Title
Testing exponentiality by comparing the empirical distribution function of the normalized spacings with that of the original data

Journal
Journal of Nonparametric Statistics, 15(6)

ISSN
1048-5252

Authors
Jammalamadaka, S Rao
Taufer, Emanuele

Publication Date
2003-12-01

DOI
10.1080/10485250310001638102

Peer reviewed

Nonparametric Statistics
Vol. 15(6), December 2003, pp. 719–729

TESTING EXPONENTIALITY BY COMPARING THE EMPIRICAL DISTRIBUTION FUNCTION OF THE NORMALIZED SPACINGS WITH THAT OF THE ORIGINAL DATA

S. RAO JAMMALAMADAKA[a] and EMANUELE TAUFER[b,*]

[a] *Department of Statistics and Applied Probability, University of California, Santa Barbara, CA 93106-3110, USA;* [b] *Department of Computer and Management Sciences, University of Trento, Via Inama 5, 38100 Trento, Italy*

(Received January 2002; Revised April 2003; In final form October 2003)

We introduce new goodness of fit tests for exponentiality by using a characterization based on normalized spacings. We provide relevant asymptotic theory for these tests and study their efficiency. An empirical power study and comparisons are also provided.

Keywords: Test for exponentiality; Kolmogorov–Smirnov statistics; Omega square statistics; Brownian Bridge; Asymptotic efficiency

1 INTRODUCTION

In this paper, we develop a goodness of fit test for exponentiality exploiting a characterization based on the 'normalized spacings'. There is considerable literature on the problem of testing for exponentiality. The reasons are many-fold and chief among these are: the watershed role played by the exponential distribution in reliability and survival analysis, its nice mathematical properties, as well as the availability of several characterizations. Most recently, Baringhaus and Henze (2000) and Taufer (2000) developed tests based on the mean residual life characterization. Alwasel (2001) and Ahmad and Alwasel (1999) exploited a characterization based on the lack of memory of the exponential distribution. This property is also the basis for tests considered earlier by Angus (1982) who used Kolmogorov–Smirnov and Cramer–von Mises type test statistics based on the difference of the empirical distribution functions (e.d.f.) $\bar{F}_n(2x)$ and $\bar{F}_n^2(x)$, where $\bar{F} = 1 - F$, denotes the survival function. Grzegorzewski and Wieczorkowski (1999) and Ebrahimi *et al.* (1992) make use of the maximum entropy characterization by considering the difference between a nonparametric estimator of entropy and the maximum likelihood estimator of entropy under exponentiality. Other omnibus tests for exponentiality have been developed by Henze (1993) and Baringhaus and Henze (1991, 1992) who used weighted distance measures between sample estimators of the Laplace

* Corresponding author. E-mail: emanuele.taufer@untin.it

ISSN 1048-5252 print; ISSN 1029-0311 online
DOI: 10.1080/10485250310001638102

transform and its counterpart under the null hypothesis. Klar (2000) provides a test against harmonic new better than used in expectation (HNBUE) alternatives by using estimates of $\int_x^\infty \bar{F}(t)\,\mathrm{d}t - \theta \mathrm{e}^{-x/\theta}$ based on the e.d.f. and the sample mean. The test exploits the property that $\int_x^\infty \bar{F}(t)\,\mathrm{d}t - \theta \mathrm{e}^{-x/\theta}$ is positive for all x for HNBUE alternatives and uses a weighted distance measure. The approach of Klar generalizes an earlier contribution of Jammalamadaka and Lee (1998) who consider a non-weighted distance measure. For a review of earlier contributions, the interested reader is referred to Ascher (1990) and Doksum and Yandell (1984).

2 TEST STATISTICS

To turn to our problem, consider a random sample $X_1, \ldots, X_n$ from an exponential distribution with density $f(x) = (1/\theta)\exp\{-x/\theta\}$, $x > 0$, $\theta > 0$ (denote this by $E(\theta)$); let $X_{(0)} = 0$ and let $X_{(1)}, \ldots, X_{(n)}$ denote the order statistics. It is well known that the so-called 'normalized spacings'

$$Y_i = (n - i + 1)(X_{(i)} - X_{(i-1)}) \quad i = 1, \ldots, n$$

are again independently and identically distributed (i.i.d.) from $E(\theta)$. Further this property characterizes the exponential distribution, as shown in Seshadri *et al.* (1969). There are several inferential procedures based on the normalized spacings. However, our goal in this paper is to use simultaneously the X and the Y variables in order to provide a goodness of fit test for exponentiality.

