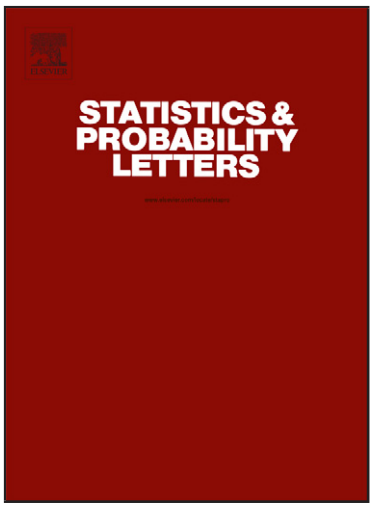
Bootstrap goodness–of–fit tests with estimated parameters based on empirical transforms

Simos Meintanis, Jan Swanepoel

Author's Accepted Manuscript



www.elsevier.com/locate/stapro

Bootstrap goodness–of–fit tests with estimated parameters based on empirical transforms

Simos Meintanis, Jan Swanepoel

PII: S0167-7152(07)00040-5
DOI: doi:10.1016/j.spl.2007.01.014
Reference: STAPRO 4573

To appear in: *Statistics & Probability Letters*

Received date: 15 December 2006
Accepted date: 17 January 2007

Cite this article as: Simos Meintanis and Jan Swanepoel, Bootstrap goodness–of–fit tests with estimated parameters based on empirical transforms, *Statistics & Probability Letters* (2007), doi:10.1016/j.spl.2007.01.014

BOOTSTRAP GOODNESS–OF–FIT TESTS WITH ESTIMATED PARAMETERS BASED ON EMPIRICAL TRANSFORMS [1]

Simos Meintanis, Jan Swanepoel

Department of Economics, National and Kapodistrian University of Athens
8 Pesmazoglou Street, 105 59 Athens, Greece
and
Department of Statistics, North–West University, Potchefstroom
South Africa

Abstract. Several test statistics have been proposed recently which employ a weighted distance that depends on an empirical transform, as well as on estimated parameters. The empirical characteristic function is a typical example, but alternative empirical transforms have also been employed, such as the empirical Laplace transform when dealing with non–negative random variables or the empirical probability generating function corresponding to discrete observations. We propose a general formulation that covers most of the transform–based test statistics which have appeared in the literature. Under this formulation, the asymptotic properties of the test statistics, such as the limiting null distribution and the consistency under general alternatives, are derived. Since large–sample critical values are extremely complicated (if not impossible) to compute, two effective bootstrap versions of the test procedures are derived, which can be used to approximate the critical values, for any given sample size, and to calculate the power under contiguous alternatives. The validity of these bootstrap procedures is shown analytically.

Keywords. Empirical characteristic function, Empirical Laplace transform, Empirical probability generating function, Asymptotic distribution, Consistency.

AMS 2000 classification numbers: 62G10, 62G20

[1] Proposed running head: BOOTSTRAP GOODNESS–OF–FIT TESTS

1 Introduction

Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (i.i.d.) random variables, with unspecified distribution function (DF) $F(x) = \mathrm{P}(X_1 \leq x)$. It is often the case that a DF depends on a parameter–vector ϑ. Then it will be denoted by F_ϑ. If ϑ is unknown, a parametric class $\mathcal{F}_\Theta = \{F_\vartheta : \vartheta \in \Theta\}$ of distributions is formed over $\vartheta \in \Theta$, where Θ is an open subset of $\mathbb{R}^k$. The Glivenko–Cantelli theorem asserts that $\sup_x |F_n(x) - F_\vartheta(x)| \to 0$ almost surely (a.s.) as $n \to \infty$, where $F_n(x)$ is the empirical DF. Hence, classical consistent tests for the composite null hypothesis,

$$H_0 : F \in \mathcal{F}_\Theta, \text{ for some } \vartheta \in \Theta,$$

are derived by defining appropriate test statistics in terms of certain distance measures between F_n and $F_{\hat{\vartheta}_n}$, where $\hat{\vartheta}_n$ denotes a consistent estimator of ϑ. This gives rise to well-known tests, such as the Kolmogorov–Smirnov, the Cramér–von Mises and the Anderson–Darling tests (e.g., see D'Agostino and Stephens, 1986).

An alternative way to tackle the composite goodness–of–fit problem is to consider a transform of the DF of the type

$$K(t) = \int_{-\infty}^{\infty} k(t,x) dF(x),$$

and the empirical version

$$K_n(t) = \int_{-\infty}^{\infty} k(t,x) dF_n(x).$$

The function $k(t,x)$ will be referred to as the kernel of the transform. Because of the Glivenko–Cantelli theorem,

