Working Paper 07-56
Statistic and Econometric Series 14
July 2007

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34-91) 6249849

A SCALE-FREE ADAPTIVE STATISTIC FOR TESTING EXPONENTIALITY AGAINST WEIBULL AND GENERALIZED PARETO DISTRIBUTIONS

Aurea Grané[1] and Josep Fortiana[2]

Abstract

In Fortiana and Grané (2002) we study a scale-free statistic, based on Hoeffding's maximum correlation, for testing exponentiality. This statistic admits an expansion along a countable set of orthogonal axes, originating a sequence of statistics. Linear combinations of a given number p of terms in this sequence can be written as a quotient of L-statistics. In this paper we propose a scale-free adaptive statistic for testing exponentiality with optimal power against a specific alternative. We obtain its exact distribution and compare it with other scale-free statistics for testing exponentiality, such as the Stephens' modification of the Shapiro-Wilk statistic, the Gini statistic and the Q_n statistic defined in Fortiana and Grané (2002).

***Keywords**:* Goodness-of-fit, exponentiality, maximum correlation, decomposition of tests statistics, exact distribution, L-statistics.
***AMS subject classification**:* 62G10, 62G30, 62E15.

[1] Statistics Department, Universidad Carlos III de Madrid, C/ Madrid 126, 28903 Getafe (Madrid), Spain, e-mail: agrane@est-econ.uc3m.es Corresponding author.
[2] Dept. de Probabilitat, Lògica i Estadística, Universitat de Barcelona, C/ Gran Via de les Corts Catalanes, 585, 08007 Barcelona, Spain, e-mail: fortiana@ub.edu

Work partially supported by Spanish grant MTM2006-09920 (Ministry of Education and Science – FEDER)

A scale-free adaptive statistic for testing exponentiality against Weibull and Generalized Pareto distributions

A. Grané(1) J. Fortiana(2)

(1) Statistics Department. Universidad Carlos III de Madrid.

(2) Dpt. de Probabilitat, Lògica i Estadística. Universitat de Barcelona.

Abstract

In Fortiana and Grané (2002), we study a scale-free statistic, based on Hoeffding's maximum correlation, for testing exponentiality. This statistic admits an expansion along a countable set of orthogonal axes, originating a sequence of statistics. Linear combinations of a given number p of terms in this sequence can be written as a quotient of L-statistics. In this paper we propose a scale-free adaptive statistic for testing exponentiality with optimal power against a specific alternative. We obtain its exact distribution and compare it with other scale-free statistics for testing exponentiality, such as the Stephens' modification of the Shapiro-Wilk statistic, the Gini statistic and the Q_n statistic defined in Fortiana and Grané (2002).

Keywords: Goodness-of-fit, exponentiality, maximum correlation, decomposition of tests statistics, exact distribution, L-statistics.

AMS subject classification: 62G10, 62G30, 62E15.

1 Introduction

Most goodness-of-fit statistics can be interpreted as measures of proximity between two distributions: empirical and hypothesized. The family of tests we are concerned with in this paper is based on Hoeffding's maximum correlation between two probability distribution functions F_1 and F_2, with second order moments, which is equal to

$$\rho^+(F_1, F_2) = \frac{1}{\sigma_1\,\sigma_2}\left(\int_0^1 F_1^-(p)\,F_2^-(p)\,dp - \mu_1\,\mu_2\right), \tag{1}$$

Work supported in part by MTM2006-09920 (Ministry of Education and Science–FEDER)
Authors address: Statistics Department, Universidad Carlos III de Madrid, c/ Madrid, 126, 28903–Getafe, Spain.
E–mail: A. Grané, agrane@est-econ@uc3m.es; J. Fortiana, fortiana@ub.edu.
Date: July 9, 2007.

where F_i^- is the left-continuous pseudoinverse of F_i, $\mu_i = E(F_i)$ and $\sigma_i^2 = var(F_i)$, $i = 1, 2$. (see, e.g., Cambanis, Simons, and Stout 1976). Here the notation $E(F)$ represents the expected value of any random variable whose probability distribution function is F, and analogously for $var(F)$. Since $\rho^+(F_1, F_2)$ equals 1 if and only if $F_1 = F_2$ (almost everywhere) up to a scale and location change, the quantity $\rho^+(F_n, F_0)$ has been used in previous works, Fortiana and Grané (2002), (2003) and Grané and Fortiana (2006), as a qualitative measure of goodness-of-fit of an iid sample, with empirical distribution function F_n, to a given distribution F_0. More precisely, in Fortiana and Grané (2002) we defined and studied the properties of

$$Q_n = \frac{s_n}{\bar{y}_n}\, \rho^+(F_n, F_0), \tag{2}$$

where $\bar{y}_n$ and s_n^2 are, respectively, the empirical sample mean and variance, as a goodness-of-fit statistic for testing the composite hypothesis of exponentiality when the scale parameter is unknown, that is

$$H_0 : F = F_0 \equiv Exp(0, \beta), \quad \beta \in \mathbb{R}_+, \tag{3}$$

where $Exp(0, \beta)$ is the exponential distribution with location parameter $\alpha = 0$ and unknown scale parameter $\beta > 0$, $F_0(x) = 1 - \exp(-x/\beta)$, $x \geq 0$. We found that Q_n has reasonably good properties as a goodness-of-fit test and we also computed its exact distribution under the null hypothesis.
In the present work we prove the following identity