One of the simplest and natural ways to do this is to compare the e.d.f. of the original variables X with that of the transformed ones, Y. More precisely, let $F_n(t)$ and $G_n(t)$ denote the e.d.f. of $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$ respectively. We can construct new tests of exponentiality by measuring the distance between these 2 e.d.f.s, using the classical Kolmogorov–Smirnov and Cramer–von Mises type distances. We then obtain the test statistics

$$T_{1,n} = \sqrt{\frac{n}{2}} \sup_{0 \le t < \infty} |F_n(t) - G_n(t)|$$

and

$$T_{2,n} = \frac{n}{2\bar{X}} \int (F_n(t) - G_n(t))^2 \mathrm{e}^{-t/\bar{X}}\,\mathrm{d}t,$$

where $\bar{X}$ is the sample mean. Clearly, under the hypothesis of exponentiality, since $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$ have identical distributions, their e.d.f. $F_n(t)$ and $G_n(t)$ should be close. Thus, one expects $T_{1,n}$ and $T_{2,n}$ to be close to zero under the null hypothesis of exponentiality, while they should be large under any alternative hypothesis.

Although the $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ are not independent, our statistics $T_{1,n}$ and $T_{2,n}$ resemble the corresponding two-sample versions and hence the computational formulae for $T_{1,n}$ and $T_{2,n}$ are easily derived. For instance, one computes $T_{1,n}$ just as it is done for the two sample Kolmogorov–Smirnov statistic, using $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ as the two 'samples'. In order to compute $T_{2,n}$, let $Z_{(1)}, \ldots, Z_{(2n)}$ denote the ordered values obtained by combining $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$. Then these $Z_{(i)}$s correspond to the jump points of $F_n(t) - G_n(t)$ and therefore

$$T_{2,n} = \frac{n}{2} \sum_{i=1}^{2n-1} [F_n(Z_{(i)}) - G_n(Z_{(i)})]^2 \left[\exp\left\{\frac{-Z_{(i)}}{\bar{X}}\right\} - \exp\left\{\frac{-Z_{(i+1)}}{\bar{X}}\right\}\right].$$

As shown in Section 3, $T_{1,n}$ and $T_{2,n}$ provide 'consistent' tests for testing exponentiality, and moreover their null distribution is free of nuisance parameters. These facts, together with their computational simplicity makes them valuable as well as practically useful tests for exponentiality.

3 ASYMPTOTIC PROPERTIES

Before we proceed to develop the relevant asymptotic theory for $T_{1,n}$ and $T_{2,n}$, we should note that the two empirical distribution functions are not independent. The dependence between the Xs and Ys can be more easily seen by representing both of them in terms of the spacings $D_i = X_{(i)} - X_{(i-1)}$. We can write

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \le t\}} = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_{(i)} \le t\}} = \frac{1}{n}\sum_{i=1}^{n} I_{\left\{\Sigma_{j=1}^{i} D_j \le t\right\}}$$

and

$$G_n(t) = \frac{1}{n}\sum_{i=1}^{n} I_{\{Y_i \le t\}} = \frac{1}{n}\sum_{i=1}^{n} I_{\{(n-i+1)(X_{(i)}-X_{(i-1)}) \le t\}} = \frac{1}{n}\sum_{i=1}^{n} I_{\{(n-i+1)D_i \le t\}}.$$

We first derive some general results on $T_{1,n}$ and $T_{2,n}$.

3.1 Consistency

For $0 \le s \le 1$, we define the two processes

$$\alpha_n(s) \equiv \sqrt{n}(F_n(F_E^{-1}(s)) - s) \quad \text{and} \quad \beta_n(s) \equiv \sqrt{n}(G_n(F_E^{-1}(s)) - s),$$

where F_E indicates the distribution function of an $E(\theta)$ random variable. Then

$$\sqrt{n}(F_n(t) - G_n(t)) = [\alpha_n(F_E(t)) - \beta_n(F_E(t))]$$

from the continuity of F_E it follows that $T_{1,n}$ and $T_{2,n}$ are distribution-free under the null hypothesis of exponentiality. From now on, without loss of generality, we will assume that the random sample $X_1, \ldots, X_n$ comes from an $E(1)$ distribution.

Next, we note that by the Glivenko-Cantelli theorem, under the null hypothesis we have that

$$\sup_{0 \le t < \infty} |F_n(t) - F_E(t)| \xrightarrow{\text{a.s.}} 0 \quad \text{and} \quad \sup_{0 \le t < \infty} |G_n(t) - F_E(t)| \xrightarrow{\text{a.s.}} 0,$$

from which we easily infer that $T_{1,n}/\sqrt{n} \xrightarrow{\text{a.s.}} 0$ and $T_{2,n}/n \xrightarrow{\text{a.s.}} 0$ under exponentiality. On the other hand, under the alternative that the observations are distributed according to some continuous distribution function (d.f.) F_A, following the results in Pyke (1965), we have that

$$F_n(t) - G_n(t) \xrightarrow{\text{a.s.}} F_A(t) - 1 + \int_0^\infty f_A(y) \exp\{-t h_A(y)\}\, \mathrm{d}y \tag{1}$$

where $f_A(t)$ denotes the density and

$$h_A(t) = \frac{f_A(t)}{1 - F_A(t)},$$

the hazard rate corresponding to F_A. It may be recalled that under the null hypothesis of exponentiality this hazard rate, $h_E(t) = 1$. This indicates that the two tests are consistent and their efficiency is tied to the hazard rate of the distribution, a point which will be studied in more detail in Section 4.