$$\sup_t |K_n(t) - K(t)| \to 0, \text{ almost surely,}$$

where t belongs to a subinterval of the interval of convergence of $K(\cdot)$. Hence, assuming that the transform $K(\cdot)$ uniquely determines the law of X_1, a goodness–of–fit approach utilizing the distance $|K_n(t) - K(t)|$ over such a subinterval, will typically yield consistent tests for the null hypothesis H_0. A standard transform is the characteristic function (CF) and the empirical CF for $k(t,x) = \exp(itx)$. The empirical CF is historically the first type of transform employed in testing problems by Heathcote (1972). While the empirical CF is appropriate for arbitrary random variables, the Laplace transform (LT) and the empirical LT for $k(t,x) = \exp(-tx)$, are most appropriate when $X_1 \geq 0$. Again when dealing with count variables, the choice $k(t,x) = t^x$ results in the probability generating function (PGF) and the empirical PGF. The

weak convergence of the process $\sqrt{n}(K_n(\cdot) - K(\cdot))$ is of utmost importance for statistical inference. For the empirical CF, weak convergence has been established in the fundamental paper of Csörgő (1981), for the empirical moment generating function (and the empirical LT) by Feuerverger (1989), while Rémillard and Theodorescu (2000) treat the case of the empirical PGF.

The rest of the paper is organized as follows. In Section 2 a general formulation is provided for the transform–based test statistics, and some cases of special interest are reported as examples. In Section 3 we study the limiting distribution of the test statistic, both under the null and alternative hypotheses, and establish its consistency against general alternatives. In Section 4 effective bootstrap procedures are proposed to calculate approximate critical values and to help in the computation of power under contiguous alternatives.

2 Goodness–of–fit tests

Let $X_1, X_2, \ldots, X_n$, be independent observations with DF, $F(x)$. A general transform–based test statistic admits the representation

$$T_{n,\beta}(\vartheta) = \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{n}} \sum_{j=1}^{n} g(Y_j, \vartheta_2; t) \right)^2 \beta(t) dt, \tag{2.1}$$

where β denotes a finite measure (often termed weight function), and $Y_j = h(X_j, \vartheta_1),\ j = 1, 2, ..., n$, with the parameter vector $\vartheta = (\vartheta_1, \vartheta_2)^T$ having being decomposed into two sub-vectors ϑ_1 and ϑ_2. For each specific empirical transform, the function g contains elements of the corresponding kernel, and we will write $g(x; t)$ instead of $g(x, \vartheta_2; t)$, whenever there is no parameter ϑ_2 involved.

Examples

1. Symmetric location–scale families on the real line: Let $\vartheta = (\vartheta_1, \vartheta_2)^T$ with $\vartheta_1 = (\theta_1, \theta_2)^T \in \mathbb{R} \times \mathbb{R}^+$, and assume that the DF of X_1 may be written as $F(x; \vartheta) = G((x - \theta_1)/\theta_2)$, for some function G. Then, if $Y_1 = h(X_1, \vartheta_1)$, with $h(x, \vartheta_1) = (x - \theta_1)/\theta_2$, the CF of Y_1 coincides with the CF of X_1, with $(\theta_1, \theta_2) = (0, 1)$. Denote this CF by $\varphi(t)$. An empirical CF–based test statistic may be written as

$$T_{n,\beta}(\vartheta) = n \int_{-\infty}^{\infty} |D_n(t)|^2 \beta(t) dt,$$

where $D_n(t)$ denotes some distance involving the empirical CF $\varphi_n(t)$ of $Y_j,\ j = 1, 2, ..., n$, and, perhaps, $\varphi(t)$. The 'direct approach' of letting $D_n(t) = (\varphi_n(t) - \varphi(t))$, was followed by Epps and Pulley (1983), Gürtler and Henze (2000a) (see also Matsui and Takemura, 2006),

Meintanis (2004a) and Meintanis(2004b) in testing for the normal, the Cauchy, the Laplace and the logistic distribution, respectively. If the weight function satisfies $\beta(t) = \beta(-t),\ t \in \mathbb{R}$, straightforward algebra leads to the representation (2.1) with $g(x;t) = \sin(tx)+\cos(tx)-\varphi(t)$.

Although employing the function $\varphi_n(t) - \varphi(t)$ is apparently the most natural approach, certain modifications may be preferable for specific problems. Such modifications where adopted by Meintanis (2004c) and Meintanis (2005), in testing for the skew normal distribution and symmetric stability, respectively. The test statistics corresponding to these modifications may also be written as in (2.1).

2. Scale families on the positive real line: Assume that the DF of X_1 may be written as $F(x;\vartheta) = G(x/\theta_1)$, for some function G and some constant $\theta_1 > 0$. Then, if $Y_1 = h(X_1, \vartheta_1)$, with $h(x, \vartheta_1) = x/\theta_1$, $\vartheta_1 = \theta_1$, the LT of Y_1 coincides with the LT of X_1 with $\theta_1 = 1$. Denote this LT by $L(t)$. An empirical LT–based test statistic may be written as

$$T_{n,\beta}(\vartheta) = n \int_0^\infty D_n^2(t)\beta(t)dt,$$

where $D_n(t)$ denotes some distance involving the empirical LT $L_n(t)$ of $Y_j,\ j = 1, 2, ..., n$, and, perhaps, $L(t)$. Several test statistics based on the empirical LT may be formulated as (2.1). In particular, when testing for exponentiality, the straightforward approach of letting $D_n(t) = L_n(t) - L(t)$ was taken up by Henze (1993), and Henze and Meintanis (2002a). This approach leads to representation (2.1) with $g(x;t) = \exp(-tx) - (1+t)^{-1}$. On the other hand, and motivated by the differential equation $(1+t)L'(t) + L(t) = 0$ which is satisfied by the LT of the unit exponential distribution, Baringhaus and Henze (1991) employed $D_n(t) = (1+t)L_n'(t) + L_n(t)$, which in turn leads to (2.1) with $g(x;t) = e^{-tx}(1 - x(1+t))$.