$$Q_n = \sum_{j \geq 0} \gamma_j\, \tilde{\Phi}_{nj}$$

where the sequence $\{\tilde{\Phi}_{nj}\}_{j \geq 0}$ appears as a decomposition of this test, analogous to those studied by Durbin and Knott (1972, 1975) and Stephens (1974). The statistics $\tilde{\Phi}_{nj}$'s are scale-free and they are obtained as a slight modification of the corresponding j-th Fourier coefficient of the pseudoinverse of F_n for the orthonormal (in $L^2[0, 1]$) sequence $\{\phi_j(t)\}_{j \geq 0}$, $t \in [0, 1]$ (see Proposition 2.1 and Proposition 2.2).
In this article, we seek to improve the performance of Q_n for a specific alternative by choosing a proper linear combination of a finite number p of terms of the scale-free sequence of statistics $\{\tilde{\Phi}_{nj}\}_{j \geq 0}$. We call this linear combination the scale-free adaptive statistic and we denote it by T_p.
Section 3 contains the construction of the T_p statistic for testing (3) against a specific alternative. Power optimization is translated into an eigenvalue-type problem with quadratic forms, functions of the first two moments of the order statistic. Since it is possible that for certain families of alternatives some of the required second order moments do not exist, in Section 4 we try to circumvent this problem using the asymptotic theory of L-statistics in order to obtain approximations of the quadratic forms involved in Section 3.

In Section 5 we obtain the exact distribution of T_p and, as an illustration, in Section 6 we perform the actual computations for the orthonormal basis of the Legendre polynomials, comparing the power of T_p with other scale-free statistics for testing exponentiality such as the Stephens' modification of the Shapiro-Wilk statistic, the Gini statistic and the Q_n statistic of (2).

2 Definition of the statistic

Consider n iid random variables with probability distribution function F, with finite second order moment, and empirical distribution function F_n. Let μ and σ^2 be the expectation and variance of F, respectively. Let $\{\phi_j(t)\}_{j\geq 0}$ be an orthonormal sequence in $L^2[0,1]$, $Exp(0,\beta)$ the exponential distribution with scale parameter $\beta > 0$ and location parameter $\alpha = 0$.

Proposition 2.1 *The Hoeffding maximum correlation has the following decomposition:*

$$\rho^+(F, Exp(0,\beta)) = \frac{1}{\sigma} \sum_{j\geq 0} \gamma_j \, \Phi_j(F),$$

where $\Phi_j(F) = \int_0^1 F^-(t)\, \phi_j(t)\, dt$ *and* $\gamma_j = \int_0^1 -(1+\log(1-t))\, \phi_j(t)\, dt$, *for* $j \geq 0$.

Proof: Noticing that $\mu = \int_0^1 F^-(t)\, dt$, formula (1) can be written as

$$\begin{aligned} \rho^+(F, Exp(0,\beta)) &= \frac{1}{\sigma\,\beta} \left(\int_0^1 -F^-(t)\, \beta \log(1-t)\, dt - \mu\, \beta \right) \\ &= \frac{1}{\sigma} \int_0^1 -(1+\log(1-t))\, F^-(t)\, dt \\ &= \frac{1}{\sigma} \sum_{j\geq 0} \gamma_j \int_0^1 F^-(t)\, \phi_j(t)\, dt, \end{aligned}$$

where we have expanded function $-(1+\log(1-t))$ in a Fourier series with respect to the orthonormal sequence $\{\phi_j(t)\}_{j\geq 0}$. □

Let $y_{(1)}, \ldots, y_{(n)}$ be the order statistic obtained from n iid$\sim Exp(0,\beta)$ random variables with empirical distribution function F_n. For $j \geq 0$ we define the sequence of L–statistics

$$\Phi_{nj} \equiv \Phi_j(F_n) = \int_0^1 F_n^-(t)\, \phi_j(t) dt = \frac{1}{n} \sum_{i=1}^{n} a_{ij}\, y_{(i)},$$

where $a_{ij} = n \int_{(i-1)/n}^{i/n} \phi_j(t)\, dt$, and the sequence of scale-free statistics is defined as $\tilde{\Phi}_{nj} = \Phi_{nj}/\bar{y}_n$.

Proposition 2.2 *With the previous notation, the statistic Q_n of (2) is scale-free and can be written as $Q_n = \sum_{j\geq 0} \gamma_j \, \tilde{\Phi}_{nj}$.*

Proof: Note that Q_n is written as a sequence of scale-free statistics, since each $\tilde{\Phi}_{nj}$ is a quotient of L-statistics. From Proposition 2.1 and using F_n instead of F, we have that

$$Q_n = \frac{s_n \, \rho^+(F_n, Exp(0,\beta))}{\bar{y}_n} = \frac{\sum_{j\geq 0} \gamma_j \, \Phi_{nj}}{\bar{y}_n} = \sum_{j\geq 0} \gamma_j \, \tilde{\Phi}_{nj}.$$

□

We define the scale-free adaptive statistic for testing exponentiality

$$T = \sum_{j\geq 0} \lambda_j \, \tilde{\Phi}_{nj} = \sum_{j\geq 0} \lambda_j \left(\frac{1}{n \, \bar{y}_n} \sum_{i=1}^{n} a_{ij} \, y_{(i)} \right) \tag{4}$$

where $\{\lambda_j\} \in \ell^1_{\mathbb{R}}$ is a sequence of real numbers.
In practice, given an alternative distribution F_1, we will use T_p, the result of truncating (4) at $j = p$, where parameters $\lambda_0, \ldots, \lambda_p$ will be determined in order to maximize power for testing $H_0 : F = Exp(0,\beta)$ with unknown scale parameter $\beta > 0$ vs. $H_1 : F = F_1$. Once coefficients λ_j's are known, it is possible to obtain the exact distribution of T_p under the null hypothesis (see Section 5) and hence to obtain the exact critical regions. It is also interesting to mention that for practical purposes T_p should be expressed directly in terms of the observed order statistic.

