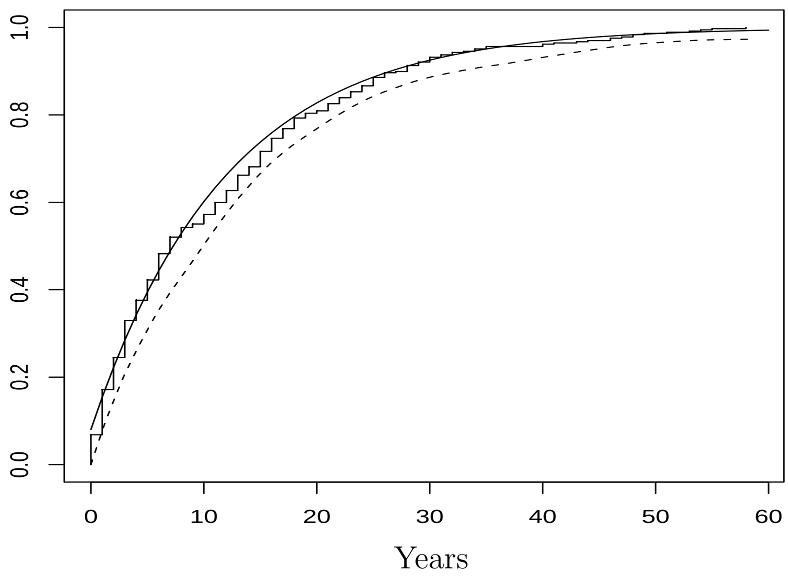

Figure 3: The empirical distribution function of durations of rule of Chinese Emperors (step function) along with an exponential approximation (solid line) and the compensator $K(x, P_n)$, introduced in §3 (dashed line).

is asymptotically distribution free, i.e. has property (a), and any statistic of omnibus nature, like the Kolmogorov-Smirnov statistic or the Cramér-von Mises or Anderson-Darling statistics, will also possess the property (b) from section 1 above.

The process w_n differs from $\hat{v}_n$ in the way one centres the empirical distribution function P_n. The idea of this different centering was originally presented in Khmaladze (1981) and is well known now, but for convenience we include a brief description. Namely, instead of subtracting from P_n its expected value P_λ with estimated parameter, or, in differential form, instead of considering $\hat{v}_n(dx) = \sqrt{n}\{P_n(dx) - P_{\hat{\lambda}}(dx)\}$, one could instead use conditional expected values and consider

$$\sqrt{n}\{P_n(dx) - E(P_n(dx)|P_n(y), y \leq x, \hat{\lambda})\}.$$

This centering of each increment, $P_n(dx)$, by the conditional expected value will render w_n to be a martingale with respect to the filtration $\{\mathfrak{F}_x, x \geq 0\}$, where each σ-algebra $\mathfrak{F}_x$ is generated by the 'past' of P_n and also the maximum likelihood estimator $\hat{\lambda} = n/\sum_{i=1}^n X_i$, i.e. $\mathfrak{F}_x = \sigma\{P_n(y), y \leq x, \sum_{i=1}^n X_i\}$. Although it may not be easy to calculate this conditional expected value exactly, it is much easier to calculate the asymptotically equivalent expression

$K(dx, P_n)$ such that

$$w_n(dx) = \sqrt{n}\{P_n(dx) - K(dx, P_n)\} \tag{11}$$

will be a process with uncorrelated increments. As the notation suggests, the centering process $K(x, P_n), x \geq 0$, is a certain modification of P_n. It is called the compensator (of P_n) – it 'compensates' P_n to the process with uncorrelated increments. The form of the compensator is suggested by the hypothetical family of distribution functions. Namely, let $\{p(x, \lambda), \lambda \in \Lambda\}$ be a family of corresponding density functions depending on some parameter λ. Consider the vector-function

$$q(x, \lambda)^T = \left[1\ ,\ \frac{\partial \ln p(x, \lambda)}{\partial \lambda}\right]$$

with the first coordinate identically equal to 1 and the second coordinate being the score function $\partial \ln p(x, \lambda)/\partial\lambda$. Using this extended score function construct the matrix

$$C(z, \lambda) = \int_z^\infty q(x, \lambda) q(x, \lambda)^T p(x, \lambda) dx,$$

which can be called the *incomplete Fisher information matrix*. Then the compensator associated with this parametric family is defined (see, e.g., Khmaladze (1981)) as

$$\begin{aligned} K(x, P_n) &= \int_0^x q(z, \lambda)^T C^{-1}(z, \lambda) \int_z^\infty q(y, \lambda) P_n(dy) P(dz) \\ &= \int_0^\infty \left\{ \int_0^{\min(x,y)} q(z, \lambda)^T C^{-1}(z, \lambda) P(dz) \right\} q(y, \lambda) P_n(dy). \end{aligned}$$

To see that $K(x, P_n)$ is really a 'modification' of $P_{\hat\lambda}(x)$, one can check that $K\{x, P_\lambda(x)\} = P_\lambda(x)$.