Modifications of the above LT–based methods may be found in Henze and Klar (2002), Henze and Meintanis (2002b), Meintanis and Iliopoulos (2003), Henze and Meintanis (2004), and Meintanis (2006), corresponding to goodness–of–fit tests for the inverse Gaussian, the exponential, the Rayleigh, the Gamma, and the strictly positive stable, distributions, respectively.

3. Discrete families: Suppose $Y_1 = h(X_1, \vartheta_1) \equiv X_1$, the identity function. Let $P(t)$ denote the PGF of X_1. A transform–based test statistic for discrete distributions is typically written as

$$T_{n,\beta}(\vartheta) = n \int_0^\infty D_n^2(t)\beta(t)dt,$$

where $D_n(t)$ denotes some distance involving the empirical PGF $P_n(t)$ of $X_j,\ j = 1, 2, ..., n$,

and, perhaps, $P(t)$. Several PGF–based tests which can be formulated as in (2.1) were developed for the Poisson distribution with mean θ. In particular, Rueda *et al.* (1991) followed the straightforward approach of letting $D_n(t) = P_n(t) - P(t)$, with weight function $\beta \equiv 1$. Their statistic was later generalized by Gürtler and Henze (2000b), by taking $\beta(t) = t^b,\ b \geq 0$. In both cases we are led to representation (2.1), with $g(x, \vartheta_2; t) = t^x - \exp(\theta(t-1)), \vartheta_2 = \theta$.

Motivated by the differential equation $P'(t) - \theta P(t) = 0$, satisfied by the PGF of the Poisson distribution, Baringhaus and Henze (1992) employed $D_n(t) = P_n'(t) - \theta P_n(t)$, with $\beta \equiv 1$. This statistic is a special case of the test of Gürtler and Henze (2000b), in which $\beta(t) = t^b$ is employed as a weight function. The differential equation approach leads in both cases to (2.1), with $g(x, \vartheta_2; t) = \theta t^x - x t^{x-1}, \vartheta_2 = \theta$.

Apart from the Poisson distribution, Rueda and O'Reilly (1999) developed statistics for testing the negative binomial distribution, but the most general PGF–based test was proposed by Meintanis and Bassiakos (2005). It corresponds to the family of invariant under convolutions count distributions, with finite variance. Members of this class are the Poisson distribution and the generalized Hermite distribution.

Remark 2.1 The proposed test statistics depend on the value of the parameter ϑ. In composite goodness–of–fit problems however, the parameter ϑ needs to be estimated. In what follows we will consider the asymptotic properties of the test statistic $\widehat{T}_{n,\beta}$, which results from $T_{n,\beta}(\vartheta)$ by replacing ϑ with a consistent estimator $\hat{\vartheta}_n$.

Remark 2.2 The transform $K_n(\cdot)$ involved in $T_{n,\beta}(\vartheta)$ is computed over an interval of the real line. In this sense the proposed test statistics may be viewed as analogous to classical continuous type tests based on the empirical DF, such as the Crámer–von Mises or the Anderson–Darling statistic (see D'Agostino and Stephens, 1986) . Tests where $K_n(\cdot)$ is computed over a finite grid of points have also been developed. These tests are of the discrete type and rely on the χ^2 principle. Hence they may be viewed as analogous to the classical χ^2 test, and as such, do not possess the feature of being consistent under general alternatives. For a general formulation of goodness–of–fit procedures of the χ^2 type based on transforms, the reader is referred to Luong and Thompson (1987).

3 Large–sample results

A convenient setting for asymptotic distribution theory is the separable Hilbert space $\mathcal{H} = L^2(\mathbb{R}, \mathcal{B}, \beta(t)dt)$ of (equivalence classes of) measurable functions $f : \mathbb{R} \to \mathbb{R}$, with $\|f\|^2 =$

$\int f^2(t)\beta(t)dt < \infty$, where $\beta(\cdot)$ denotes a finite measure. Notice that in the Hilbert space framework,

$$\widehat{T}_{n,\beta} = \|Z_n\|^2, \text{ with } Z_n(t) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n} g(\hat{Y}_j, \hat{\vartheta}_{2n}; t), \tag{3.1}$$

where $\hat{Y}_j = h(X_j, \hat{\vartheta}_{1n})$, and $\hat{\vartheta}_n = (\hat{\vartheta}_{1n}, \hat{\vartheta}_{2n})^T$ denotes a consistent estimator of ϑ. Here and in what follows, the notation $\xrightarrow{\mathcal{D}}$ means convergence in distribution of random elements and random variables, $\xrightarrow{\mathrm{P}}$ means convergence in probability, $o_P(1)$ stands for convergence in probability to 0, $O_P(1)$ denotes boundedness in probability.