3 Optimization of the power function

Let $\mathbf{x} = (x_{(1)}, \ldots, x_{(n)})'$ be the order statistic of n iid$\sim$ $Exp(0,1)$ random variables and consider the following auxiliary parametric function

$$\mathcal{L}^r_{np} = \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{j=0}^{p} a_{ij} \, \lambda_j - r \right) x_{(i)}, \tag{5}$$

where $a_{ij} = n \int_{(i-1)/n}^{i/n} \phi_j(t) \, dt$, $j = 0, \ldots, p$, and r, $\lambda_0, \ldots, \lambda_p$ are real coefficients to be determined.
$\mathcal{L}^r_{np}$ is a linear combination of the order statistic and, in matrix notation, it can be written as $\mathcal{L}^r_{np} = \mathbf{L}' \, \mathbf{A}' \, \mathbf{x}$, where $\mathbf{L} = (r, \lambda_0, \ldots, \lambda_p)'$ and

$$\mathbf{A} = \frac{1}{n} \begin{pmatrix} -1 & a_{10} & \ldots & a_{1p} \\ -1 & a_{20} & \ldots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ -1 & a_{n0} & \ldots & a_{np} \end{pmatrix}.$$

Its expectation and variance are given by $\mu_i = \mathbf{L}'\,\mathbf{A}'\,\mathbf{m}_i$, $\sigma_i^2 = \mathbf{L}'\,\mathbf{A}'\,\mathbf{\Sigma}_i\,\mathbf{A}\,\mathbf{L}$, where $\mathbf{m}_i = E(\mathbf{x}|H_i)$, $\mathbf{\Sigma}_i = Var(\mathbf{x}|H_i)$, $i = 0, 1$, and applying the general theory of L-statistics, it is asymptotically normal distributed (see, for example, Shorack and Wellner 1986).
In order to find the coefficients $\lambda_0, \ldots, \lambda_p$ such that the statistic $T_p = \sum_{j=0}^{p} \lambda_j\,\tilde{\Phi}_{nj}$ has maximum power to test $H_0 : F = Exp(0, \beta)$ with unknown $\beta > 0$ vs. $H_1 : F = F_1$, we impose the following conditions, for a given significance level $\varepsilon \in (0, 1)$:

$$P(T_p > r|H_0) = \varepsilon, \quad P(T_p > r|H_1) \text{ is maximum.}$$

Using the auxiliary parametric function introduced in (5) these conditions are written:

$$P(\mathcal{L}_{np}^r > 0|H_0) = \varepsilon \quad \Rightarrow \quad \mathbf{L}'\,\mathbf{A}'\,\mathbf{m}_0 = -c_\varepsilon(\mathbf{L}'\,\mathbf{A}'\,\mathbf{\Sigma}_0\,\mathbf{A}\,\mathbf{L})^{1/2}, \qquad (6)$$

where c_ε is the $(1 - \varepsilon) \times 100$ percentile of the standard normal distribution, and the probability to be maximized can be asymptotically approximated by

$$\Psi(\mathbf{L}) = 1 - F_Z\left(\frac{-\mathbf{L}'\,\mathbf{A}'\,\mathbf{m}_1}{(\mathbf{L}'\,\mathbf{A}'\,\mathbf{\Sigma}_1\,\mathbf{A}\,\mathbf{L})^{1/2}}\right) \qquad (7)$$

where F_Z is the standard normal distribution function. To maximize the asymptotic power function is equivalent to find the extremes of the following quotient of quadratic forms:

$$\frac{\mathbf{L}'\,\mathbf{A}'\,\mathbf{m}_1\,\mathbf{m}_1'\,\mathbf{A}\,\mathbf{L}}{\mathbf{L}'\,\mathbf{A}'\,\mathbf{\Sigma}_1\,\mathbf{A}\,\mathbf{L}}, \qquad (8)$$

which is also equivalent to find the extremes of $\mathbf{L}'\,\mathbf{A}'\,\mathbf{m}_1\,\mathbf{m}_1'\,\mathbf{A}\,\mathbf{L}$ constrained to $\mathbf{L}'\,\mathbf{A}'\,\mathbf{\Sigma}_1\,\mathbf{A}\,\mathbf{L} = 1$. Since the numerator of (8) is a matrix of rank one, the solution of (7) is the (unique with non-null eigenvalue) eigenvector of the generalized eigenvalue problem: $\mathbf{A}'\,\mathbf{m}_1\,\mathbf{m}_1'\,\mathbf{A}\,\mathbf{L} = \xi\,\mathbf{A}'\,\mathbf{\Sigma}_1\,\mathbf{A}\,\mathbf{L}$, normalized so that $\mathbf{L}'\,\mathbf{A}'\,\mathbf{\Sigma}_1\,\mathbf{A}\,\mathbf{L} = 1$, whenever $\mathbf{A}'\,\mathbf{\Sigma}_1\,\mathbf{A}$ is positive defined, that is:

$$\mathbf{L} = (\mathbf{A}'\,\mathbf{\Sigma}_1\,\mathbf{A})^{-1}\,\mathbf{A}'\,\mathbf{m}_1, \quad \xi = \mathbf{m}_1'\,\mathbf{A}\,(\mathbf{A}'\,\mathbf{\Sigma}_1\,\mathbf{A})^{-1}\,\mathbf{A}'\,\mathbf{m}_1.$$

