Ann. Inst. Statist. Math.
Vol. 55, No. 1, 137–151 (2003)

TESTS OF FIT FOR THE RAYLEIGH DISTRIBUTION BASED ON THE EMPIRICAL LAPLACE TRANSFORM

SIMOS MEINTANIS[1] AND GEORGE ILIOPOULOS[2]

[1] *Department of Engineering Sciences, University of Patras, 261 10 Patras, Greece*
[2] *Department of Mathematics, University of the Aegean, 83 200 Samos, Greece*

(Received September 10, 2001; revised March 4, 2002)

Abstract. In this paper a class of goodness-of-fit tests for the Rayleigh distribution is proposed. The tests are based on a weighted integral involving the empirical Laplace transform. The consistency of the tests as well as their asymptotic distribution under the null hypothesis are investigated. As the decay of the weight function tends to infinity the test statistics approach limit values. In a particular case the resulting limit statistic is related to the first nonzero component of Neyman's smooth test for this distribution. The new tests are compared with other omnibus tests for the Rayleigh distribution.

Key words and phrases: Rayleigh distribution, goodness-of-fit test, empirical Laplace transform, smooth test.

1. Introduction

Next to the exponential law, the Rayleigh distribution is the most widely known special case of the Weibull distribution. It arises from the Weibull density when the shape parameter is set equal to two. Also, the square root of a chi-squared χ^2_ν random variable with $\nu = 2$, that is of an exponential random variable, follows the Rayleigh distribution. The Rayleigh distribution was originally derived in connection with a problem in acoustics, and has been used in modelling certain features of electronic waves and as the distance distribution between individuals in a spatial Poisson process. Most frequently however it appears as a suitable model in life testing and reliability theory. For more details on the Rayleigh distribution the reader is referred to Johnson *et al.* (1994). The appropriateness of the Rayleigh distribution as a model for non-negative measurements can be assessed by testing goodness of fit of the squared data to the exponential distribution. Hence, by applying this transformation to the data, all exponentiality tests can be utilized for the purpose of testing the goodness-of-fit to the Rayleigh distribution. Apart from such tests one can find in the literature a few additional procedures, often restricted though to a subset of alternatives to the Rayleigh model. See for example, Castillo and Puig (1997) and Auinger (1990).

To fix notation, the Rayleigh distribution with density $(2x/\theta^2)\exp(-x^2/\theta^2)$, $x \geq 0$, will be denoted by $Ral(\theta)$. Suppose $X_1, \dots, X_n$, are independent copies of a nonnegative random variable X with unknown distribution. On the basis of $X_1, \dots, X_n$, the hypothesis to be tested is

$$H_0 : \text{The law of } X \text{ is } Ral(\theta) \text{ for some } \theta > 0.$$

Our tool for testing H_0 will be the empirical Laplace transform (*ELT*),

$$l_n(t) = \frac{1}{n}\sum_{j=1}^{n} \exp(-tX_j).$$

The *ELT* has been employed in estimation problems by, among others, Feigin *et al.* (1983), Gawronski and Stadtmüller (1985), Laurence and Morgan (1987), Csörgő and Teugels (1990) and Yao and Morgan (1999). Baringhaus and Henze (1991) apparently initiated the *ELT*-approach in the context of goodness-of-fit testing, which was followed up by Baringhaus and Henze (1992), Henze (1993) and Henze and Meintanis (2002*a*). These authors utilize the *ELT* of properly scaled data for testing exponentiality.

In this paper we study a family of omnibus tests for H_0 that are based on the *ELT*

$$L_n(t) = \frac{1}{n}\sum_{j=1}^{n} \exp(-tY_j),$$

of the scaled data $Y_j = X_j/\hat{\theta}_n$, $j = 1, \dots, n$, where $\hat{\theta}_n$ denotes a consistent estimator of the scale parameter θ. To this end, note that the Laplace transform of $Ral(1)$ is

$$L(t) = 1 - \frac{\sqrt{\pi}}{2} t \exp\left(\frac{t^2}{4}\right)\left[1 - \Phi\left(\frac{t}{2}\right)\right],$$

where $\Phi(\cdot)$ denotes the error function. Then the Laplace transform of $Ral(\theta)$ is $l(t) = L(\theta t)$. Our approach is motivated by the observation that $l(t)$ is the unique solution of the differential equation $ty'(t) - [1 + (\theta^2 t^2/2)]y(t) + 1 = 0$, subject to the condition $\lim_{t\to\infty} y(t) = 0$. Consequently, the random function $tl_n'(t) - [1 + (\hat{\theta}_n^2 t^2/2)]l_n(t) + 1$, $t \geq 0$, should be close to the zero function under H_0, provided that we employ a reasonable estimator $\hat{\theta}_n$ of θ.

In the spirit of Baringhaus and Henze (1991) we propose the statistic

$$T_{n,a} = n\int_0^{\infty} D_n^2(t)\exp(-at)dt, \tag{1.1}$$

for testing the null hypothesis H_0, where $D_n(t) = tL_n'(t) - [1 + (t^2/2)]L_n(t) + 1$ and $a > 0$ is a constant. Rejection of H_0 is for large values of $T_{n,a}$. A closed-form expression for $T_{n,a}$, obtained by straightforward manipulations of integrals, is

$$\begin{aligned} T_{n,a} = {} & \frac{n}{a} + \frac{1}{n}\sum_{j,k=1}^{n}\left[\frac{1}{(Y_j+Y_k+a)} + \frac{Y_j+Y_k}{(Y_j+Y_k+a)^2} + \frac{2Y_jY_k+2}{(Y_j+Y_k+a)^3}\right. \\ & \left. + \frac{3(Y_j+Y_k)}{(Y_j+Y_k+a)^4} + \frac{6}{(Y_j+Y_k+a)^5}\right] \\ & - 2\sum_{j=1}^{n}\left[\frac{1}{Y_j+a} + \frac{Y_j}{(Y_j+a)^2} + \frac{1}{(Y_j+a)^3}\right]. \end{aligned} \tag{1.2}$$

This expression shows that $T_{n,a}$, like each of the statistics dealt with in this paper, depends on $X_1, \dots, X_n$ solely via $Y_1, \dots, Y_n$ and thus has the desirable feature of being invariant with respect to scale changes. Consequently, the null distribution of $T_{n,a}$ does not depend on the parameter θ of the Rayleigh distribution. The 'free' parameter a figuring in (1.2) offers great flexibility with regard to the power of a test based on $T_{n,a}$. From Tauberian theorems on Laplace transforms (see Baringhaus and Henze (1991), p. 552), it may be anticipated that choosing a small value of a, which means letting the weight function decay slowly, will give high power against alternative distributions having a point mass or infinite density at zero. On the other hand, a large value of a means putting most of the mass of the weight function near zero, which should give high power against alternatives that greatly differ in tail behavior with respect to the Rayleigh distribution.