3.2 Asymptotic Null Distribution

Now we consider the problem of determining the asymptotic null distribution of $T_{1,n}$ and $T_{2,n}$. It is well known, see for example Csörgő and Horváth (1993, p. 114), that there exist sequences of Brownian Bridges $B_n^1(t)$ and $B_n^2(t)$ such that

$$\sup_{0\le t\le 1} |\alpha_n(t) - B_n^1(t)| \overset{\text{a.s.}}{=} \mathrm{O}(n^{-1/2}\log n)$$

and

$$\sup_{0\le t\le 1} |\beta_n(t) - B_n^2(t)| \overset{\text{a.s.}}{=} \mathrm{O}(n^{-1/2}\log n).$$

However, in our case, $\alpha_n(s)$ and $\beta_n(s)$ are not independent. Therefore, in order to determine the asymptotic behavior of $T_{1,n}$ and $T_{2,n}$, we need to find a joint approximation for the processes $\alpha_n(s)$ and $\beta_n(s)$. Such joint behavior is discussed in Barbe (1994). It turns out that the two processes are asymptotically independent, which enables us to prove consistency of our test statistics and to study their asymptotic efficiency.

For $0 \le s, u \le 1$, let

$$G_K^{\alpha}(s) = \int_0^1 [K(u,s) - sK(u,1)]\,\mathrm{d}F^{-1}(u)$$
$$G_K^{\beta}(s) = K(s,1)$$

where $K(\cdot,\cdot)$ denotes a Kiefer process. For details on this process the reader is referred to Csörgö and Révész (1981). Note that under exponentiality $F^{-1}(u) = -\log(1-u)$ and $G_K^{\alpha}(s)$ and $G_K^{\beta}(s)$ have the same distribution as a Brownian Bridge $B(s)$; moreover the joint process $(G_K^{\alpha}(s), G_K^{\beta}(s))$ is Gaussian (Barbe, 1994). Also, since for any choice of $0 \le s, u \le 1$,

$$\mathrm{cov}(G_K^{\alpha}(s), G_K^{\beta}(u)) = 0,$$

we see that the two processes are independent. We provide the joint approximation in the following theorem.

THEOREM 1 *Under exponentiality one can construct the sequences $\alpha_n(s)$ and $\beta_n(s)$ and a sequence of Kiefer processes K_n $(\cdot,\cdot)$ on the same probability space, such that*

$$\sup_{0\le s\le 1} |\alpha_n(s) - G_{K_n}^{\alpha}(s)| = \mathrm{O}_p(n^{-1/2}\phi_n\log^2 n)$$
$$\sup_{0\le s\le 1} |\beta_n(s) - G_{K_n}^{\beta}(s)| = \mathrm{O}_p(n^{-1/2}\log^2 n).$$

where

$$\phi_n = \left|F^{-1}\left(\frac{1}{n}\right)\right| \vee \left|F^{-1}\left(1-\frac{1}{n}\right)\right|$$

Proof The proof is adapted from that of Theorem 2.1 in Barbe (1994) which in turn relies on a result from Komlós *et al.* (1975). Consider first the process $\alpha_n(s)$, we just need to note that it is equivalent to the process $\alpha_n^U(s)$ defined in Barbe (1994) and hence we can directly apply the approximation provided therein with the function $\phi_n = \log n$ and $F^{-1}(t) = -\log(1-t)$ in the case of exponentiality.

As far as the process $\beta_n(s)$ is concerned, we may apply directly the following result of Komlós *et al.* (1975) which provides an approximation of the uniform empirical process by a Kiefer process such that

$$\max_{1\le k\le n} \sup_{0\le s\le 1} |k(F_k(s)-s) - K(s,k)| \overset{\text{a.s.}}{=} \mathrm{O}(\log^2 n)$$

where $F_k(s) = k^{-1}\sum_{i=1}^{k} I_{\{Y_i\le t\}}$. Then, it holds that

$$\beta_n(s) = \frac{K(s,n)}{n^{1/2}} + \mathrm{O}_p(n^{-1/2}\log^2 n).$$

Reasoning again as in Barbe (1994), we obtain the approximation given in the theorem.

As a consequence of the theorem, we see that the processes $\alpha_n(s)$ and $\beta_n(s)$ are asymptotically independent. From this result, consistency and the asymptotic null distribution of $T_{1,n}$ and $T_{2,n}$ follow at once. We state the following Corollary.

COROLLARY *Under exponentiality the process* $\sqrt{(1/2)}[\alpha_n(s) - \beta_n(s)]$ *converges weakly to a Brownian Bridge* $B(s)$, $0\le s\le 1$.