Observe that when the first element of the orthonormal basis $\{\phi_j(t)\}_{j\geq 0}$ is equal to 1, that is $\phi_0(t) = 1$, then $\mathbf{A}'\,\mathbf{\Sigma}_1\,\mathbf{A}$ will be positive semidefined, since the first two columns of matrix $\mathbf{A}$ will be proportional. In this case, the solution of (7) is

$$\mathbf{L} = \mathbf{B}^-\,\mathbf{A}'\,\mathbf{m}_1 + \mathbf{N}\,\mathbf{h}, \quad \xi = \mathbf{m}_1'\,\mathbf{A}\,\mathbf{B}^-\,\mathbf{A}'\,\mathbf{m}_1,$$

where $\mathbf{B}^-$ is the Moore-Penrose pseudoinverse of $\mathbf{B} = \mathbf{A}'\,\mathbf{\Sigma}_1\,\mathbf{A}$, $\mathbf{N} = \mathbf{I} - \mathbf{B}\,\mathbf{B}^-$ and $\mathbf{h}$ is a $(p+2) \times 1$ arbitrary vector (see McDonald, Torii, and Nishisato 1979). Condition (6) helps to determine vector $\mathbf{h}$.
Note that to solve this problem it is necessary that all the first and second order moments of the order statistic involved do exist. This could be not the case when the expectation of the distribution does not exist due to singularities in 0 or 1 (see David 1981).

4 Generic alternatives

In this section we find approximations of the expectation and variance (under H_1) of the parametric function defined in (5), in order to compute the optimal vector $\mathbf{L}$ of Section 3. We will suppose that the alternative probability distribution function F has a pseudoinverse of the form:

$$F^{-}(t) = \sum_{k=0}^{q} \gamma_k \, \psi_k(t), \tag{9}$$

where γ_k are real numbers and $\{\psi_k(t)\}_{k\geq 0}$ is an orthonormal sequence in $L^2[0,1]$, possibly different from $\{\phi_j(t)\}_{j\geq 0}$.
Given an arbitrary F the first q Fourier terms of F^{-} yield such an expression. In the present context this is more natural than expanding F or the probability density function, since the moments of the order statistic can be advantageously expressed in terms of F^{-}, e.g.,

$$\begin{aligned} E(x_{(i)}|H_1) &= i\binom{n}{i}\int_0^1 F^{-}(t)\, t^{i-1}\,(1-t)^{n-i}\,dt \\ &= i\binom{n}{i}\sum_{k=0}^{q}\gamma_k\int_0^1 \psi_k(t)\, t^{i-1}\,(1-t)^{n-i}\,dt. \end{aligned} \tag{10}$$

To solve (7) we must determine $\mathbf{m}_1$ and $\mathbf{\Sigma}_1$. Formula (10) gives the entries in $\mathbf{m}_1$, but for several reasons, an exact $\mathbf{\Sigma}_1$ could not be available. Instead we can determine $\mathbf{A}'\mathbf{\Sigma}_1\mathbf{A}$ from the asymptotic approximation given in the following proposition.

Proposition 4.1 *Let $\{\phi_j(t)\}_{j\geq 0}$ and $\{\psi_k(t)\}_{k\geq 0}$ be two orthonormal systems $L^2[0,1]$ such that $\phi_0(t) = \psi_0(t) = 1$, $\mathcal{L}^r_{np}$ the parametric function defined in (5) and $(\ell_1, \ell_2, \ldots, \ell_{p+2}) = (r, \lambda_0, \ldots, \lambda_p)$. We have the following convergences in law*

$$\sqrt{n}[\mathcal{L}^r_{np} - \mu_{\mathcal{L}}] \xrightarrow[n\to\infty]{\mathcal{L}} N(0, \sigma^2_{\mathcal{L}}), \tag{11}$$

$$\sqrt{n}\frac{[\mathcal{L}^r_{np} - \mu_{\mathcal{L}}]}{\sigma_n} \xrightarrow[n\to\infty]{\mathcal{L}} N(0, 1), \tag{12}$$

where

$$\mu_{\mathcal{L}} = (-\ell_1 + \ell_2)\,\gamma_0 + \sum_{j=1}^{p}\sum_{k=1}^{q}\ell_{j+2}\,\gamma_k\int_0^1 \phi_j(t)\,\psi_k(t)\,dt, \tag{13}$$

$$\sigma_n^2 = \frac{1}{n}\sum_{j=1}^{p+2}\sum_{l=1}^{p+2}\ell_j\,\ell_l\,\sigma_{n,jl}, \quad \sigma_{n,jl} = \sum_{k=0}^{q}\sum_{m=0}^{q}\gamma_k\,\gamma_m\,I_{jklm}, \tag{14}$$

where

$$I_{jklm} = \begin{cases} \int_0^1 \int_0^1 K(s,t)\, \psi_k'(s)\, \psi_m'(t)\, ds\, dt, & j = l = 1, \\ \int_0^1 \int_0^1 K(s,t)\, \phi_{l-2}(t)\, \psi_k'(s)\, \psi_m'(t)\, ds\, dt, & j = 1, l \geq 2, \\ \int_0^1 \int_0^1 K(s,t)\, \phi_{j-2}(t)\, \psi_k'(s)\, \psi_m'(t)\, ds\, dt, & l = 1, j \geq 2, \\ \int_0^1 \int_0^1 K(s,t)\, \phi_{j-2}(s)\, \psi_k'(s)\, \phi_{l-2}(t)\, \psi_m'(t)\, ds\, dt, & j \geq 2, l \geq 2, \end{cases}$$

with $K(s,t) = \min(s,t) - s\,t$ *and* $\psi_k'(t)$ *denotes the derivative of* $\psi_k(t)$.