The paper is organized as follows. Section 2 deals with the weak convergence of $T_{n,a}$ under H_0, and the consistency of the test based on $T_{n,a}$. The theoretical properties of $T_{n,a}$ are derived for two choices on the estimator of the scale parameter, namely the maximum likelihood estimator and the moment estimator. In both cases we show that, under general conditions, the class of test statistics $(T_{n,a})_{a>0}$ is 'closed at the boundary $a = \infty$' by establishing a 'limit statistic'. The limit statistic corresponding to the moment estimator for θ is related to Neyman's smooth test for the Rayleigh distribution (see Section 6.3 of Rayner and Best (1989) for an account on smooth tests of fit). In Section 3 we present the results of a Monte Carlo study on the power of the new tests in comparison with several goodness-of-fit tests for the Rayleigh distribution. The final section illustrates the applicability of the proposed procedures on real data sets.

2. Theoretical results

In what follows, $\to^{\mathcal{D}}$ denotes weak convergence of random variables or stochastic processes, $\to^{\mathrm{P}}$ is convergence in probability, $o_P(1)$ stands for convergence in probability to 0, and i.i.d. means 'independent and identically distributed'. Finally, recall the notation $Y_j = X_j/\hat{\theta}_n$ from Section 1. The reasoning below follows similar lines as the proof of Theorem 2.1 in Henze and Meintanis (2002*a*). The starting point for asymptotics is the representation

$$T_{n,a} = \int_0^\infty Z_n^2(t)\exp(-at)dt,$$

where

$$Z_n(t) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left[\left(1 + tY_j + \frac{t^2}{2}\right)\exp(-tY_j) - 1\right], \qquad 0 \le t < \infty. \tag{2.1}$$

The process Z_n is a random element of the set $C[0,\infty)$ of continuous functions on $[0,\infty)$, equipped with the metric $\rho(g,h) = \sum_{k=1}^{\infty} 2^{-k}\min[1,\rho_k(g,h)]$, where $\rho_k(g,h) = \max_{0\le t\le k}|g(t) - h(t)|$.

THEOREM 2.1. *Let $X_1,\ldots,X_n$ be a sequence of i.i.d. random variables with distribution $Ral(\theta)$, and assume that θ is estimated either by the method of maximum likelihood (ML) or by the method of moments (MO). Then $Z_n \to^{\mathcal{D}} Z$ in $C[0,\infty)$, where Z is a zero mean Gaussian process in $C[0,\infty)$ with covariance kernel $K(s,t)$.*

(a) *If θ is estimated by the method of ML, the covariance kernel is given by*

$$K(s,t) = \frac{s^2t^2}{4}[L(s+t) - L(s)L(t)] \qquad (s,t \ge 0). \tag{2.2}$$

(b) *If θ is estimated by the method of MO, the covariance kernel is given by*

$$K(s,t) = \frac{s^2t^2}{4}L(s+t) + 2s^2L(s)\left[\frac{1}{\sqrt{\pi}}\{tL(t) - L'(t)\} - \frac{1}{2}\right] + 2t^2L(t)\left[\frac{1}{\sqrt{\pi}}\{sL(s) - L'(s)\} - \frac{1}{2}\right] + \left(\frac{4-\pi}{\pi}\right)s^2t^2L(s)L(t) \qquad (s,t \geq 0). \tag{2.3}$$

PROOF. (a) Without loss of generality assume that $\theta = 1$ and let $\hat{\theta}_n = \sqrt{n^{-1}\sum_{j=1}^n X_j^2}$ be the *ML* estimator of θ. Fix an integer k, since weak convergence in $(C[0,\infty), \rho)$ is weak convergence on each interval $[0,k]$, $k \in \mathbb{N}$. For $0 \leq t < \infty$, let

$$Z_n^*(t) = \frac{1}{\sqrt{n}}\sum_{j=1}^n \left[\left(1 + tX_j + \frac{t^2}{2}\right)\exp(-tX_j) - 1\right] + \left(\frac{t}{2}\right)^2 A(t)U_n(t), \tag{2.4}$$

where $U_n(t) = n^{-1/2}\sum_{j=1}^n(X_j^2 - 1)$ and $A(t) = E[(2X+t)X\exp(-tX)] = 2L(t)$. We first prove

$$\max_{0\leq t\leq k}|Z_n(t) - Z_n^*(t)| = o_P(1) \tag{2.5}$$

and thus $\rho(Z_n, Z_n^*) = o_P(1)$. To this end, a Taylor expansion of $h(u) = \exp(-tu)$, $u > 0$, around $u = X_j$ gives

$$\left(1 + tY_j + \frac{t^2}{2}\right)e^{-tY_j} = \left(1 + tX_j + \frac{t^2}{2}\right)e^{-tX_j} - \Delta_j(2X_j + t)\frac{t^2}{2}e^{-tX_j} + \varepsilon_{n,j}(t),$$

where $\Delta_j = Y_j - X_j$ and $\max_{0\leq t\leq k}|n^{-1/2}\sum_{j=1}^n \varepsilon_{n,j}(t)| = o_P(1)$. Consequently,

$$\max_{0\leq t\leq k}|Z_n(t) - \tilde{Z}_n(t)| = o_P(1), \tag{2.6}$$

where

$$\tilde{Z}_n(t) = \frac{1}{\sqrt{n}}\sum_{j=1}^n\left[\left\{\left(1 + tX_j + \frac{t^2}{2}\right)e^{-tX_j} - 1\right\} - \Delta_j(2X_j + t)\frac{t^2}{2}\exp(-tX_j)\right].$$