Naturally, asymptotic properties of the test statistic will depend on the estimator $\hat{\vartheta}_n$ of the parameter involved, and on the assumptions regarding the functions g and h in (2.1), as well as on certain properties of the measure β. In particular, for the asymptotic results of the test statistic $\widehat{T}_{n,\beta}$ we rely on the following assumptions:

(A.1) Under H_0, for some (true) value $\vartheta_0 = (\vartheta_{10}, \vartheta_{20})^T \in \Theta$,

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta_0) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n} \mathbf{l}(X_j; \vartheta_0) + \mathbf{o}_P(1),$$

where $\mathbf{l}$ denotes a measurable column vector valued function of dimension k, such that $\mathrm{E}_{\vartheta_0}[\mathbf{l}(\mathrm{X}_1; \vartheta_0)] = 0$ and $\mathrm{E}_{\vartheta_0}[\mathbf{l}(\mathrm{X}_1; \vartheta_0)\mathbf{l}(\mathrm{X}_1; \vartheta_0)^{\mathrm{T}}] < \infty$; here $\mathbf{o}_P(1)$ is a vector consisting of k $o_P(1)$ elements.

(A.2) For all $x \in \mathbb{R}$, the function $h(x, \vartheta_1)$ is continuous at ϑ_{10}. Furthermore,
$g(h(x, \vartheta_1), \vartheta_2; \cdot) \in \mathcal{H}$ for $x \in \mathbb{R}$ and $\vartheta = (\vartheta_1, \vartheta_2)^T \in \Theta_0 \subset \Theta$, and each component of the k-dimensional column vector

$$q(x, \vartheta; \cdot) := \frac{\partial}{\partial \vartheta} g(h(x, \vartheta_1), \vartheta_2; \cdot)$$

is, for each $x \in \mathbb{R}$, continuous at $\vartheta_0 = (\vartheta_{10}, \vartheta_{20})^T$ in $\mathcal{H}$, where Θ_0 is any open neighborhood of the true value $\vartheta_0 = (\vartheta_{10}, \vartheta_{20})^T \in \Theta$.

(A.3) $\mathrm{E}_{\vartheta_0}[g(Y_{10}, \vartheta_{20}; t)] = 0, t \in \mathbb{R}$, $\mathrm{E}_{\vartheta_0}[\|g(Y_{10}, \vartheta_{20}; \cdot)\|^2] < \infty$, $\mathrm{E}_{\vartheta_0}[\|q(Y_{10}, \vartheta_{20}; \cdot)\|^2] < \infty$, where $Y_{10} = h(X_1, \vartheta_{10})$.

In the following theorem, the asymptotic null distribution of the test statistic is derived.

Theorem 3.1 *Let $X_1, X_2, \ldots, X_n$ be i.i.d. observations with DF, $F \in \mathcal{F}_\Theta$. Assume that representation (3.1) holds for the test statistic $\widehat{T}_{n,\beta}$, $\hat{\vartheta}_n$ satisfies (A.1) and $\hat{\vartheta}_n \to \vartheta_0$ a.s. as $n \to \infty$. Suppose the functions g and h, as well as the measure β in (2.1), satisfy (A.2) and (A.3). Then*

$$\widehat{T}_{n,\beta} = ||W_n||^2 + o_{\mathrm{P}}(1), \quad \text{as} \quad n \to \infty, \text{ where} \tag{3.2}$$

$$W_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} W_j^0(t), \ \textit{with} \tag{3.3}$$

$$W_j^0(t) = g(Y_{j0}, \vartheta_{20}; t) + (\mathrm{E}_{\vartheta_0}[q(Y_{10}, \vartheta_{20}; t)])^T \, \mathbf{I}(X_j; \vartheta_0), \ \text{and } Y_{j0} = h(X_j, \vartheta_{10}). \tag{3.4}$$

Moreover, as $n \to \infty$, there is a zero–mean continuous Gaussian process $\mathcal{W} = \{\mathcal{W}(t);\ t \in \mathbb{R}\}$ such that

$$W_n \xrightarrow{\mathcal{D}} \mathcal{W}, \ \textit{and} \tag{3.5}$$

$$\widehat{T}_{n,\beta} \xrightarrow{\mathcal{D}} \|\mathcal{W}\|^2 := T_\beta. \tag{3.6}$$

Proof. Our approximation will be accomplished in three consecutive steps, starting with the process $Z_n(\cdot)$ of (3.1) in the first step, and arriving at the process $W_n(\cdot)$ after the third step. In each step a process U_n will be replaced by a process V_n, provided that

$$U_n(t) = V_n(t) + \epsilon_n(t), \ \text{with } \|\epsilon_n\|^2 = o_P(1). \tag{3.7}$$