Proof: The parametric function of (5) can be written as

$$\mathcal{L}_{np}^r = \frac{1}{n} \sum_{i=1}^n J(i/n)\, x_{(i)}$$

where

$$J(i/n) = \sum_{j=0}^p \lambda_j\, a_{ij} - r = -B_0(i/n)\, \ell_1 + \sum_{j=0}^p B_j(i/n)\, \ell_{j+2},$$

where $(\ell_1, \ell_2, \ldots, \ell_{p+2}) = (r, \lambda_0, \ldots, \lambda_p)$ and

$$B_j(i/n) = n \int_{(i-1)/n}^{i/n} \phi_j(t)\, dt = \frac{b_j(i/n) - b_j((i-1)/n)}{1/n}, \quad j = 0, \ldots, p,$$

verifying that B_j tends to ϕ_j, as n tends to infinity. We can use the asymptotic approximation

$$J(t) \approx -\phi_0(t)\, \ell_1 + \sum_{j=0}^p \phi_j(t)\, \ell_{j+2}, \quad t \in (0,1).$$

Since $J(t)$ is a continuous and bounded a.s. (F^-) function, we can compute the asymptotic expectation of $\mathcal{L}_{np}^r$ under H_1 as

$$\begin{aligned} \mu_{\mathcal{L}} &= \int_0^1 J(t)\, F^-(t)\, dt = \int_0^1 \left(-\phi_0(t)\, \ell_1 + \sum_{j=0}^p \phi_j(t)\, \ell_{j+2} \right) \sum_{k=0}^q \gamma_k\, \psi_k(t)\, dt \\ &= (-\ell_1 + \ell_2)\, \gamma_0 + \sum_{j=1}^p \sum_{k=1}^q \ell_{j+2}\, \gamma_k \int_0^1 \phi_j(t)\, \psi_k(t)\, dt, \end{aligned}$$

where we have used that $\phi_0(t) = \psi_0(t) = 1$ and $\int_0^1 \phi_j(t)\, dt = \int_0^1 \psi_k(t)\, dt = 0$ for $j \geq 1$ and $k \geq 1$. The asymptotic variance is given by

$$\sigma_{\mathcal{L}}^2 = \int_0^1 \int_0^1 J(s)\, J(t)\, K(s,t)\, dF^-(s)\, dF^-(t),$$

where $K(s,t) = \min(s,t) - st$ (see, for example, Shorack and Wellner 1986). Formula (14) is obtained substituting the expressions for function J and for the derivative of F^- and proceeding analogously as for the computation of the expectation.
The convergence of (11) and (12) are obtained applying the general theory of L-statistics described in Shorack and Wellner (1986). □
Comparing the expression for $\sigma_1^2 = \mathbf{L}'\,\mathbf{A}'\,\mathbf{\Sigma}_1\,\mathbf{A}\,\mathbf{L}$ of Section 3 with (14), we see that the entries in $\mathbf{A}'\,\mathbf{\Sigma}_1\,\mathbf{A}$ are the limit $\sigma_{jl} = \lim_{n\to\infty} \sigma_{n,jl}$, but some computational examples suggest that a better approximation is obtained with $\sigma_{n,jl}$.
As a final comment to this section, note that the orthonormal system $\{\psi_k\}_{k\geq 0}$ should be chosen carefully so that a good approximation for F^- is guaranteed.

5 Exact distribution of T_p under H_0

Proposition 5.1 *Given $\lambda_0, \ldots, \lambda_p \in \mathbb{R}$, the statistic $T_p = \sum_{j=0}^{p} \lambda_j\, \tilde{\Phi}_{nj}$ does not depend on the scale parameter, and can be written as*

$$T_p = \frac{\sum_{i=1}^{n} c_i\, z_i}{\sum_{i=1}^{n} z_i},$$

where $c_i = \sum_{k=i}^{n} \frac{1}{n-i+1} \sum_{j=0}^{p} a_{kj}\, \lambda_j$, $i = 1, \ldots, n$, $a_{kj} = n \int_{(k-1)/n}^{k/n} \phi_j(t)\, dt$, $j = 1, \ldots, p$ and $z_1, \ldots, z_n$ are iid$\sim$ $Exp(0,1)$ random variables.