The mean value theorem for $h(u) = \sqrt{u}$, $u > 0$, yields $\hat{\theta}_n - 1 = (2\theta^*)^{-1}(n^{-1}\sum_{j=1}^n X_j^2 - 1)$, where θ^* lies between $\hat{\theta}_n$ and 1. Now use the compactness of $[0,k]$, the consistency of $\hat{\theta}_n$, the continuity of $n^{-1}\sum_{j=1}^n(2X_j + t)X_j\exp(-tX_j)$ and the law of large numbers to obtain

$$\max_{0\leq t\leq k}\left|(\hat{\theta}_n\theta^*)^{-1}\frac{1}{n}\sum_{j=1}^n(2X_j + t)X_j\exp(-tX_j) - A(t)\right| = o_P(1),$$

which in turn implies $\rho(\tilde{Z}_n, Z_n^*) = o_P(1)$. In view of (2.6), (2.5) follows. It thus remains to prove $Z_n^* \to^{\mathcal{D}} Z$ in $C[0,\infty)$. Since the finite-dimensional distributions of Z_n^* converge to centered multivariate normal distributions with a covariance structure given by the kernel $K(\cdot,\cdot)$ in (2.2), the proof is finished if we can show tightness of the sequence $Z_n^*(\cdot)$. To this end, let $g(x,t) = [1+tx+(t^2/2)]e^{-tx} - 1 + (t/2)^2 A(t)(x^2-1)$. Then

$$\max_{0\le t,s\le k} |g(X,s) - g(X,t)| \le |s-t|M,$$

where for $\lambda = k + k^2/2$, $M = (k+\lambda)X^2 + (\lambda - k + 2)|X| + (k+\lambda)$ is a (positive) random variable satisfying $E(M^2) < \infty$. Hence $Z_n^*(\cdot)$ is tight and the proof is completed.

(b) Let $\hat{\theta}_n = (2/\sqrt{\pi})n^{-1}\sum_{j=1}^n X_j$ be the *MO* estimator of θ. The proof follows along similar lines, with Z_n^* produced by replacing $(t/2)^2$ by t^2 and $U_n(t)$ by $n^{-1/2}\sum_{j=1}^n[(X_j/\sqrt{\pi})-1/2]$, in the second term of the right hand side of (2.4). The value of M in the tightness bound is replaced by $kX^2 + [2-k+\{1+(2/\sqrt{\pi})\}\lambda]|X| + (k+\lambda)$. □

The next result can be readily obtained by adapting the reasoning in the proof of Theorem 2.2 of Henze and Wagner (1997).

COROLLARY 2.1. *Under the conditions of Theorem 2.1, we have*

$$T_{n,a} = \int_0^\infty Z_n^2(t)\exp(-at)dt \xrightarrow{\mathcal{D}} T_a = \int_0^\infty Z^2(t)\exp(-at)dt.$$

Remark 2.1. The distribution of T_a is that of $\sum_{j\ge1}\nu_j(a)N_j^2$, where $N_1, N_2, \ldots$ are independent unit normal random variables and $(\nu_j(a))_{j\ge1}$ are the nonzero eigenvalues of the integral operator O defined by

$$Og(s) = \int_0^\infty K(s,t)g(t)\exp(-at)dt.$$

There is little hope to solve the equation $Og(s) = \nu g(s)$ and thus to determine $\nu_j(a)$ explicitly. However, we can obtain the expectation of T_a, via the relation

$$E[T_a] = \int_0^\infty K(t,t)\exp(-at)dt.$$

Let us denote by T_a^L (resp. T_a^M) the asymptotic test statistic T_a corresponding to Z with covariance kernel given by (2.2) (resp. (2.3)). Then by straightforward manipulations of integrals we have:

$$E[T_a^L] = \frac{\sqrt{\pi}}{4}[2^5\{2\mathcal{L}_5(2a) - \sqrt{\pi}\Lambda_6(2a)\} - \mathcal{L}_5(a)]$$

and

$$\begin{aligned} E[T_a^M] = {} & \left(\frac{96}{\pi} - 18\right)\frac{1}{a^5} + \frac{12}{a^4\sqrt{\pi}} - \frac{4}{a^3} - \frac{\sqrt{\pi}}{4}\mathcal{L}_5(a) + 2^4[\mathcal{L}_2(2a) + \sqrt{\pi}\mathcal{L}_3(2a)] \\ & + 2^6\left[\left(\sqrt{\pi} - \frac{4}{\sqrt{\pi}}\right)\mathcal{L}_5(2a) - \mathcal{L}_4(2a)\right] \\ & - 2^4\sqrt{\pi}[\Lambda_3(2a) - 2\Lambda_5(2a)] + 2^5(4-\pi)\Lambda_6(2a), \end{aligned}$$

where

$$\mathcal{L}_\nu(a) = \int_0^\infty f_\nu(a,t)dt \quad \text{and} \quad \Lambda_\nu(a) = \int_0^\infty f_{\frac{\nu}{2}}^2\left(\frac{a}{2},t\right)dt,$$

with $f_\nu(a,t) = t^\nu e^{t^2-at}[1-\Phi(t)]$.

Let us now denote by $T_{n,a}^L$ (resp. $T_{n,a}^M$) the test statistic $T_{n,a}$ when θ is estimated by the method of ML (resp. method of MO). The next result studies the asymptotic behavior of the (suitably rescaled) test statistics $T_{n,a}^L$ and $T_{n,a}^M$. The asymptotic test statistic corresponding to $T_{n,a}^M$ is related to the first nonzero component of Neyman's smooth test. For further examples on the connection between weighted integral test statistics and components of smooth tests of fit, see Baringhaus *et al.* (2000).