Applying Donsker's theorem we obtain the asymptotic distribution of $T_{1,n}$ and $T_{2,n}$.

THEOREM 2 *Under the null hypothesis of exponentiality,*

$$T_{1,n} \xrightarrow{\mathcal{D}} \sup_{0<s<1} |B(s)|,$$

$$T_{2,n} \xrightarrow{\mathcal{D}} \int_0^1 |B(s)|^2\,\mathrm{d}s,$$

where $B(s), 0\le s\le 1$ *denotes a Brownian Bridge. Thus*

$$\lim_{n\to\infty} P(T_{1,n} > t) = 2\sum_{k=1}^{\infty}(-1)^{k+1}\exp\{-2k^2t^2\}$$

and

$$\lim_{n\to\infty} P(T_{2,n} > t) = \frac{1}{\pi}\sum_{k=1}^{\infty}(-1)^{k+1}\int_{(2k-1)^2\pi^2}^{(2k)^2\pi^2}\frac{1}{y}\sqrt{\frac{-\sqrt{y}}{\sin\sqrt{y}}}\exp\left(-\frac{ty}{2}\right)\mathrm{d}y.$$

4 APPROXIMATE BAHADUR EFFICIENCY

Result (1) tells us that the performance of our test statistics is closely connected with the hazard rate of the distribution and this is even more evident if we write the distribution function F in terms of the hazard rate, *i.e.*

$$F(t) = 1 - \exp\left\{-\int_0^t h(y)\,\mathrm{d}y\right\}.$$

In order to investigate in more depth, we will find the dominant term in Eq. (1) for a sequence of alternatives that approach the null hypothesis. To do so, suppose that under the alternative, the observations are distributed according to some continuous d.f. $F(t,\theta)$, $\theta \geq 0$ which coincides with the exponential d.f. only for $\theta = 0$ and denote, for convenience, the RHS of Eq. (1) as $H(t,\theta)$.

We expand formally $H(t,\theta)$ in Taylor series for θ around $\theta = 0$ retaining only the first few terms. To this end, we need some regularity conditions on $F(t,\theta)$ and $f(t,\theta)$ and these are implicit as in *e.g.* Nikitin (1996). Let the prime denote the derivative with respect to θ, *i.e.* for any function $g(\cdot,\theta)$

$$g'(\cdot,\theta) = \frac{\partial g(\cdot,\theta)}{\partial\theta}.$$

If we compute the derivative of the function $H(t,\theta)$ with respect to θ we obtain

$$\begin{aligned} H'(t,\theta) = F'(t,\theta) &+ \int_0^\infty f'(y,\theta)\exp\{-th(y,\theta)\}\,\mathrm{d}y \\ &- t\int_0^\infty f(y,\theta)h'(y;\theta)\exp\{-th(y,\theta)\}\,\mathrm{d}y. \end{aligned}$$

To evaluate $H'(t,0)$ we recall that $h(t,0) = h_E(t) = 1$ and note that, by definition and regularity conditions

$$h'(t,0) = \mathrm{e}^t[f'(t,0) + F'(t,0)], \qquad \int_0^\infty f'(t,0)\,\mathrm{d}t = 0.$$

After some simplifications we obtain

$$H'(t,0) = F'(t,0) - t\mathrm{e}^{-t}\int_0^\infty F'(y,0)\,\mathrm{d}y.$$

Together with the fact that $H(t,0) = 0$ we finally have that

$$H(t,\theta) = \left[F'(t,0) - t\mathrm{e}^{-t}\int_0^\infty F'(y,0)\,\mathrm{d}y\right]\theta + \mathrm{O}(\theta^2). \tag{2}$$

This result may be used to compute the approximate local Bahadur slope in order to make some comparisons with other test statistics efficiencies. Recall that if a sequence $\{T_n\}$ satisfies $T_n/\sqrt{n} \xrightarrow{P_\theta} b(\theta), \theta > 0$ and, under the null hypothesis, the limiting distribution $F(t) = \lim_{n\to\infty} P(T_n \leq t)$ satisfies $\log[1 - F(t)] = -((at^2)/2)(1 + \mathrm{o}(1))$, as $t \to \infty$, then the approximate Bahadur slope of the standard sequence T_n is defined as $c_T(\theta) = a[b(\theta)]^2$.