Proof: Suppose $x_j = y_j/\beta$, $j = 1, \ldots, n$, then after truncating formula (4) at $j = p$ we get:

$$T_p = \sum_{j=0}^{p} \lambda_j \left(\frac{\sum_{i=1}^{n} a_{ij}\, x_{(i)}}{\sum_{i=1}^{n} x_{(i)}} \right) = \frac{\sum_{i=1}^{n} b_i\, x_{(i)}}{\sum_{i=1}^{n} x_{(i)}}, \tag{15}$$

where $b_i = \sum_{j=0}^{p} \lambda_j\, a_{ij}$, $i = 1, \ldots, n$. From David (1981), for $i = 1, \ldots, n$, the transformation $z_i = (n - i + 1)(x_{(i)} - x_{(i-1)})$, where, by convention, $x_{(0)} = 0$, gives n iid$\sim$ $Exp(0,1)$ random variables, and the i–th order statistic from the standard distribution can be written as

$$x_{(i)} = \sum_{k=1}^{i} \frac{z_k}{n-k+1} \tag{16}$$

Substituting (16) in (15),

$$T_p = \frac{\sum_{i=1}^{n} \sum_{k=1}^{i} b_i\, \frac{z_k}{(n-k+1)}}{\sum_{i=1}^{n} \sum_{k=1}^{i} \frac{z_k}{(n-k+1)}} = \frac{\sum_{i=1}^{n} \frac{z_i}{(n-i+1)} \sum_{k=i}^{n} b_k}{\sum_{i=1}^{n} z_i} = \frac{\sum_{i=1}^{n} c_i\, z_i}{\sum_{i=1}^{n} z_i},$$

where $c_i = \sum_{k=i}^{n} \frac{b_k}{n-i+1} = \sum_{k=i}^{n} \left(\frac{1}{n-i+1} \sum_{j=0}^{p} a_{kj}\, \lambda_j \right)$, $i = 1, \ldots, n$. □
From now on we will suppose that coefficients c_i are non-null and that $c_i \neq c_j$ for all $i \neq j$, in order to simplify notation and also because this is exactly what happens in the applications considered in Section 6. Let $N = \sum_{i=1}^{n} c_i\, z_i$ and $D = \sum_{i=1}^{n} z_i$ so that $T_p = N/D$.

Lemma 5.1 *The characteristic function of the vector* (N, D) *is*

$$\varphi_{(N,D)}(t_1, t_2) = \prod_{j=1}^{n} \left(1 - i(c_j\, t_1 + t_2)\right)^{-1},$$

with $(t_1, t_2) \in \mathbb{R}^2$, $i = \sqrt{-1}$.

Proof: In matrix notation, $N = \mathbf{c}'\, \mathbf{z}$, $D = \mathbf{1}'\, \mathbf{z}$, where $\mathbf{c} = (c_1, \ldots, c_n)'$, $\mathbf{z} = (z_1, \ldots, z_n)'$ and $\mathbf{1} = (1, \ldots, 1)'$. We also consider the matrix $\mathbf{C} = (\mathbf{c}, \mathbf{1})'$. Then,

$$\begin{pmatrix} N \\ D \end{pmatrix} = \mathbf{C}\, \mathbf{z}.$$

From the transformation formula of characteristic functions by the action of an affine map, we have that $\varphi_{\mathbf{Cz}}(\mathbf{t}) = \varphi_{\mathbf{z}}(\mathbf{C}'\, \mathbf{t})$, for $\mathbf{t} = (t_1, t_2)' \in \mathbb{R}^2$. Since $z_1, \ldots, z_n$ are iid$\sim$ $Exp(0, 1)$ random variables,

$$\varphi_{\mathbf{z}}(\mathbf{C}'\, \mathbf{t}) = \prod_{j=1}^{n} \varphi_{z_j}((c_j\, t_1, t_2)) = \prod_{j=1}^{n} \left(1 - i(c_j\, t_1 + t_2)\right)^{-1}.$$

□
Since T_p is a quotient of linear combinations of iid$\sim$ $Exp(0, 1)$ random variables, to obtain its exact probability density function we can adapt the technique used in Dwass (1961) and Matsunawa (1985).

Proposition 5.2 *The exact probability density function of* T_p *is given by*

$$f(t) = (n-1) \sum_{l=1}^{n} sgn(c_l) \prod_{k=1, k\neq l}^{n} (c_l - c_k)^{-1}\, (c_l - t)^{n-2}\, \chi(t/c_l)\, \chi(1 - t/c_l),$$

where $\chi(s)$ *is the indicator of the interval* $[s > 0]$.

Proof: using Lemma 2.2 from Matsunawa (1985), we obtain the characteristic function of (N, D):

$$\varphi_{(N,D)}(t_1, t_2) = \prod_{j=1}^{n} c_j^{-1} \sum_{l=1}^{n} e_l\, c_l\, \left((1 - i\, t_2) - i\, c_l\, t_1\right)^{-1} (1 - i\, t_2)^{1-n},$$

where

$$e_l = \prod_{k=1, k\neq l}^{n} \left(\frac{1}{c_k} - \frac{1}{c_l} \right)^{-1}.$$

Inverting this characteristic function, following formula 2.1 in Matsunawa (1985):

$$f_{(N,D)}(w,s) = \sum_{l=1}^{n} e_l \left(\prod_{j=1}^{n} c_j^{-1}\right) \frac{\exp(-s)}{\Gamma(n-1)} \left(s - \frac{w}{c_l}\right)^{n-2} \chi\left(\frac{w}{c_l}\right) \chi\left(s - \frac{w}{c_l}\right),$$

where $\chi(s)$ is the indicator of the interval $[s > 0]$, for $(w,s) \in \mathbb{R}^2$. Applying a change of variables and noticing that

$$e_l \left(\prod_{j=1}^{n} c_j^{-1}\right) = c_l^{n-2} \prod_{k=1, k\neq l} (c_l - c_k)^{-1},$$

we get the probability density function of T_p, $f(t) = \int_{-\infty}^{+\infty} |s| \, f_{(N,D)}(ts, s)\, ds$.
□
We have developed a *Mathematica* program implementing this algorithm, which also computes the exact critical values for a given significance level.