THEOREM 2.2. *Assume that n is fixed, and let $\bar{Y}_n = n^{-1}\sum_{j=1}^n Y_j$ and $\bar{Y}_n^k = n^{-1}\sum_{j=1}^n Y_j^k$, $k \geq 2$. Then we have,*

$$\text{(a)} \quad T_n^L := \lim_{a\to\infty} a^7 T_{n,a}^L = 20n(2\bar{Y}_n^3 - 3\bar{Y}_n)^2,$$

and

$$\text{(b)} \quad T_n^M := \lim_{a\to\infty} a^5 T_{n,a}^M = 6n(\bar{Y}_n^2 - 1)^2.$$

PROOF. (a) Observe that $T_{n,a} = \int_0^\infty g(t)\exp(-at)dt$, where $g(t) = Z_n^2(t)$. A Taylor expansion gives

$$Z_n(t) = -\frac{t^2}{2}\frac{1}{\sqrt{n}}\sum_{j=1}^n (Y_j^2 - 1) + \frac{t^3}{6}\frac{1}{\sqrt{n}}\sum_{j=1}^n (2Y_j^3 - 3Y_j) + O(t^4), \qquad \text{as} \quad t \to 0^+.$$

Since the ML estimator is employed in Y_j, we have $\sum_{j=1}^n (Y_j^2 - 1) = 0$. Hence

$$g(t) \sim \frac{t^6}{36n}\left[\sum_{j=1}^n (2Y_j^3 - 3Y_j)\right]^2, \qquad \text{as} \quad t \to 0^+.$$

Application of Proposition 1.1 of Baringhaus *et al.* (2000), yields the first asymptotic result.

(b) The Taylor expansion now gives

$$g(t) \sim \frac{t^4}{4n}\left[\sum_{j=1}^n (Y_j^2 - 1)\right]^2, \qquad \text{as} \quad t \to 0^+.$$

Application of the same proposition as in part (a) yields the second asymptotic result. □

Remark 2.2. Note that as $n \to \infty$, T_n^L measures the deviation of $2E[(X/\theta)^3]$ from $3E[(X/\theta)]$ whereas T_n^M measures the deviation of $E[(X/\theta)^2]$ from unity, both being zero under the null hypothesis. Moreover, T_n^M enjoys an interesting relation with the first nonzero component of Neyman's smooth test for H_0: The first three orthogonal polynomials for the Rayleigh density are $h_0(x;\theta) = 1$, $h_1(x;\theta) = c_1[(x/\theta) - \sqrt{\pi}/2]$, and

$$h_2(x;\theta) = c_2\left[\left(\frac{x}{\theta}\right)^2 - \frac{\sqrt{\pi}}{4-\pi}\frac{x}{\theta} + \frac{3\pi - 8}{2(4-\pi)}\right],$$

where $c_1 = 2/\sqrt{(4-\pi)}$ and $c_2 = 2\sqrt{(4-\pi)/(16-5\pi)}$. As it was pointed out by a referee, these polynomials are a special case of the so-called speed polynomials that emerge in the solution of the Boltzmann and the Fokker–Planck equation (see Clarke and Shizgal (1993)). The k-th component of Neyman's smooth test $\hat{u}_{nk} = n^{-1/2}\sum_{j=1}^{n} h_k(X_j, \hat{\theta}_n)$ vanishes for $k = 1$, when θ is estimated by the method of MO. The corresponding asymptotic test statistic can be written as $T_n^M = (6/c_2^2)\hat{u}_{n2}^2$. Hence T_n^M, apart from a constant factor, coincides with the square of the first nonzero component of Neyman's smooth test based on the polynomials that are orthogonal with respect to $Ral(\theta)$.

We now consider the asymptotic behavior of $T_{n,a}$ in a more general, nonparametric, setting. Our result is a weak limit law for $T_{n,a}$ under fixed alternatives to H_0.

THEOREM 2.3. *Assume that the distribution of the nonnegative random variable X is not degenerate at zero.*

(a) *If $E(X^2) < \infty$, we have*

$$n^{-1}T_{n,a}^L \xrightarrow{\mathrm{P}} \int_0^\infty \left[tL'\left(\frac{t}{\sqrt{E(X^2)}}\right) - \left(1+\frac{t^2}{2}\right)L\left(\frac{t}{\sqrt{E(X^2)}}\right) + 1\right]^2 \exp(-at)dt, \tag{2.7}$$

whereas if $E(X^2) = \infty$, we have

$$n^{-1}T_{n,a}^L \xrightarrow{\mathrm{P}} 6a^{-5}. \tag{2.8}$$

(b) *If $E(X) < \infty$, we have*

$$n^{-1}T_{n,a}^M \xrightarrow{\mathrm{P}} \int_0^\infty \left[tL'\left(\frac{t\sqrt{\pi}}{2E(X)}\right) - \left(1+\frac{t^2}{2}\right)L\left(\frac{t\sqrt{\pi}}{2E(X)}\right) + 1\right]^2 \exp(-at)dt, \tag{2.9}$$

whereas if $E(X) = \infty$, we have

$$n^{-1}T_{n,a}^M \xrightarrow{\mathrm{P}} 6a^{-5}. \tag{2.10}$$

PROOF. (a) Starting with (1.1) and using $D_n^2(t) \leq [3 + (t^2/2)]^2$, dominated convergence and Fubini's theorem yield the convergence of $E(n^{-1}T_{n,a}^L)$ to the right-hand side of (2.7) or (2.8) respectively, according to whether $E(X^2) < \infty$ or $= \infty$. Also $\mathrm{Var}(n^{-1}T_{n,a}^L) \to 0$. Notice that $\hat{\theta}_n \to \infty$ almost surely if $E(X^2) = \infty$.

(b) It follows the same reasoning as the proof of part (a). □

Since the right-hand sides of (2.8) and (2.10) are always positive, and the right-hand sides of (2.7) and (2.9) are positive if X does not follow the Rayleigh distribution, it follows from Corollary 2.1 and Theorem 2.3 that a level α-test that rejects H_0 for large values of $T_{n,a}^L$ or $T_{n,a}^M$, is consistent against each fixed alternative distribution not degenerate at zero.