As our first observation in this section we note that the form of $K(x, P_n)$ in the case of exponential distributions is quite straightforward:

$$\begin{aligned} K(x, P_n) &= \hat\lambda \int_0^\infty \left(2 + \frac{\hat\lambda}{2}\min(x, y) - \hat\lambda y\right) \min(x, y) P_n(dy) \\ &= \hat\lambda \int_0^x \left(2 - \frac{\hat\lambda}{2} y\right) y P_n(dy) + \hat\lambda \left(2 + \frac{\hat\lambda}{2} x\right) x\{1 - P_n(x)\} - \hat\lambda^2 x \int_x^\infty y P_n(dy) \end{aligned}$$

or

$$K(x, P_n) = \frac{\hat\lambda}{n} \sum_{i: X_i \leq x} \left(2X_i - \frac{\hat\lambda}{2} X_i^2\right) + \hat\lambda \left(2 + \frac{\hat\lambda}{2} x\right) x\{1 - P_n(x)\} - x\frac{\hat\lambda^2}{n} \sum_{i: X_i > x} X_i\,.$$

As our second observation we note that the rate of convergence of w_n to a standard Brownian motion seems quite good; at least, as good as the convergence of the uniform empirical process to a Brownian bridge in the classical theory. We present some numerical results that support this claim in Section 4.

For testing exponentiality there exists another interesting form of empirical process, which is also asymptotically distribution free. In particular, using the equality $1 - P_\lambda(bx) = \{1 - P_\lambda(x)\}^b$ more fully than via a single statistic (cf. Deshpande (1983) for the latter case) one can consider the empirical process

$$\alpha_n(x) = -\sqrt{n}\left(1 - P_n(bx) - \{1 - P_n(x)\}^b\right)$$

and test if it is asymptotically a zero-mean process. It is easy to see that

$$\alpha_n(x) = \sqrt{n}\{P_n(bx) - P(bx)\} + \sqrt{n}\left[\{1 - P_n(x)\}^b - \{1 - P(x)\}^b\right]$$

and hence,

$$\alpha_n(x) = v_n(bx) - b\{1 - P_\lambda(x)\}^{b-1} v_n(x) + o_p(1). \tag{12}$$

Following, e.g., Angus (1982) and Nikitin (1996), choose $b = 2$. After the time transformation $t = 1 - P_{\hat{\lambda}}(x)$ the process α_n will converge in distribution to the process β,

$$\beta(t) = u(t^2) - 2tu(t), \tag{13}$$

where u is a standard Brownian bridge on $[0, 1]$. Hence we obtain a standard, or distribution free, process as the limit.

A more general version of this last process was studied earlier by Koul (1977, 1978). Starting with the equality $1 - P_\lambda(x + y) = \{1 - P_\lambda(x)\}\{1 - P_\lambda(y)\}$, Koul considered the process

$$\alpha_n(x, y) = -\sqrt{n}\left[1 - P_n(x + y) - \{1 - P_n(x)\}\{1 - P_n(y)\}\right].$$

The asymptotic form of Koul's process is

$$\alpha_n(x, y) = v_n(x + y) - \{1 - P(x, \lambda)\}v_n(y) - \{1 - P(y, \lambda)\}v_n(x) + o_p(1)$$

and therefore, after the time transformation $t = 1 - P_{\hat{\lambda}}(x), s = 1 - P_{\hat{\lambda}}(y)$, it converges in distribution to β^*,

$$\beta^*(t, s) = u(ts) - tu(s) - su(t), \tag{14}$$

which is again a distribution free process (in t and s). We note that the process (14) is of a very appealing structure: it can be written as Πu where, for a function $f(t,s),\ (t,s) \in [0,1]^2$,

$$\Pi f(t,s) = f(t,s) - tf(1,s) - sf(t,1) + tsf(1,1)$$

is the projection of f on the class of functions (on $[0,1]^2$) equal to zero everywhere on the boundary of the unit square. This does not imply as yet, however, that Πu does not define u uniquely, because in (14) Π is applied to a very narrow class of functions of t and s given by $f(t,s) = u(ts)$. However, we will presently see that the definition of u is not unique.