6 Power of T_p

We have considered families of the form $F(x;\theta_1,\theta_2)$. In each case, one of the θ_1, θ_2 is the scale parameter and can be fixed to a given arbitrary value, without loss of generality.
A1. The first family of alternatives is the Weibull distribution with shape parameter α and scale parameter β, whose density function is:

$$f(x;\alpha,\beta) = \frac{\alpha}{\beta^\alpha}\, x^{\alpha-1} \exp\{-(x/\beta)^\alpha\}, \quad \alpha,\beta > 0,\ x \geq 0.$$

Note that $\alpha = 1$ corresponds to H_0 and that the Weibull distribution is more similar to the exponential distribution when $\alpha < 1$.
A2. The second family of alternatives is the Generalized Pareto, whose distribution function is:

$$F(x;a,k) = 1 - \left(1 - \frac{k}{a}\, x\right)^{1/k}, \quad a > 0,$$

with $0 \leq x < \infty$, if $k \leq 0$, and $0 \leq x \leq a/k$ if $k > 0$. Observe that $\lim_{k\to 0} F(x;a,k) = Exp(0,a)$, and that the Generalized Pareto distribution is similar to the exponential distribution only when $k < 1$.
As examples of construction of the test statistic T_p and motivated by the reasons explained above, we have considered the A1 and A2 alternatives for some specific values of the parameters, for which it is also possible to compute the exact values of $\mathbf{m}_1$ and $\mathbf{\Sigma}_1$ of Section 3. For a sample size of $n = 10$ and a 5% significance level we have determined coefficients $\lambda_0, \ldots, \lambda_p$ for

$p = 4$ using the orthonormal basis obtained from the sequence of Legendre polynomials $\phi_0(t) = 1$, $\phi_1(t) = \sqrt{3}(2t-1)$, and recurrence relation:

$$\phi_{j+1}(t) = \frac{\sqrt{(2j+3)(2j+1)}}{j+1}(2t-1)\,\phi_j(t) - \frac{\sqrt{2j+3}}{\sqrt{2j-1}}\frac{j}{j+1}\,\phi_{j-1}(t), \quad j \geq 1.$$

Applying Proposition 5.2 we have computed the exact critical region of the test. Table 1 contains the power comparisons of the (unilateral) test of exponentiality at the 5%-significance level based on T_p with those based on other scale-free statistics such as the Q_n statistic defined in Fortiana and Grané (2002):

$$Q_n = \frac{\sum_{i=1}^n l_i\, x_{(i)}}{n\,\overline{x}},$$

where $l_i = (n-i)\log(n-i) - (n-i+1)\log(n-i+1) + \log(n)$, $i = 1,\ldots,n$ and $0\log 0 \equiv 0$, the Stephens' modification of the Shapiro–Wilk statistic (Stephens 1978):

$$W_s = \frac{(\sum_{i=1}^n x_i)^2}{n(n+1)\sum_{i=1}^n x_i^2 - n(\sum_{i=1}^n x_i)^2}$$

and the Gini statistic (Gail and Gastwirth 1978):

$$G = \frac{\sum_{i=1}^n (2i-n-1)x_{(i)}}{n(n-1)\overline{x}}.$$

These powers have been estimated from $N = 1000$ simulated samples of size $n = 10$ as the relative frequency of values of the corresponding statistic in the critical region. As a general comment, it can be said that T_p improves,

Table 1: Power comparisons for A1 and A2 families

A1 family ($\beta = 1$)				
α	T_p	Q_n	W_S	G
0.25	0.998	0.955	0.956	0.980
0.5	0.789	0.641	0.653	0.768
2	0.735	0.715	0.765	0.755
3	0.993	0.987	0.991	0.994

A2 family ($a = 1$)				
k	T_p	Q_n	W_S	G
-0.2	0.161	0.154	0.126	0.152
0.25	0.126	0.111	0.115	0.109
0.5	0.235	0.207	0.211	0.208

in general, the performance of the statistics compared, but its principal drawback is that its coefficients $(\lambda_0,\ldots,\lambda_p)$ should be computed anew for each parameter value. Although this procedure can be made automatically, we would like to find a set of coefficients that make T_p optimal enough for very practical purposes.

Practical issues

For each family of alternative distributions, A1 and A2, we have constructed the T_p statistic for several parameter values and we have selected the set of coefficients $(\lambda_0, \ldots, \lambda_p)$ for which T_p was globally more powerful. These selected sets, which are shown in Table 2, correspond to values of $\alpha = 0.5$ and $\alpha = 1.5$ for A1 family and to $k = -0.2$ and $k = 0.5$ for A2 family. Table 2 also contains the range of the parameter of the alternative distribution for which the test based on T_p can be considered as "globally optimal" (from a practical viewpoint) and the critical region of the test.

Table 2: Coefficients and critical values of the T_p statistic constructed for $n = 10$ and $p = 4$.

A1 family	T_p coefficients					c. region
$0 < \alpha < 1$	0.250882	-0.185657	0.065586	-0.025221	0.002454	$(-\infty, 0.091024)$
$\alpha > 1$	0.435419	0.140576	0.016431	0.0181301	0.010238	$(-\infty, 0.511821)$
A2 family	**T_p coefficients**					**c. region**
$k < 0$	0.035870	0.043223	0.024558	0.006038	-0.00018	$(0.059948, +\infty)$
$0 < k < 1$	0.079384	0.102235	0.068476	0.025487	0.00407	$(-\infty, 0.136742)$

As it was mentioned in Section 2, in a practical situation T_p should be expressed directly in terms of the observed order statistic, for example using formula (15). Figure 1 contains the different vectors of weights of the order statistic, associated to the adaptive statistics described in Table 2, which exactly correspond to the b_i's in the numerator of formula (15). We have represented them by solid and dotted lines (and not points) just for better comparison.