Table 1. Percentage points based on 50 000 Monte Carlo samples of size $n = 20$ (first line) and $n = 50$ (second line), for $a \in \{0.5, 1.0, 2.0, 5.0, 10.0\}$ and significance level α.

$a =$	0.5		1.0		2.0		5.0		10.0	
$\alpha =$	0.05	0.10	0.05	0.10	0.05	0.10	0.05	0.10	0.05	0.10
$T^L_{n,a}$	4.33	3.03	0.37	0.27	0.021	0.015	0.21^3	0.15^3	0.39^5	0.28^5
	4.66	3.34	0.38	0.28	0.021	0.015	0.22^3	0.16^3	0.42^5	0.30^5
$T^M_{n,a}$	3.35	2.31	0.22	0.16	0.82^2	0.59^2	0.35^4	0.26^4	0.22^5	0.16^5
	3.63	2.58	0.23	0.17	0.84^2	0.60^2	0.37^4	0.26^4	0.22^5	0.16^5
BH^L	0.58	0.41	0.30	0.22	0.13	0.097	0.030	0.022	0.71^2	0.51^2
	0.56	0.41	0.31	0.22	0.14	0.099	0.033	0.023	0.82^2	0.58^2
BH^M	0.37	0.27	0.18	0.13	0.070	0.051	0.022	0.016	0.014	0.98^2
	0.37	0.27	0.18	0.13	0.072	0.052	0.022	0.016	0.014	0.010
HE^L	0.18	0.13	0.059	0.043	0.014	0.99^2	0.11^2	0.79^3	0.10^2	0.71^3
	0.19	0.14	0.060	0.043	0.014	0.010	0.12^2	0.85^3	0.12^3	0.83^4
HE^M	0.088	0.064	0.022	0.016	0.37^2	0.27^2	0.32^3	0.24^3	0.10^3	0.72^4
	0.088	0.064	0.022	0.016	0.39^2	0.28^2	0.32^3	0.24^3	0.12^3	0.80^4
HM^L_1	16.0	13.4	2.30	1.82	0.27	0.21	0.70^2	0.51^2	0.17^3	0.10^3
	16.2	13.4	2.28	1.81	0.27	0.21	0.81^2	0.59^2	0.24^3	0.16^3
HM^M_1	15.5	13.0	2.04	1.62	0.20	0.16	0.51^2	0.35^2	0.43^3	0.31^3
	15.6	13.0	2.04	1.62	0.21	0.16	0.52^2	0.38^2	0.43^3	0.31^3
HM^L_2	1.62	1.20	0.49	0.36	0.11	0.086	0.011	0.74^2	0.15^2	0.84^3
	1.58	1.19	0.49	0.36	0.13	0.093	0.014	0.010	0.22^2	0.14^2
HM^M_2	1.23	0.92	0.32	0.24	0.072	0.050	0.99^2	0.63^2	0.27^2	0.19^2
	1.23	0.93	0.33	0.25	0.076	0.056	0.98^2	0.68^2	0.27^2	0.20^2

$0.a^b$ denotes the number $0.a \times 10^{-b}$.

3. Simulations

This section presents the results of a Monte Carlo study conducted to assess the power of the new tests. We compare the new tests with alternative procedures which were initially developed in order to test goodness-of-fit to the exponential distribution. However, tests of fit for the Rayleigh distribution result if we apply these procedures to the squared data Y_j^2, instead of Y_j, $j = 1, 2, \ldots, n$. All calculations were done at the Department of Engineering Sciences, University of Patras, using double precision arithmetic in FORTRAN and routines from the IMSL library, whenever available. The proposed procedures are compared with the following tests for several values of the weight parameter a:

i) *The tests of Baringhaus and Henze* (1991),

$$BH = n \int_0^\infty [(1+t)\psi_n'(t) + \psi_n(t)]^2 \exp(-at)dt,$$

Table 2. Percentage of rejection for 10 000 Monte Carlo samples of size $n = 20$ at significance level $\alpha = 0.05$.

	$T^L_{n,a}$	BH^L	HE^L	HM^L_1	HM^L_2	$T^L_{n,a}$	BH^L	HE^L	HM^L_1	HM^L_2
altern.			$a = 1.0$					$a = 2.0$		
$W(1.0)$	97	93	93	86	83	96	92	91	85	83
$W(2.0)$	5	5	5	5	5	5	5	5	5	5
$W(3.0)$	40	52	53	45	41	47	51	50	47	27
$G(1.5)$	76	71	71	56	57	75	69	68	59	61
$G(2.0)$	43	43	43	30	36	44	43	43	36	41
$IG(0.5)$	98	98	98	96	96	98	98	98	95	96
$IG(1.5)$	48	63	64	51	61	57	65	65	60	65
$LN(0.8)$	66	75	75	62	70	71	75	75	70	74
$LN(1.5)$	⋆	⋆	⋆	99	99	⋆	⋆	⋆	99	99
$GO(0.5)$	84	70	69	55	47	80	65	63	52	49
$GO(1.5)$	57	32	30	24	15	46	25	24	18	15
$PW(1.0)$	43	14	12	23	9	30	8	7	13	6
$PW(2.0)$	99	91	90	86	64	98	84	82	75	54
$LF(2.0)$	70	52	51	38	33	63	47	46	35	36
$LF(4.0)$	57	38	36	26	23	49	34	32	24	26
$EP(1.0)$	70	47	46	36	26	61	40	40	30	27
$EP(2.0)$	16	28	29	26	35	22	30	29	35	24
$PE(3.0)$	88	72	71	55	50	83	67	66	53	53
$PE(4.0)$	60	43	43	29	29	54	40	40	30	34

⋆ denotes power 100%.

and Henze (1993)

$$HE = n \int_0^\infty \left(\psi_n(t) - \frac{1}{1+t}\right)^2 \exp(-at)dt,$$

where $\psi_n(t)$ denotes the ELT of Y_j^2. We denote by BH^L (resp. BH^M) and by HE^L (resp. HE^M) the tests in which the data Y_j are computed by employing the ML estimator (resp. MO estimator) for θ. Computationally simple forms for the tests statistics can be found in Henze (1993).