By the mean–value theorem,

$$\begin{aligned} Z_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} g(Y_{j0}, \vartheta_{20}; t) + \left(\frac{1}{n} \sum_{j=1}^{n} q(Y_{j0}, \vartheta_{20}; t) \right)^T \sqrt{n}(\hat{\vartheta}_n - \vartheta_0) \\ + \left(\frac{1}{n} \sum_{j=1}^{n} (q(\widetilde{Y}_{j0}, \tilde{\vartheta}_{20}; t) - q(Y_{j0}, \vartheta_{20}; t)) \right)^T \sqrt{n}(\hat{\vartheta}_n - \vartheta_0), \end{aligned}$$

where $Y_{j0} = h(X_j, \vartheta_{10})$, $\widetilde{Y}_{j0} = h(X_j, \tilde{\vartheta}_{10})$, and $\|\tilde{\vartheta}_0 - \vartheta_0\| \le \|\hat{\vartheta}_n - \vartheta_0\| \to 0$ a.s. as $n \to \infty$. From (A.1), (A.2), the multivariate CLT and the Cauchy–Schwarz inequality it follows that $\sqrt{n}(\hat{\vartheta}_n - \vartheta_0) = \mathbf{O}_P(1)$, and

$$\left(\frac{1}{n} \sum_{j=1}^{n} \Big(q(\widetilde{Y}_{j0}, \tilde{\vartheta}_{20}; \cdot) - q(Y_{j0}, \vartheta_{20}; \cdot) \Big) \right)^T \sqrt{n}(\hat{\vartheta}_n - \vartheta_0) = o_P(1)$$

in $\mathcal{H}$. Hence, (3.7) holds with U_n replaced by Z_n, and V_n replaced by

$$\widetilde{Z}_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} g(Y_{j0}, \vartheta_{20}; t) + \left(\frac{1}{n} \sum_{j=1}^{n} q(Y_{j0}, \vartheta_{20}; t) \right)^T \sqrt{n}(\hat{\vartheta}_n - \vartheta_0).$$

The second step consists of replacing $\widetilde{Z}_n$ by the process

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} g(Y_{j0}, \vartheta_{20}; t) + (\mathrm{E}_{\vartheta_0}[q(Y_{10}, \vartheta_{20}; t)])^T \sqrt{n}(\hat{\vartheta}_n - \vartheta_0), \tag{3.8}$$

where $Y_{10} = h(X_1, \vartheta_{10})$. This is permissible because of assumption (A.3) and the Hilbert Space Strong Law of Large Numbers. Thus (3.7) holds with U_n replaced by $\widetilde{Z}_n$ and V_n replaced by the process in (3.8).

In the last approximation, $\sqrt{n}(\hat{\vartheta}_n - \vartheta_0)$ in (3.8) is replaced by the asymptotically non–negligible part of the right–hand side of (A.1). Hence (3.7) holds with U_n replaced by the process in (3.8), and V_n replaced by the process W_n specified by (3.3)–(3.4). Then under the standing assumptions, and from the previous steps, it follows that $Z_n(t) = W_n(t) + \epsilon_n(t)$, with $\epsilon_n(\cdot)$ as in (3.7). Therefore by means of (3.1) one has,

$$\widehat{T}_{n,\beta} = \|W_n\|^2 + \|\epsilon_n\|^2 + 2 < W_n, \epsilon_n >, \tag{3.9}$$

with $< f, g > = \int f(t)g(t)\beta(t)dt$. In turn, an application of the Cauchy–Schwarz inequality yields

$$< W_n, \epsilon_n > \leq (\|W_n\|^2)^{1/2}(\|\epsilon_n\|^2)^{1/2}. \tag{3.10}$$

Notice however that $\|\epsilon_n\|^2 = o_P(1)$, and also that $\{W_n(t),\ t \in \mathbb{R}\}$ is the sum of the zero–mean i.i.d. random variables $W_j^0(t)$, with $\mathrm{E}_{\vartheta_0}[\|W_1^0\|^2] < \infty$ (apply (A.1), (A.3) and Fubini's Theorem). Therefore assertion (3.5) is a consequence of the Hilbert Space Central Limit Theorem for i.i.d. random elements (see, e.g., van der Vaart and Wellner 1996, §1.8). Moreover, the Continuous Mapping Theorem implies that $\|W_n\|^2 \xrightarrow{\mathcal{D}} \|\mathcal{W}\|^2$, and hence $\|W_n\|^2 = O_P(1)$. Then utilizing (3.9) and (3.10), (3.2) follows, which in turn implies (3.6), and the proof of the theorem is completed. ■

Remark 3.1 Typically, the finite–sample as well as the asymptotic distribution of the test statistic depends on the true value ϑ_0 of the parameter $\vartheta = (\vartheta_1, \vartheta_2)^T$. In location–scale families however the incorporation of a suitable estimator $\hat{\vartheta}_{1n} = (\hat{\theta}_{1n}, \hat{\theta}_{2n})^T$, for the parameter $\vartheta_1 = (\theta_1, \theta_2)^T$, renders both distributions independent of the true value ϑ_{10}. In particular we employ an estimator, say $\hat{\theta}_{1n} = \hat{\theta}_{1n}(X_1, X_2, ..., X_n)$, of the location parameter θ_1 which is location and scale equivariant, i.e., for each $\delta \in \mathbb{R}$ and $c > 0$ we have,

$$\hat{\theta}_{1n}(\delta + cX_1, \delta + cX_2, ..., \delta + cX_n) \ = \ \delta + c\ \hat{\theta}_{1n}(X_1, X_2, ..., X_n).$$