Given the asymptotic independence of the processes based on the observations and the normalized spacings and result in Eq. (1) it is possible to compute the approximate Bahadur slope for the two statistics $T_{1,n}$ and $T_{2,n}$ (we take the root of $T_{2,n}$ in order to have a normalized sequence). The value of the constant a may be determined from the corresponding results for two sample Kolmogorov–Smirnov and Cramer–von Mises statistics; they can be found, for example, in Nikitin (1995, chapter 3). Let $c_{T_1}(\theta)$ and $c_{T_2}(\theta)$ denote the approximate Bahadur slopes of $T_{1,n}$ and $T_{2,n}$, respectively. Then

$$c_{T_1}(\theta) = [\sup_{0\leq t<\infty} |H(t,\theta)|]^2$$

and

$$c_{T_2}(\theta) = \frac{\pi^2}{4}\left[\int_0^\infty H(t,\theta)^2 \mathrm{e}^{-t}\,\mathrm{d}t\right].$$

Next, we can use result (2) to compute the approximate slope as $\theta \to 0$. In order to have a first battery of results, we now compare the approximate Bahadur slopes of the statistics $T_{1,n}$ and $T_{2,n}$ with those of Kolmogorov–Smirnov (KS) and Cramer–von Mises (ω^2) one sample statistics of the simple hypothesis of exponentiality where the mean is given; next we will consider a test (G) based on the Gini's statistic (Gail and Gastwirth, 1978) and a test (A_1) based on a loss of memory type functional equation which has been proposed by Angus (1982).

It should be remarked that comparisons are made in order to provide some complementary information and exemplify computations for our tests statistics for some common distributions. Moreover, note that in the case of KS and ω^2 the slopes are computed on the basis of a simple hypothesis, *i.e.* the mean θ is given; in the case of Gini's and Angus' statistics the exact slopes are available as they have been computed by Nikitin and Tchirina (1996) and Nikitin (1996), respectively.

We utilize the linear failure rate, Makeham and Weibull alternatives, which are often considered in the evaluation of the performance of tests for exponentiality (Doksum and Yandell, 1984; Nikitin, 1996).

Example 1 Consider a distribution with density $f(t,\theta) = (1+\theta t)\exp\{-t-(1/2)\theta t^2\}$ (linear increasing failure rate). In this case we have

$$F'(t,\theta) = \frac{1}{2}t^2 \exp\left\{-\left[t + \frac{\theta t^2}{2}\right]\right\}$$

and hence

$$F'(t,0) - t\mathrm{e}^{-t}\int_0^\infty F'(y,0)\,\mathrm{d}y = \frac{1}{2}t\mathrm{e}^{-t}(t-2).$$

From this we obtain that $c_{T_1}(\theta) = 0.0532\theta^2$ and $c_{T_2}(\theta) = 0.061\theta^2$ as $\theta \to 0$. To obtain the corresponding results for one sample KS and ω^2 statistics we expand $F(t,\theta)$ around $\theta = 0$ to obtain

$$F(t,\theta) - F(t,0) = F'(t,0)\theta + \mathrm{O}(\theta^2).$$

Applying standard techniques for computing the approximate slope of these statistics (see, *e.g.* Nikitin, 1995, chapter 2) we obtain $c_{\mathrm{KS}}(\theta) = 0.2930\theta^2$ and $c_{\omega^2}(\theta) = 0.2437\theta^2$ as $\theta \to 0$. From this we see that the relative efficiencies are $e^B_{T_1,\mathrm{KS}} = c_{T_1}(\theta)/c_{\mathrm{KS}}(\theta) \simeq 0.18$ and $e^B_{T_2,\omega^2} = c_{T_2}(\theta)/c_{\omega^2}(\theta) \simeq 0.25$. Next, using the results in Nikitin and Tchirina (1996) and Nikitin (1996) we obtain $c_G(\theta) = 0.75\theta^2$ and $c_{A_1}(\theta) = 0.0733\theta^2$ as $\theta \to 0$. Considering $T_{1,n}$, we obtain $e^B_{T_1,G} \simeq 0.07$ and $e^B_{T_1,A_1} \simeq 0.73$; slightly better results hold for $T_{2,n}$.

Example 2 Take a Makeham density $f(t,\theta) = [1+\theta(1-\mathrm{e}^{-t})]\exp\{-[t+\theta\,(t+\mathrm{e}^{-t}-1)]\}$. Then

$$F'(t,\theta) = (t+\mathrm{e}^{-t}-1)\exp\{-[t+\theta(t+\mathrm{e}^{-t}-1)]\}$$

from which

$$F'(t,0) - t\mathrm{e}^{-t}\int_0^\infty F'(y,0)\,\mathrm{d}y = \mathrm{e}^{-t}\left[\mathrm{e}^{-t} + \frac{t}{2} - 1\right],$$

and we have $c_{T_1}(\theta) = 0.0077\theta^2$ and $c_{T_2}(\theta) = 0.008\theta^2$ as $\theta \to 0$. Also, $c_{\mathrm{KS}}(\theta) = 0.1048\theta^2$ and $c_{\omega^2}(\theta) = 0.1\theta^2$ as $\theta \to 0$, and hence $e^B_{T_1,\mathrm{KS}} \simeq 0.07$ and $e^B_{T_2,\omega^2} \simeq 0.08$. Again, we obtain $c_G(\theta) = 0.0833\theta^2$ and $c_{A_1}(\theta) = 0.0156\theta^2$ as $\theta \to 0$ and consequently $e^B_{T_1,G} \simeq 0.09$ and $e^B_{T_1,A_1} \simeq 0.49$ with analogue results for $T_{2,n}$.