Indeed, we can clarify the situation with the processes $\alpha_n(x)$ and $\alpha_n(x,y)$ further, and respectively with the processes (13) and (14), if we observe the following. As can be seen, the relationship between $\hat{v}_n$ and v_n is given by

$$\hat{v}_n(x) = v_n(x) - \frac{d}{d\lambda}P_\lambda(x)\sqrt{n}(\hat{\lambda} - \lambda) + o_p(1) = v_n(x) + x\exp(-\lambda x)\frac{1}{\lambda}\int_0^\infty x v_n(dx) + o_p(1)$$

if $\hat{\lambda} = 1/\bar{X}$. One can show that the main term on the right hand side above is a projection of v_n (see the general description of parametric empirical processes as projections in Khmaladze (1979) and similarly for the parametric regression process in Khmaladze and Koul (2004)). At the same time the functions $const \cdot dP_\lambda(x)/d\lambda = const \cdot x\exp(-\lambda x)$ are annulated by the transformation given by the right hand side of (12). These are the only functions which are annulated by this transformation: for a function $f(x)$,

$$\text{if} \quad f(2x) - 2\exp(-\lambda x)f(x) = 0, \quad \text{then} \quad \exp(2\lambda x)f(2x) = 2\exp(\lambda x)f(x)$$

which implies that $\exp(\lambda x)f(x) = const \cdot x$ and hence $f(x) = const \cdot x\exp(-\lambda x)$. Therefore, although the process α_n looks like a nonparametric object, unconnected with and free from any estimation of the parameter λ, asymptotically it is actually a one-to-one transformation of $\hat{v}_n$. It follows that (14) annulates $const \cdot t\ln t$ and β^* is not a one-to-one transformation of u.

Angus (1982) and Koul (1977) calculated some limiting critical values of Kolmogorov-Smirnov statistics from $\alpha_n(x)$ and $\alpha_n(x,y)$ respectively. However, the distribution of the processes (13) and (14) is rather complicated, or at least, is less convenient than the distribution of a standard Brownian motion.

4 Assessment of test statistics based on the process w_n

It is not entirely clear how to best evaluate the convergence of distribution of a sequence of processes – in our case, the sequence of w_n – to the limiting process. We chose to consider several statistics with different behaviour: one-sided and two-sided Kolmogorov-Smirnov statistics,

$$d_n^+ = \sup_{0 \le x < \infty} w_n(x), \qquad d_n^- = -\inf_{0 \le x < \infty} w_n(x),$$
$$d_n = \max(d_n^+, d_n^-) = \sup_{0 \le x < \infty} |w_n(x)|,$$

the Cramér-von Mises statistic,

$$\omega_n^2 = \int_0^\infty w_n^2(x) dP_\lambda(x),$$

and a version of the Anderson-Darling statistic,

$$A_n^2 = \int_0^\infty \frac{w_n^2(x)}{P_\lambda(x)} dP_\lambda(x) .$$

We evaluated the distribution functions of all five statistics for finite n and compared these distribution functions with their limits in a simulation study.

The limit distribution function of d_n^+ and d_n^- is $2\Phi(z)-1$, where Φ is the standard normal distribution function, which follows from the reflection principle; see, e.g., Feller (1971). The limit distribution of d_n is given, e.g., in Shiryaev (1999) and Borodin and Salminen (2002), and was calculated by E. Shinjikashvili using the numerical method suggested and studied in Khmaladze and Shinjikashvili (2001). Tables of the limit distribution of ω_n^2 were given in Orlov (1972) and Martynov (1977). The limiting percentage points of A_n^2 are given in Deheuvels and Martynov (2003). In fact, we have now made all these limiting distributions, and others, available on the web site of the School of Mathematics, Statistics and Computer Science, VUW:

`http://www.mcs.vuw.ac.nz/~ray/Brownian/`

Random samples of sizes 50, 100 and 200 were generated from the exponential distribution with parameter $\lambda = 1$, such that $P = \exp(-x)$. We considered two situations: the less realistic case with λ assumed known and the situation where λ was estimated, using the maximum likelihood estimator, $\hat{\lambda} = 1/\bar{X}$. Table 1 reports the empirical sizes of the five named statistics above obtained from 50,000 replications, for nominal sizes of 5% and 10%, with $\lambda = 1$ assumed. Table 2 gives the corresponding empirical sizes, obtained with $\lambda = \hat{\lambda}$.