ii) *The tests of Henze and Meintanis* (2002*b*),

$$HM = n \int_0^\infty [S_n(t) - tC_n(t)]^2 w(t)dt,$$

where $C_n(t) = n^{-1}\sum_{j=1}^n \cos(tY_j^2)$, $S_n(t) = n^{-1}\sum_{j=1}^n \sin(tY_j^2)$, and $w(t)$ is a weight function. When $w(t) = \exp(-at)$, HM_1^L (resp. HM_1^M) denotes the test corresponding to the ML estimator (resp. the MO estimator) for θ. For $w(t) = \exp(-at^2)$, the resulting tests are denoted by HM_2^L and HM_2^M. Computationally simple forms for the test statistics can be found in Henze and Meintanis (2002*b*).

Empirical critical values for these test statistics were computed based on 50 000 Monte Carlo replications and are given in Table 1 for significance level $\alpha = 0.05$ and $\alpha = 0.10$.

Table 3. Percentage of rejection for 10 000 Monte Carlo samples of size $n = 20$ at significance level $\alpha = 0.05$.

	$T_{n,a}^L$	BH^L	HE^L	HM_1^L	HM_2^L	$T_{n,a}^L$	BH^L	HE^L	HM_1^L	HM_2^L
altern.			$a = 5.0$					$a = 10.0$		
$W(1.0)$	95	90	90	83	81	93	88	82	79	78
$W(2.0)$	5	5	5	5	5	5	5	5	5	5
$W(3.0)$	52	46	46	22	0	51	41	41	0	0
$G(1.5)$	74	67	67	62	61	72	66	69	59	58
$G(2.0)$	45	43	44	42	43	45	43	57	41	41
$IG(0.5)$	98	98	97	96	95	98	97	94	94	94
$IG(1.5)$	63	66	67	67	67	65	67	74	65	64
$LN(0.8)$	75	76	76	74	74	76	75	79	72	72
$LN(1.5)$	$\star$	$\star$	99	99	99	$\star$	99	97	98	98
$GO(0.5)$	74	59	59	49	47	69	56	58	44	43
$GO(1.5)$	34	20	20	14	14	29	18	32	12	12
$PW(1.0)$	13	5	5	4	1	9	3	5	1	0
$PW(2.0)$	93	73	72	51	41	88	65	43	37	34
$LF(2.0)$	56	43	43	36	36	52	41	51	34	33
$LF(4.0)$	41	30	31	26	26	38	29	44	25	24
$EP(1.0)$	51	34	34	26	25	45	31	42	23	22
$EP(2.0)$	28	29	29	19	0	28	25	25	0	0
$PE(3.0)$	77	62	62	54	52	72	59	63	50	49
$PE(4.0)$	48	38	39	35	35	45	37	52	33	33

$\star$ denotes power 100%.

At the suggestion of a referee, we have also included in the comparisons the Kolmogorov–Smirnov (KS) exponentiality test performed on the squared data, implemented via Algorithm 2 of Edgeman and Scott (1987).

For the nominal level 5%, Tables 2–6 show power estimates of the tests under discussion. The entries are the percentages of 10 000 Monte Carlo samples that resulted in rejection of H_0, rounded to the nearest integer.

The following alternative distributions are considered, all concentrated on the positive half-line:

- the Weibull distribution with density $\theta x^{\theta-1}\exp(-x^{\theta})$, denoted by $W(\theta)$,
- the gamma distribution with density $\Gamma(\theta)^{-1}x^{\theta-1}\exp(-x)$, denoted by $\Gamma(\theta)$,
- the inverse Gaussian law $IG(\theta)$ with density $(\theta/2\pi)^{1/2}x^{-3/2}\exp[-\theta(x-1)^2/2x]$,
- the lognormal law $LN(\theta)$ with density $(\theta x)^{-1}(2\pi)^{-1/2}\exp[-(\log x)^2/(2\theta^2)]$,
- the Gompertz law $GO(\theta)$ having distribution function $1-\exp[\theta^{-1}(1-e^x)]$,
- the power distribution $PW(\theta)$ with density $\theta^{-1}x^{(1-\theta)/\theta}$, $0 \le x \le 1$,
- the linear increasing failure rate law $LF(\theta)$ with density $(1+\theta x)\exp(-x-\theta x^2/2)$,
- the exponential-power $EP(\theta)$ law having distribution function $1-\exp[1-\exp(x^{\theta})]$,
- the Poisson-exponential law $PE(\theta)$, which is the distribution of $E_1+\cdots+E_N$, where $N, E_1, E_2, \ldots$ are independent, N has a Poisson distribution with $E[N]=\theta$, and for $j \ge 1$, E_j is exponentially distributed with parameter equal to one.

These distributions comprise widely used models in reliability and life testing, areas

Table 4. Percentage of rejection for 10 000 Monte Carlo samples of size $n = 20$ at significance level $\alpha = 0.05$.

	$T_{n,a}^M$	BH^M	HE^M	HM_1^M	HM_2^M	$T_{n,a}^M$	BH^M	HE^M	HM_1^M	HM_2^M
altern.			$a = 1.0$					$a = 2.0$		
$W(1.0)$	96	94	94	80	83	96	95	94	84	86
$W(2.0)$	5	5	5	5	5	5	5	5	5	5
$W(3.0)$	35	47	47	41	16	43	44	41	35	0
$G(1.5)$	72	73	72	47	58	75	75	73	57	63
$G(2.0)$	37	44	45	23	38	43	48	47	34	44
$IG(0.5)$	96	98	98	93	96	98	99	98	96	97
$IG(1.5)$	33	62	64	38	62	50	67	69	57	69
$LN(0.8)$	55	74	75	50	69	66	77	78	66	76
$LN(1.5)$	⋆	⋆	⋆	99	99	⋆	⋆	⋆	99	99
$GO(0.5)$	84	75	72	52	53	83	75	71	54	56
$GO(1.5)$	60	40	37	25	22	54	38	34	24	23
$PW(1.0)$	50	24	19	28	22	37	21	19	26	24
$PW(2.0)$	99	96	94	87	82	99	95	93	85	84
$LF(2.0)$	70	57	55	35	37	68	58	54	38	41
$LF(4.0)$	58	43	41	24	27	54	43	39	27	30
$EP(1.0)$	71	55	52	35	34	67	54	49	36	35
$EP(2.0)$	13	24	25	25	20	18	24	24	29	24
$PE(3.0)$	88	76	73	50	53	86	76	72	54	57
$PE(4.0)$	61	47	45	25	32	57	49	45	31	37

⋆ denotes power 100%.

where the Rayleigh distribution is most frequently encountered, and include densities f with increasing and decreasing hazard rates $f(x)/(1 - F(x))$ as well as models with U-shaped and inverted U-shaped hazard functions.