Example 3 Consider a Weibull density of the form $f(t,\theta) = (\theta+1)t^\theta \exp\{-t^{1+\theta}\}$. Calculations lead to

$$F'(t,\theta) = t^{1+\theta}\exp\{-t^{1+\theta}\}\log t$$

from which

$$F'(t,0) - t\mathrm{e}^{-t}\int_0^\infty F'(y,0)\,\mathrm{d}y = t\mathrm{e}^{-t}[\log t + \gamma - 1]$$

where γ is Euler's constant. It turns out that $c_{T_1}(\theta) = 0.1321\theta^2$ and $c_{T_2}(\theta) = 0.1377\theta^2$ as $\theta \to 0$. Also, $c_{\mathrm{KS}}(\theta) = 0.2916\theta^2$ and $c_{\omega^2}(\theta) = 0.3113\theta^2$ as $\theta \to 0$, and hence $e^B_{T_1,\mathrm{KS}} \simeq 0.45$ and $e^B_{T_2,\omega^2} \simeq 0.44$. Finally we have $c_G(\theta) = 1.44\theta^2$ and $c_{A_1}(\theta) = 0.2601\theta^2$ as $\theta \to 0$ and consequently $e^B_{T_1,G} \simeq 0.09$ and $e^B_{T_1,A_1} \simeq 0.51$ with analogue results for $T_{2,n}$.

From the three examples, we note that the approximate slopes of $T_{1,n}$ and $T_{2,n}$ are always close to each other, but fall short of the comparisons with the other test statistics. Part of the explanation, for classical Kolmogorov–Smirnov and ω^2 one sample statistics, may be found in the fact that these are computed under a simple hypothesis of exponential with given mean, whereas for our statistics the slope has been calculated under the composite hypothesis where the mean is not specified. Note, in fact, that there is no need to estimate the mean in computing $T_{1,n}$; a version of the quadratic statistics $T_{2,n}$ with analogue characteristics could be developed also.

Further, approximate Bahadur slopes are only a rough measure of the asymptotic efficiency and this might partly explain results as far as Gini's statistic and $A_{1,n}$ are concerned. Coming to the question of computing the exact slopes, given this partial evidence and the impossibility of applying standard results it seems an uneven task. In Section 5, we examine some simulated power values of our statistics where we will see, power for moderate sample sizes gives satisfactory results.

5 MONTE CARLO POWER COMPARISONS

In this section, we are going to compare the power performance of $T_{1,n}$ and $T_{2,n}$ with those of other tests statistics that have appeared in the literature. In particular, we consider traditional one sample Kolmogorov–Smirnov and ω^2 statistics with estimated parameters and other tests based on a characterization of the exponential distribution that have been proposed in the literature such as those proposed by Angus (1982) and based on a loss of memory type functional equation (actually on an order statistics characterization) but also, on this subject one can consult Ahmad and Alwasel (1999) which propose a different test statistics based on the same functional equation. Next, we consider the proposal of Baringhaus and Henze (2000) which define Kolmogorov–Smirnov and Cramer–von Mises type statistics relying on a characterization via mean residual life. The last two alternative statistics considered are those of Ebrahimi *et al.* (1992) (see also Grzegorzewski and Wieczorkowski, 1999) whose test is based on entropy and Gail and Gastwirth (1978) who propose a test for exponentiality based on the Gini's index which is constructed from the area under the Lorenz curve; this test, also

TABLE I Goodness of Fit Tests Studied.

Symbol	*Definition/characterization*	*Reference*
KS_n	Two sided Kolmogorov–Smirnov test with estimated mean	Durbin (1975)
ω_n^2	Cramer–von Mises statistic with estimated mean	Darling (1957)
$A_{1,n}$	KS type based on loss of memory functional equation	Angus (1982)
$A_{2,n}$	CVM type based on loss of memory functional equation	Angus (1982)
$H_{1,n}$	KS type based on mean residual life function	Baringhaus and Henze (2000)
$H_{2,n}$	Cramer–von Mises type based on mean residual life	Baringhaus and Henze (2000)
$V_{m,n}$	Based on Vasicek's (1976) entropy estimator	Ebrahimi *et al.* (1992)
G_n	Based on Gini's index	Gail and Gastwirth (1978)

in the light of the recent results of Nikitin and Tchirina (1996) is a test which performs well in a variety of situations. Table I summarizes the relevant information about the alternative test statistics used.

As far as the alternative distributions considered in the Monte Carlo experiment, they are reported in Table II. We have tried to encompass some common distributions frequently considered in other power studies as well as to have a certain variety of life distributions differing from the point of view of the hazard rate. The choice of the parameters has been made in such a way that the resulting density is not too far away in shape from the exponential one.