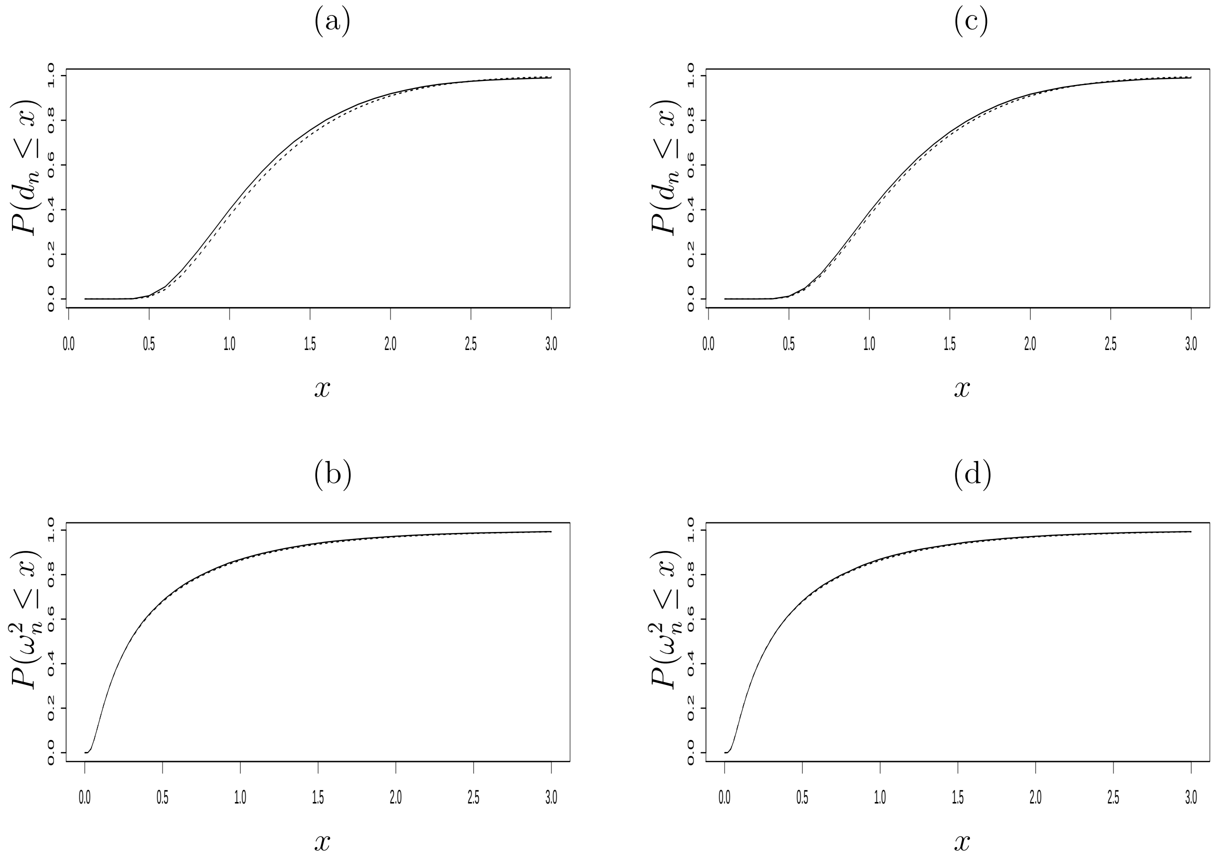

Figure 4: Comparison of empirical distribution function (solid line) and corresponding limit distribution (dashed line) for two goodness-of-fit test statistics based on the process w_n, calculated in each case from 50,000 standard exponential samples of sizes 50 or 100 with estimated parameters: (a) size 50, d_n; (b) size 50, ω_n^2; (c) size 100, d_n; (d) size 100, ω_n^2.

It is clear from the Tables 1 and 2 that all five statistics have accurate empirical sizes, both when λ is assumed known and, more realistically, when it is estimated. Such performance is required for a test statistic to be distribution-free. Figure 4 illustrates the good convergence of the entire empirical distribution functions of two of the test statistics to their limits: we present graphical comparisons for each of d_n and ω_n^2, for 50,000 samples of sizes 50 and 100, with $\lambda = \hat{\lambda}$. Convergence was comparable for the other test statistics, and slightly better with $\lambda = 1$ assumed known.