The main conclusions that can be drawn from the simulation results are the following:

1. In most cases and for the same value of a, the test statistic in which the MO estimator is employed is more powerful than the one in which the ML estimator is employed. Differences in power are more pronounced when testing against one of the distributions in the second part of the tables (GO, PW, LF, EP, PE).

2. Under alternatives belonging to the first part of the tables (W, G, IG, LN) and for $a = 1.0$ or $a = 2.0$, $T_{n,a}^L$ and $T_{n,a}^M$ are either less powerful or have a slight advantage over the most powerful of the other tests, this being the Baringhaus and Henze (1991) or the Henze (1993) test. For the same values of a but for alternative distributions belonging to the second part of the tables, the tests proposed herein are in most cases the most powerful (at times by a wide margin), usually followed by the Baringhaus and Henze (1991) test.

3. For $a = 5.0$ or $a = 10.0$ and for alternative distributions contained in the first part of the tables, the situation reported above persists, the only difference being that now HM_1^L is in some cases the best test. For the same values of a but for alternative distributions belonging to the second part, $T_{n,a}^L$ outperforms its competitors in the great

Table 5. Percentage of rejection for 10 000 Monte Carlo samples of size $n = 20$ at significance level $\alpha = 0.05$.

	$T^M_{n,a}$	BH^M	HE^M	HM^M_1	HM^M_2	$T^M_{n,a}$	BH^M	HE^M	HM^M_1	HM^M_2
altern.			$a = 5.0$					$a = 10.0$		
$W(1.0)$	96	96	96	92	92	96	96	96	96	95
$W(2.0)$	5	5	5	5	5	5	5	5	5	5
$W(3.0)$	45	40	37	14	2	44	40	41	37	32
$G(1.5)$	77	77	76	70	70	77	77	78	77	75
$G(2.0)$	49	49	48	48	46	47	48	54	49	47
$IG(0.5)$	99	99	98	98	97	99	99	98	99	98
$IG(1.5)$	62	66	65	70	67	60	62	68	65	64
$LN(0.8)$	75	77	76	78	76	74	75	77	76	75
$LN(1.5)$	⋆	⋆	⋆	⋆	⋆	⋆	⋆	⋆	⋆	⋆
$GO(0.5)$	80	79	80	68	68	81	80	81	80	78
$GO(1.5)$	46	44	50	32	34	47	45	51	46	46
$PW(1.0)$	25	25	41	30	33	26	25	33	29	33
$PW(2.0)$	97	97	98	94	96	98	97	98	97	98
$LF(2.0)$	65	63	64	50	52	65	64	67	64	62
$LF(4.0)$	50	48	50	38	39	50	49	54	49	47
$EP(1.0)$	62	60	63	46	48	62	61	65	61	60
$EP(2.0)$	21	19	17	17	17	20	18	18	17	14
$PE(3.0)$	83	82	82	70	72	83	82	84	82	81
$PE(4.0)$	55	54	54	45	45	55	54	60	54	52

⋆ denotes power 100%.

Table 6. Percentage of rejection for 10 000 Monte Carlo samples of size $n = 20$ at significance level $\alpha = 0.05$.(1)

	W	G	IG	LN	GO	PW	LF	EP	PE
T^L_n	91 5 46	69 44	98 66	76 ⋆	62 22	5 79	46 33	38 28	67 41
T^M_n	96 5 39	78 50	99 66	77 ⋆	79 41	18 96	63 48	59 18	81 55
KS	86 5 39	57 32	97 56	67 ⋆	56 23	16 87	38 26	35 25	56 30

⋆ denotes power 100%.

(1)distributions from left to right as they appear in Tables 2–5.

majority of cases. When the MO estimator is employed in the test statistics, then $T^M_{n,a}$ is either the best or (with the exception of testing against the $PW(1.0)$ distribution) the second best test, outperformed only by the HE^M test.

4. It can be seen from Table 6 that, as $a \to \infty$, the resulting 'limit tests' based on T^L_n and T^M_n retain the characteristics already revealed in Tables 2–5. For example under the $G(2.0)$ alternative, the power of both the $T^L_{n,a}$ and the $T^M_{n,a}$ tests is not significantly affected by the value of a. Then the corresponding 'limit statistic' T^L_n (resp. T^M_n) has a similar power to the test based on $T^L_{n,a}$ (resp. $T^M_{n,a}$), uniformly in a. In other cases however, such as testing against the $PW(1.0)$ distribution, the power of T^L_n (resp.

T_n^M) greatly differs from that of the test based on $T_{n,a}^L$ (resp. $T_{n,a}^M$) for $a = 1.0$ or 2.0. Despite the fact that $T_{n,a}^L$ and $T_{n,a}^M$ are overall more powerful than the corresponding 'limit statistics', the tests based on T_n^L and T_n^M have considerable power against specific alternatives, and they are certainly more powerful than the KS test.

By taking into account the competitive performance of the BH, the HE and the HM test reported in Baringhaus and Henze (1991), Henze (1993) and Henze and Meintanis (2002*b*), respectively, against classical procedures, including the Kolmogorov–Smirnov and the Cramér–von Mises procedures, we conclude that $T_{n,a}^L$ and $T_{n,a}^M$ (perhaps with a compromise value for a, such as $a = 2.0$), constitute serious competitors for the existing goodness-of-fit tests for the Rayleigh distribution.