In Table III, we find the power estimates for samples of moderate size ($n = 20$) for the goodness fit tests considered in Table I and the alternative distributions of Table II. The estimates have been obtained by rounding to the nearest integer the percentage of statistics declared significant out of 10,000 samples of size 20. In order to check our routines, we have reproduced the power studies contained in the original papers of Table I, obtaining substantially equal results. The test statistics of Ebrahimi *et al.* (1992) required the choice of a window parameter m. Also in the light of Taufer (2002), we chose a value $m = 4$ for all situations.

This power study provides some elements to make an overall comparison among tests for exponentiality based on characterizations.

Analogously to what happened in the computation of the approximate Bahadur slope, $T_{1,n}$ and $T_{2,n}$ show very close power values however, as we can see, they compare well in most cases also with the tests statistics considered in Section 4. The tests do not perform as well as most of its competitors for Weibull alternatives with $\theta < 1$ (and also Gamma). Actually, the same happens to the test statistics $A_{1,n}$ and $A_{2,n}$ based on the loss of memory characterization. Apart from this case, the test statistics $T_{1,n}$ and $T_{2,n}$ compare well with most of the other tests although they do not uniformly outperform any of the other tests for the cases considered. Although there is clearly not a best test from the point of view of power for all situations, we see that the statistics based on mean residual life and entropy show good power for most alternatives.

TABLE II Alternative Distributions Considered.

Symbol	*F(t) [f(t) when indicated]*	*Support*	*Failure rate*
Pareto	$1 - [(\theta-2)/(\theta-1)]^{\theta-1}t^{1-\theta}$	$t \geq (\theta-2)/(\theta-1)$	Decreasing
Weibull	$1 - \exp\{-t^{\theta}\}$	$t \geq 0$	Increasing $\theta > 1$, decreasing $\theta < 1$
Lognormal	$f(t) = (1/(\sqrt{2\pi}t\theta))\exp\{\log^2(t)/(2\theta^2)\}$	$t \geq 0$	Hump shaped
Shifted exponential	$1 - \exp\{-(t-\theta)\}$	$t \geq \theta$	Constant
Linear failure rate	$1 - \exp\{-t - \theta t^2/2\}$	$t \geq 0$	Increasing
Dhillon	$1 - \exp\{1 - e^{t^{\theta}}\}$	$t \geq 0$	Bathtub shaped, $\theta < 1$

TABLE III Monte Carlo Power Estimates Based on 10,000 Samples of Size 20.

Distribution	$T_{1,n}$	KS_n	$A_{1,n}$	$H_{1,n}$	$V_{4,n}$	G_n	$T_{2,n}$	ω_n^2	$A_{2,n}$	$H_{2,n}$
Pareto(2.2)	87	93	99	83	99	44	67	91	97	74
Pareto(2.5)	97	98	99	95	99	53	91	98	99	86
Weibull(0.8)	04	17	02	13	04	24	04	20	04	21
Weibull(1.4)	27	29	25	36	37	36	26	35	29	37
Lognormal(0.6)	73	84	81	84	90	80	74	89	84	84
Lognormal(0.8)	28	30	42	29	42	24	26	34	38	27
Shifted exp(0.2)	28	25	42	23	59	22	25	28	34	23
LFR(0.5)	12	10	09	14	13	10	12	11	11	11
LFR(3.0)	30	30	23	39	36	38	30	37	29	38
Dhillon(0.7)	06	06	03	06	05	05	06	06	05	05
Dhillon(0.9)	17	15	10	21	20	17	17	17	14	18

Note: Significance level $\alpha = 0.05$.

6 EXAMPLES

As examples of application we use the data set given in Grubbs (1971) and Wadsworth (1990, p. 611). These data sets have been used in Ebrahimi *et al.* (1992), Ahmad and Alwasel (1999) and Shapiro (1995) to show the application of various tests for exponentiality. The first data (Tab. IV) are the times between arrivals of 25 customers at a facility, their quantile plot shows a clear departure from the exponentiality hypothesis, the hypothesis of exponentiality is rejected with p-value smaller than 0.01 by all the tests considered in Shapiro (1995). The second data set (Tab. V) represents mileages for 19 military personal carriers that failed in service. These data have been used by Ebrahimi *et al.* (1992) and Ahmad and Alwasel (1999), and they come to the conclusion that the hypothesis of exponentiality is tenable in this case.

Let us begin with the data set of Wadsworth (1990), the inter arrival times have been ordered in Table IV.

From the data we obtain $T_{1,n} = 2.83$ and $T_{2,n} = 2.87$. Both statistics reject the hypothesis of exponentiality with an estimated p-value well below the 0.001 level.

As far as the second data set is concerned, the mileages have been ordered in Table V.