We remark that calculation of Kolmogorov-Smirnov statistics d_n^+, d_n^- and d_n becomes easy and quick, yet exact, if we note the following: within each interval $(X_{(j)}, X_{(j+1)})$ formed by adjacent order statistics the compensator $K(x, P_n)$ is simply a quadratic function in x and its minimum is attained at the point

$$x_j^0 = \frac{\sum_{i:X_i > X_{(j)}} X_i}{n-j} - 2\hat{\lambda} = \frac{\sum_{i:X_i > X_{(j)}} X_i}{n-j} - 2\bar{X},$$

while $P_n(x)$ stays constant. It follows that the maximum or minimum of the difference $P_n(x) - K(x, P_n)$ on each such interval is attained either at the end-points or at x_j^0, provided $x_j^0 \in (X_{(j)}, X_{(j+1)})$. In other words, $d_n^+ = \max(a_n, b_n)$ where

$$a_n = \max_j \left\{ \left(\frac{j}{n} \right) - K(X_{(j)}, P_n) \right\}, \quad \text{and} \quad b_n = \max_{j: x_j^0 \in (X_{(j)}, X_{(j+1)})} \left\{ \left(\frac{j}{n} \right) - K(x_j^0, P_n) \right\}$$

while

$$d_n^- = -\min_j \left\{ \left(\frac{j-1}{n} \right) - K(X_{(j)}, P_n) \right\}.$$

5 Concluding remarks

In this paper we have demonstrated a simple transformation of the empirical process under the hypothesis of exponentiality. Our transformed empirical process allows the construction of a whole class of goodness-of-fit test statistics, which have the desirable poperties of power against any local alternative and asymptotic distribution-freeness. An especially appealing feature of our particular transformed empirical process is the convenience of the limit distributions of any constructed test statistics, which are all simply functionals from the standard Brownian motion.

We have also explored the asymptotically linear functional form of several well known test statistics and shown that they are essentially equivalent. Some of these asymptotically linear functionals have been recommended by others as 'omnibus' goodness-of-fit statistics. As we have shown, in fact such linear statistics can have almost no power against many local alternatives. In contrast, true omnibus test statistics constructed from our transformed empirical process demonstrate superior power, which further amplifies the value of that approach when testing the hypothesis of exponentiality against general alternatives.

Acknowledgements

We thank Ray Brownrigg for some assistance with statistical computing and Boris Andreyanov for useful discussions.

References

Ascher, S. (1990) A surevey of tests for exponentiality. *Communications in Statistics – Theory and Methods*, **19(5)**, 1811–1825.

Angus, J. E. (1982) Goodness-of-fit tests for exponentiality based on a loss-of-memory type functional equation. *Journal of Statistical Planning and Inference*, **6(3)**, 241–251.

Bandyopadhyay, D. and Basu, A. P. (1989) A note on tests for exponentiality by Deshpande. *Biometrika*, **76**, 403–405.

Baringhaus, L. and Henze, N. (1991) A class of consistent tests for exponentiality based on the empirical Laplace transform. *Annals of the Institute of Statistical Mathematics*, **43**, 179–192.

Baringhaus, L. and Henze, N. (2000) Tests of fit for exponentiality based on a characterization via the mean residual life function. *Statistical Papers*, **41**, 225–236.

Borodin, A. and Salminen, P. (2002) *Handbook of Brownian Motion: Facts and Formulae (Probability and its Applications)*, 2nd edn. Basel: Birkhauser Verlag AG.

Cox, D. R. and Oakes, D. (1984) *Analysis of Survival Data.* London: Chapman and Hall.

Deheuvels, P. and Martynov, G. (2003) Karhunen-Loève expansions for weighted Wiener processes and Brownian bridges via Bessel functions. *Progress in Probability*, **55**, 57–93.

Deshpande, J. V. (1983) A class of tests for exponentiality against increasing failure rate average alternatives. *Biometrika*, **70**, 514–518.

Epps, T. W. and Pulley, L. B. (1986) A test for exponentiality vs. monotone hazard alternatives derived from the empirical characteristic function. *Journal of the Royal Statistical Society* B, **48**, 206–213.

Feller, W. (1971) *An Introduction to Probability Theory and its Applications, Volume II*, 2nd edn. New York: Wiley.

Gail, M. H. and Gastwirth, J. L. (1978) A scale-free goodness-of-fit test for exponentiality based on the Gini statistic. *Journal of the Royal Statistical Society* B, **40**, 350–357.