Likewise, let $\hat{\theta}_{2n} = \hat{\theta}_{2n}(X_1, ..., X_n)$ be an estimator of the scale parameter θ_2 which is location invariant and scale equivariant, i.e., for each $\delta \in \mathbb{R}$ and $c > 0$, we have

$$\hat{\theta}_{2n}(\delta + cX_1, \delta + cX_2, ..., \delta + cX_n) \ = \ c\ \hat{\theta}_{2n}(X_1, X_2, ..., X_n).$$

Then if the test statistic employs the data $Y_j = (X_j - \hat{\theta}_{1n})/\hat{\theta}_{2n},\ j = 1, 2, ..., n$ (or some trasnsformation thereof), the law of $\hat{T}_{n,\beta}$, and consequently the corresponding law of T_β, are independent of the true value $\vartheta_{10} = (\theta_{10}, \theta_{20})^T$. Therefore without loss of generality we may assume that $(\theta_{10}, \theta_{20}) = (0, 1)$.

Remark 3.2 The exact distribution of T_β is complicated. It results from the theory of integral operators. Specifically, let $\omega(s,t)$ denote the covariance kernel of $\mathcal{W}$ which may be calculated from (3.4). Then, the distribution of T_β coincides with that of $\sum_{j\geq 1} \lambda_j(\beta) N_j^2$, where $N_1, N_2, ...,$ are i.i.d. standard normal variates, and $\lambda_j\ ,\ j = 1, 2, ...,$ are the nonzero eigenvalues of the integral equation $\int \omega(s,t) G_j(t) \beta(t) dt = \lambda_j G_j(s)$ with corresponding eigenfunctions $G_j(\cdot)$.

There are several basic problems in computing percentiles of T_β and perform a large–sample test thereof. The first is to find the eigenvalues, and then write the CF of T_β as an infinite product involving $\lambda_j,\ j = 1, 2, ...$. Secondly, one has to approximate this infinite product and invert the corresponding approximation to obtain the asymptotic distribution of the test statistic. In classical goodness–of–fit tests with known parameter values (or with unknown location and scale parameters), such as those based on the Cramér–von Mises or the Anderson–Darling stastistic, the eigenvalues have in some cases been found, and the corresponding CF was inverted to yield asymptotic percentiles. However, in most cases of composite goodness–of–fit, the test has been performed by simulating the distribution of the test statistic by Monte Carlo methods, particularly so when, apart from location and scale, a shape parameter is involved. The interested reader is referred to D'Agostino and Stephens (1986), and the references therein.

The problem of approximating the limit null distribution is even more complicated for transform–based statistics. It is only very recently that Matsui and Takemura (2005) presented sophisticated numerical techniques in determining percentage points for T_β, for composite goodness–of–fit with Cauchy distributions. These techniques have recently been extended to cover the case of symmetric stable distributions with unspecified characteristic exponent (Matsui and Takemura, 2006). In our view however, the best way to circumvent this problem is to follow the line of approach taken in classical statistics based on the empirical DF, and utilize Monte Carlo in determining the finite–sample critical points empirically. This approach is in fact pretty much straightforward in location–scale families (see for instance Epps, 2005, for a unified treatment), but the presence of an extra shape parameter necessitates the use of the parametric bootstrap, and the consequent justification of this procedure

on both theoretical and applied grounds.

In the rest of this section we will be concerned with the asymptotic behavior of the test statistic under alternatives. To this end, recall from (3.1) that

$$\frac{\widehat{T}_{n,\beta}}{n} = \int_{-\infty}^{\infty} D_n^2(t)\beta(t)dt,$$

where $D_n(t) = n^{-1}\sum_{j=1}^{n} g(\hat{Y}_j, \hat{\vartheta}_{2n}; t)$. Informally, it is often the case that since the function g contains elements of the kernel $k(t,x)$ which is bounded, g is also bounded so that $D_n^2(t) \leq \widetilde{D}_n^2(t)$, with $\|\widetilde{D}_n\|^2 < \infty$. Then provided that for each t, $D_n(t)$ converges almost surely to a quantity that is not identically zero under a fixed alternative, a version of Lebesgue's Dominated Convergence Theorem yields the consistency of the test that rejects H_0 for large values of $\widehat{T}_{n,\beta}$ against such alternative distributions. In the following theorem, a stronger result is proved.