4. Real data examples

In this section we apply the proposed procedures to two data sets which have been recently employed by researchers, and compare their conclusions based on alternative methods, to the conclusions reached by our methods. The first data set represents 26 fracture toughness measurements for steels at given temperatures, and appears in Bowman and Shenton (2001), Section 4. The authors report a satisfactory fit for this data set to a two-parameter (location-scale) Weibull model with known shape parameter equal to two (the Rayleigh distribution), and estimated location parameter equal to 25.85. After subtracting the estimate of the location parameter, we have applied the proposed procedures to the toughness data and the results are shown in Fig. 1. They represent in logarithmic scale, the values of the two test statistics $T_{n,a}^L$ and $T_{n,a}^M$, for several values of the weight parameter and the corresponding critical points (CR^L and CR^M) computed by simulation. Regardless of the choice of a and the estimation method for θ (note that $T_{n,a}^L$ and $T_{n,a}^M$ are almost identical), the figure reveals a satisfactory fit, which is in agreement with the conclusions of Bowman and Shenton (2001) reached by a classical χ^2 test.

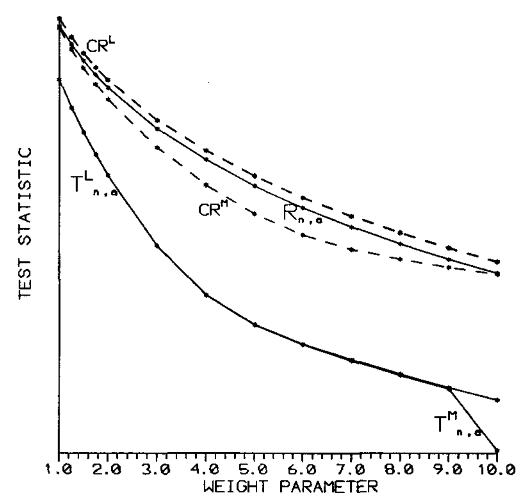

Fig. 1. Values of the test statistics and critical points for the fracture toughness and the mileage data.

The second set of data represents mileages for 19 military personnel carriers that failed in service. Based on a Kullback–Leibler information test, Ebrahimi *et al.* (1992) satisfactorily fitted an exponential distribution to this data set. As suggested by a referee, it is of interest to consider what kind of conclusions will be reached by an exponentiality test which results from our ELT-tests following a square root transformation of the data. After scaling the data by their sample mean, we have applied the proposed procedure to the square root of the mileage data. The results for the exponentiality statistic denoted by $R_{n,a}$ are shown in Fig. 1. The corresponding critical points, computed by simulation, essentially coincide with those of the $T^L_{n,a}$-test. We conclude, as Ebrahimi *et al.* (1992) did, that we can not reject the null hypothesis of exponentiality.

Acknowledgements

The authors wish to thank the referees for their constructive criticism.

References

Auinger, K. (1990). Quasi goodness of fit tests for lifetime distributions, *Metrika*, **37**, 97–116.

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, *Comm. Statist. Theory Methods*, **21**, 969–978.

Baringhaus, L., Gürtler, N. and Henze, N. (2000). Weighted integral test statistics and components of smooth tests of fit, *Austral. New Zeal. J. Statist.*, **42**, 179–192.

Bowman, K. O. and Shenton, L. R. (2001). Weibull distributions when the shape parameter is defined, *Comput. Statist. Data Anal.*, **36**, 299–310.

Castillo, J. D. and Puig, P. (1997). Testing departures from gamma, Rayleigh and truncated normal distributions, *Ann. Inst. Statist. Math.*, **49**, 255–269.

Clarke, A. S. and Shizgal, B. (1993). On the generation of orthogonal polynomials using asymptotic methods for recurrence coefficients, *J. Comput. Phys.*, **104**, 140–149.

Csörgő, S. and Teugels, J. (1990). Empirical Laplace transform and approximation of compound distributions, *J. Appl. Probab.*, **27**, 88–101.

Ebrahimi, N., Habibullah, M. and Soofi, E. S. (1992). Testing exponentiality based on Kullback–Leibler information, *J. Roy. Statist. Soc. Ser. B*, **54**, 739–748.

Edgeman, R. L. and Scott, R. C. (1987). Lilliefors's tests for transformed variables, *Brazilian Journal of Probability Statistics*, **1**, 101–112.

Feigin, P. D., Tweedie, R. L. and Belyea, C. (1983). Weighted area techniques for explicit parameter estimation in multi-stage models, *Austral. J. Statist.*, **25**, 1–16.

Gawronski, W. and Stadtmüller, U. (1985). Parameter estimation for distributions with regularly varying tails, *Statist. Decisions*, **3**, 297–316.

Henze, N. (1993). A new flexible class of omnibus tests for exponentiality, *Comm. Statist. Theory Methods*, **22**, 115–133.

Henze, N. and Meintanis, S. (2002*a*). Tests of fit for exponentiality based on the empirical Laplace transform, *Statistics*, **36**, 147–161.

Henze, N. and Meintanis, S. (2002*b*). Goodness-of-fit tests based on a new characterization of the exponential distribution, *Comm. Statist. Theory Methods*, **31**, 1479–1497.

Henze, N. and Wagner, Th. (1997). A new approach to the BHEP tests for multivariate normality, *J. Multivariate Anal.*, **62**, 1–23.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, Vol. **1**, Wiley, New York.

Laurence, A. F. and Morgan, B. J. T. (1987). Selection of the transformation variable in the Laplace transfom method of estimation, *Austral. J. Statist.*, **29**, 113–127.

Rayner, J. C. W. and Best, D. J. (1989). *Smooth Tests of Goodness of Fit*, Oxford University Press, New York.

Yao, Q. and Morgan, B. J. T. (1999). Empirical transform estimation for indexed stochastic models, *J. Roy. Statist. Soc. Ser. B*, **61**, 127–141.