Henze, N. (1993) A new flexible class of omnibus tests for exponentiality. *Communications in Statistics – Theory and Methods*, **22**, 115–133.

Henze, N. and Meintanis, S. G. (2005) Recent and classical tests for exponentiality: a partial review with comparisons. *Metrika*, **61(1)**, 29–45.

Khmaladze, E. V. (1979) The use of Omega-square tests for testing parametric hypotheses. *Theory of Probability and its Applications*, **24(2)**, 283–302.

Khmaladze, E. V. (1981) Martingale approach in the theory of goodness of fit tests. *Theory of Probability and its Applications*, **26(1)**, 240–257.

Khmaladze, E. V., Brownrigg, R. D. and Haywood, J. (2005) On Roman Emperors, their Chinese counterparts, and statistics of their reign. *Mimeo.* Victoria University of Wellington.

Khmaladze, E. V. and Koul, H. L. (2004) Martingale transforms goodness-of-fit tests in regression models. *The Annals of Statistics*, **32(3)**, 995–1034.

Khmaladze, E. V. and Shinjikashvili, E. (2001) Calculation of noncrossing probabilities for Poisson processes and its corollaries. *Advances in Applied Probability*, **33**, 702–716.

Koul, H. L. (1977) A test for new better than used. *Communications in Statistics: Theory and Methods*, **A6**, 563–574.

Koul, H. L. (1978) A class of tests for testing "new is better than used". *The Canadian Journal of Statistics*, **6(2)**, 249–271.

Janssen, A. (2000) Global power functions of goodness of fit tests. *The Annals of Statistics*, **28(1)**, 239–253.

Martynov, G. (1977) *Omega Square (Cramér-von Mises) Test* (in Russian). Moscow: Nauka.

Moran, P. A. P. (1951) The random division of an interval—part II. *Journal of the Royal Statistical Society* B, **13**, 147–150.

Nikitin, Ya. (1996) Bahadur efficiency of a test of exponentiality based on a loss-of-memory type functional equation. *Journal of Nonparametric Statistics*, **6**, 13–26.

Orlov, A. I. (1972) On testing the symmetry of distributions. *Theory of Probability and its Applications*, **17**, 357–361.

Shiryaev, A. N. (1999) *Essentials of Stochastic Finance. Facts, Models, Theory.* Singapore: World Scientific.

Stephens, M. A. (1986) Tests for the exponential distribution. In R. B. D'Agostino and M. A. Stephens (Eds.), *Goodness-of-Fit Techniques*, 421–459. New York: Marcel Dekker.

Tchirina, A. V. (2002) Bahadur efficiency and local optimality of a test for exponentiality based on the Moran statistics. *Zap. Nauchn. Semin. POMI*, **291**, 245–259.

Test statistic	Nominal size	Size of sample	Empirical size (%)
d_n^+	10%	50	9.260
		100	9.700
		200	10.132
	5%	50	5.852
		100	5.850
		200	5.794
d_n^-	10%	50	9.204
		100	9.248
		200	9.268
	5%	50	3.956
		100	4.036
		200	4.268
d_n	10%	50	9.810
		100	9.892
		200	10.062
	5%	50	5.562
		100	5.498
		200	5.504
ω_n^2	10%	50	9.770
		100	9.918
		200	9.962
	5%	50	4.954
		100	5.010
		200	5.060
A_n^2	10%	50	9.888
		100	10.064
		200	10.046
	5%	50	4.984
		100	5.128
		200	5.152

Table 1: Empirical sizes for the test statistics considered, for various sample sizes, with exponential parameter known. 50,000 replications for each sample size, with all statistics calculated for each replicate.

Test statistic	Nominal size	Size of sample	Empirical size (%)
d_n^+	10%	50	7.428
		100	8.204
		200	9.220
	5%	50	4.468
		100	4.712
		200	5.242
d_n^-	10%	50	10.572
		100	10.420
		200	10.090
	5%	50	4.430
		100	4.368
		200	4.462
d_n	10%	50	8.898
		100	9.086
		200	9.704
	5%	50	4.540
		100	4.790
		200	5.094
ω_n^2	10%	50	9.618
		100	9.502
		200	9.710
	5%	50	4.604
		100	4.662
		200	4.870
A_n^2	10%	50	9.772
		100	9.652
		200	9.778
	5%	50	4.786
		100	4.750
		200	4.854

Table 2: Empirical sizes for the test statistics considered, for various sample sizes, with exponential parameter estimated. 50,000 replications for each sample size, with all statistics calculated for each replicate.