Theorem 3.2 *Suppose that for a specific fixed alternative (A.1)–(A.3) hold with ϑ_0 replaced by $\tilde{\vartheta} = (\tilde{\vartheta}_1, \tilde{\vartheta}_2)^T$ and $\mathbf{l}$ replaced by some $\tilde{\mathbf{l}}$ (which often is of the same type as $\mathbf{l}$). Assume that $g(h(x,\vartheta_1),\vartheta_2;\cdot)$ is continuous at $\tilde{\vartheta} = (\tilde{\vartheta}_1, \tilde{\vartheta}_2)^T$ in $\mathcal{H}$, uniformly in x, and that $\hat{\vartheta}_n \to \tilde{\vartheta}$ a.s. as $n \to \infty$. Let $\widetilde{T}_\beta = \|\mathrm{E}_{\tilde{\vartheta}}[g(\widetilde{Y}_1, \tilde{\vartheta}_2;\cdot)]\|^2$, with $\widetilde{Y}_1 = h(X_1, \tilde{\vartheta}_1)$, and $0 < \widetilde{T}_\beta < \infty$. Then, for some $0 < \sigma^2 < \infty$,*

$$\sqrt{n}\left(\frac{\widehat{T}_{n,\beta}}{n} - \widetilde{T}_\beta\right) \xrightarrow{\mathcal{D}} N(0,\sigma^2), \text{ as } n \to \infty. \tag{3.11}$$

Proof. Let $\widehat{\mathcal{T}}_{n,\beta} = \sqrt{n}(\widehat{T}_{n,\beta}/n - \widetilde{T}_\beta)$, and write

$$\widehat{\mathcal{T}}_{n,\beta} = <\sqrt{n}(D_n - \widetilde{D}), D_n + \widetilde{D}>, \tag{3.12}$$

where $\widetilde{D}(t) = \mathrm{E}_{\tilde{\vartheta}}[g(\widetilde{Y}_1, \tilde{\vartheta}_2; t)]$. Application of the Cauchy–Schwarz inequality, Fubini's Theorem, the Hilbert Space Strong Law of Large Numbers and the standing assumptions yield that $\|D_n - \widetilde{D}\|^2 = o_P(1)$ in $\mathcal{H}$. Hence, it follows from (3.12) that the asymptotic distribution of $\widehat{\mathcal{T}}_{n,\beta}$ is the same as that of $<\sqrt{n}(D_n - \widetilde{D}), 2\widetilde{D}>$. It remains to show that

$$\sqrt{n}(D_n - \widetilde{D}) \xrightarrow{\mathcal{D}} \widetilde{\mathcal{D}}, \text{ as } n \to \infty, \tag{3.13}$$

where $\widetilde{\mathcal{D}}$ is a zero–mean Gaussian process. Then the Continuous Mapping Theorem would give $\widehat{\mathcal{T}}_{n,\beta} \xrightarrow{\mathcal{D}} <\widetilde{\mathcal{D}}, 2\widetilde{D}>$. To prove (3.13) we proceed by using a completely analogous reasoning as the one used in the proof of Theorem 3.1, and therefore the details are omitted.

In particular, using the notation $\approx$ when (3.7) holds, it follows that under the standing assumptions the process $\sqrt{n}(D_n - \widetilde{D})$ is asymptotically equivalent to each of the following processes (compare with the three–step approximation of Z_n in the proof of Theorem 3.1),

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\{g(\widetilde{Y}_j,\tilde{\vartheta}_2;t)-\widetilde{D}(t)\}+\left(\frac{1}{n}\sum_{j=1}^{n}q(\widetilde{Y}_j,\tilde{\vartheta}_2;t)\right)^T\sqrt{n}(\hat{\vartheta}_n-\tilde{\vartheta})$$

$$\approx \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\{g(\widetilde{Y}_j,\tilde{\vartheta}_2;t)-\widetilde{D}(t)\}+\left(\mathrm{E}_{\tilde{\vartheta}}\left[q(\widetilde{Y}_1,\tilde{\vartheta}_2;t)\right]\right)^T\sqrt{n}(\hat{\vartheta}_n-\tilde{\vartheta})$$

$$\approx \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\widetilde{W}_j(t),$$

with $\widetilde{Y}_j = h(X_j,\tilde{\vartheta}_1),\ j=1,2,...,n$, and

$$\widetilde{W}_j(t)=g(\widetilde{Y}_j,\tilde{\vartheta}_2;t)-\widetilde{D}(t)+\left(\mathrm{E}_{\tilde{\vartheta}}\left[q(\widetilde{Y}_1,\tilde{\vartheta}_2;t)\right]\right)^T\tilde{\mathrm{l}}(\widetilde{Y}_j;\tilde{\vartheta}).$$

Hence (3.13) is an immediate consequence of the Hilbert Space Central Limit Theorem, and the proof of the theorem is completed. ■

Remark 3.3 The covariance kernel $\tilde{\omega}(s,t)$ of $\widetilde{\mathcal{D}}$ can be calculated as $\mathrm{E}[\widetilde{W}_1(s)\widetilde{W}_1(t)]$. Then the variance figuring in (3.11) is given by

$$\sigma^2 = 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\tilde{\omega}^2(s,t)\widetilde{D}(s)\widetilde{D}(t)\beta(s)\beta(t)dsdt.$$

4 Bootstrap procedures

As mentioned in Remark 3.2, several nonparametric goodness–of–fit test statistics are defined in terms of estimated nuisance parameters. In such a case, it is well known that critical values shift, and the asymptotic null distribution of the test statistic may depend in a complex way on the unknown parameters.