Performing the computations we have the values $T_{1,n} = 0.65$ with an estimated p-value of 0.44 $T_{2,n} = 0.1283$ with an estimated p-value of 0.18.

TABLE IV Inter Arrival Times.

1.80	3.43	3.98	4.23	4.65
2.89	3.48	4.06	4.34	4.84
2.93	3.57	4.11	4.37	4.91
3.03	3.85	4.13	4.53	4.99
3.15	3.92	4.16	4.62	5.17

TABLE V Mileages.

162	508	884	1603
200	539	1003	1984
271	629	1101	2355
320	706	1182	2880
393	778	1463	

Acknowledgements

The authors would like to thank a referee who pointed out a critical error in our earlier draft and helped improve the paper's readability.

References

Ahmad, I. A. and Alwasel, I. (1999). A goodness of fit test for exponentiality based on the memoryless property. *J. Royal Stat. Soc. B*, **61**, 681–689.

Alwasel, I. (2001). On goodness of fit testing for exponentiality using the memoryless property. *J. of Nonparametric Statistics*, **13**, 569–581.

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

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

Barbe, P. (1994). Joint approximation of processes based on spacings and order statistics. *Stochastic Processes and Their Applications*, **53**, 339–349.

Baringhaus, L. and Henze, N. (1991). A class of consistent tests for exponentiality based on the empirical Laplace transform. *Ann. Inst. Statist. Math.*, **43**, 551–564.

Baringhaus, L. and Henze, N. (1992). An adaptive omnibus test for exponentiality. *Communications in Statistics – Theory and Methods*, **21**, 969–978.

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.

Csörgö, M. and Horváth, L. (1993). *Weighted Approximations in Probability and Statistics*. Wiley, New York.

Csörgö, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics*. Academic Press, Budapest.

Darling, D. (1957). The Kolmogorov Smirnov Cramer von Mises tests. *Ann. Math. Stat.*, **28**, 823–838.

Doksum, K. A. and Yandell, B. S. (1984). Test for exponentiality. *Handbook of Statistics*, Vol. 4, North Holland, pp. 579–612.

Durbin, J. (1975). Kolmogorov Smirnov test when parameters are estimated with applications to test to exponentiality and test on spacings. *Biometrika*, **62**, 5–22.

Ebrahimi, N., Soofi, E. S. and Habibullah, M. (1992). Testing exponentiality based on Kullback-Leibler information. *Journal of the Royal Statistical Society B*, **54**, 739–748.

Gail, M. H. and Gastwirth, J. L. (1978). A scale free goodness of fit test for the exponential distribution based on the Gini statistic. *Journal of the Royal Stat. Soc.*, **40**, 350–357.

Grzegorzewski, P. and Wieczorkowski, R. (1999). Entropy-based goodness of fit test for exponentiality. *Communications in Statistics – Theory and Methods*, **28**, 1183–1202.

Grubbs, F. E. (1971). Fiducial bounds on reliability for the two parameter negative exponential distribution. *Technometrics*, **13**, 873–876.

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

Jammalamadaka, S. R. and Lee, E. S. (1998). Testing for Harmonic New Better than Used in Expectation. *Probab. Eng. Inform. Sci.*, **12**, 409–416.

Klar, B. (2000). A class of tests for exponentiality against HNBUE alternatives. *Statistics and Prob. Letters*, **47**, 199–207.

Komlós, J., Major, P. and Tusnády. (1975). An approximation of partial sums of independent R.V.'s and the sample D.F. *Zeit. Wahrsch. Verv. Gebiete*, **32**, 111–131.

Nikitin, Y. Y. (1995). *Asymptotic Efficiency of Nonparametric Tests*. Cambridge University Press, New York.

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

Nikitin, Y. Y. and Tchirina, A. V. (1996). Bahadur efficiency and local optimality of a test for the exponential distribution based on the Gini statistic. *J. Ital. Stat. Soc.*, **1**, 163–175.

Pyke, R. (1965). Spacings. *J. Royal Stat. Soc. B*, **3**, 395–436.

Seshadri, V., Csörgö, M. and Stephens, M. A. (1969). Tests for the exponential distribution using Kolmogorov-type statistics. *J. Royal Stat. Soc. B*, **31**, 499–509.

Shapiro, S. S. (1995). Goodness of fit test. In: Balakrishnan, N. and Basu, A. P. (Eds.), *The Exponential Distribution: Theory Methods and Applications*. Gordon and Breach, Amsterdam.

Taufer, E. (2000). A new test for exponentiality against omnibus alternatives. *Stochastic Modelling and Applications*, **3**, 23–36.

Taufer, E. (2002). On entropy based test for exponentialiy. *Communications in Statistics – Simulation and Computation*, **31**, 189–200.

Wadsworth, H. M. (1990). Handbook of Statistical Methods for Engineers and Scientists. McGraw-Hill, New York.