Figure 1: Vectors of weights for the observed order statistic associated to the adaptive statistics described in Table 2.

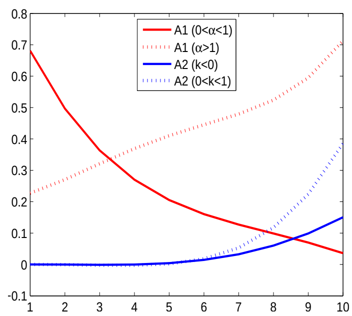
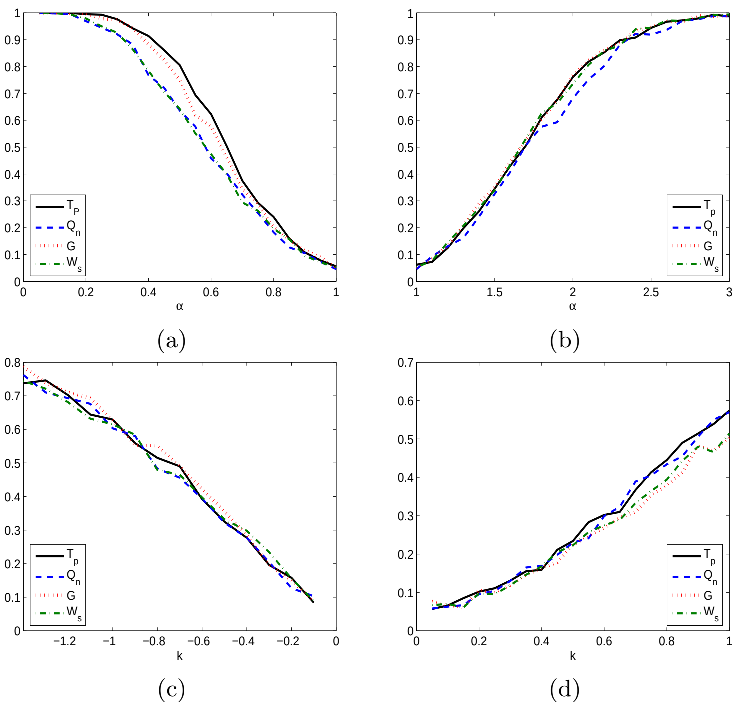


Figure 2 contains the power functions obtained using the T_p statistic described in Table 2, and the comparisons with the corresponding power functions of the tests based on Q_n, G and W_s. Power against an alternative distribution has been estimated by the relative frequency of values of the corresponding statistic in the critical region for $N = 1000$ simulated samples of size $n = 10$. For each family we have taken 40 different values of the parameter. As a general comment, it can be said that the adaptive statistics perform better for the A1 family when $\alpha < 1$ and for the A2 family when $0 < k < 1$. In the other two cases, the behavior of the statistics compared is rather similar.

Figure 2: Comparison of the tests based on T_p, Q_n, G and W_s for (a)-(b) the A1 family and (c)-(d) the A2 family.



References

Cambanis, S., G. Simons, and W. Stout (1976). Inequalities for $\mathcal{E}\, k(x, y)$ when the marginals are fixed. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 36*, 285–294.

David, H. A. (1981). *Order statistics (2nd ed.)*. New York: John Willey & Sons, Inc.

Durbin, J. and M. Knott (1972). Components of Cramér–Von Mises statistics. I. *Journal of the Royal Statistical Society B 34*, 290–307.

Durbin, J. and M. Knott (1975). Components of Cramér–Von Mises statistics. II. *Journal of the Royal Statistical Society B 37*, 216–237.

Dwass, M. (1961). The distribution of linear combinations of random divisions of an interval. *Trabajos de Estadística e Investigación Operativa 12*, 11–17.

Fortiana, J. and A. Grané (2002). A scale-free goodness-of-fit statistic for the exponential distribution based on maximum correlations. *Journal of Statistical Planning and Inference 108*, 85–97.

Fortiana, J. and A. Grané (2003). Goodness–of–fit tests based on maximum correlations and their orthogonal decompositions. *Journal of the Royal Statistical Society B 65 (1)*, 115–126.

Gail, M. and J. Gastwirth (1978). A scale–free goodness–of–fit test for the exponential distribution based on the Gini statistic. *Journal of the Royal Statistical Society B 40 (3)*, 350–357.

Grané, A. and J. Fortiana (2006). An adaptive goodness–of–fit test. *Communications in Statistics 35 (6)*, 1141–1155.

Matsunawa, T. (1985). The exact and approximate distributions of linear combinations of selected order statistics from a uniform distribution. *Annals of the Institute of Statistical Mathematics 37*, 1–16.

McDonald, R. P., Y. Torii, and S. Nishisato (1979). Some results on proper eigenvalues and eigenvectors with applications to scaling. *Psychometrika 44*(2), 211–227.

Shorack, G. R. and J. A. Wellner (1986). *Empirical processes with applications to statistics*. New York: John Wiley & Sons.

Stephens, M. A. (1974). Components of goodness–of–fit statistics. *Annales de l'Institut Henri Poincaré, Section B 10*, 37–54.

Stephens, M. A. (1978). On the W test for exponentiality with origin known. *Technometrics 20 (1)*, 33–35.