In this section we apply the bootstrap to estimate the null distribution of the statistic $\widehat{T}_{n,\beta}$ defined in (3.1). We shall consider both parametric and nonparametric bootstrap methods. It will be shown that under some regularity conditions both bootstrap procedures lead to consistent estimates of the percentiles of the true null distribution of the test statistic when the parameters involved are estimated. The nonparametric bootstrap will also help in the computation of power under contiguous alternatives $\lambda_n = \vartheta_0 + \lambda n^{-1/2}$. However, in the case of nonparametric bootstrap we have to make a correction for the bias.

To describe the bootstrap procedure, let $X_1^*, X_2^*, \ldots, X_n^*$ be i.i.d. random variables from $\widehat{F}_n$, where $\widehat{F}_n$ is an estimator of the DF, F, based on the random sample $X_1, X_2, \ldots, X_n$. Let $\hat{\vartheta}_n^* = (\hat{\vartheta}_{1n}^*, \hat{\vartheta}_{2n}^*)^T$, where $\hat{\vartheta}_{1n}^* = \hat{\vartheta}_{1n}(X_1^*, X_2^*, \ldots, X_n^*)$ and $\hat{\vartheta}_{2n}^* = \hat{\vartheta}_{2n}(X_1^*, X_2^*, \ldots, X_n^*)$. The resampling method is called parametric bootstrap (denoted by (P)) if $\widehat{F}_n = F_{\hat{\vartheta}_n}$, and it is called nonparametric bootstrap (denoted by (N)) if $\widehat{F}_n = F_n$, the empirical DF.

Using the same notation as in Section 3, we state the following set of assumptions required for Theorem 4.1 below.

(B.1) The local asymptotic expansion in (A.1) holds in P_{ϑ_0}–probability, where $\mathbf{l}$ now satisfies $\mathrm{E}_\vartheta[\mathbf{l}(\mathrm{X}_1;\vartheta)] = 0$ for $\vartheta \in \Theta_0$, and $L(\vartheta) := \mathrm{E}_\vartheta[\mathbf{l}(\mathrm{X}_1;\vartheta)\mathbf{l}(\mathrm{X}_1;\vartheta)^{\mathrm{T}}]$ is a finite positive definite matrix for $\vartheta \in \Theta_0$, where Θ_0 is any open neighborhood of the true value $\vartheta_0 = (\vartheta_{10}, \vartheta_{20})^T$ under H_0. Furthermore, $L(\vartheta)$ is continuous at ϑ_0.

(B.2) For all $x \in \mathbb{R}$, the function $h(x, \vartheta_1)$ is uniformly continuous in $\vartheta_1 \in \Theta_{10}$, for a given open neighborhood Θ_{10} of ϑ_{10}. Furthermore, $g(h(x, \vartheta_1), \vartheta_2; \cdot) \in \mathcal{H}$ for $x \in \mathbb{R}$ and $\vartheta = (\vartheta_1, \vartheta_2)^T \in \Theta_0 \subset \Theta$, and each component of the k-dimensional column vector

$$q(x, \vartheta; \cdot) := \frac{\partial}{\partial \vartheta} g(h(x, \vartheta_1), \vartheta_2; \cdot)$$

is, for each $x \in \mathbb{R}$, uniformly continuous on Θ_0, in $\mathcal{H}$.

(B.3) $\mathrm{E}_\vartheta[g(Y_1, \vartheta_2; t)] = 0, t \in \mathbb{R}$, $\mathrm{E}_\vartheta[\|g(Y_1, \vartheta_2; \cdot)\|^2] < \infty$, $\mathrm{E}_\vartheta[\|q(Y_1, \vartheta_2; \cdot)\|^2] < \infty$ for $\vartheta \in \Theta_0$, where $Y_1 = h(X_1, \vartheta_1)$.

(B.4) As $\gamma \to \infty$,

$$\sup_{\vartheta \in \Theta_0} \int_{\{\|\mathbf{l}(x;\vartheta)\| > \gamma\}} \|\mathbf{l}(x;\vartheta)\|^2 \, dF_\vartheta(x) \to 0.$$

(P) Under the bootstrap probability measure P^*,

$$\sqrt{n}(\hat{\vartheta}_n^* - \hat{\vartheta}_n) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \mathbf{l}(X_j^*; \hat{\vartheta}_n) + \mathbf{o}_{P^*}(1).$$

(N) Under the bootstrap probability measure P^*,

$$\sqrt{n}(\hat{\vartheta}_n^* - \hat{\vartheta}_n) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \mathbf{l}(X_j^*; \vartheta_0) - \frac{1}{\sqrt{n}} \sum_{j=1}^n \mathbf{l}(X_j; \vartheta_0) + \mathbf{o}_{P^*}(1).$$

The meaning of $\mathbf{R}_n^* := \mathbf{o}_{P^*}(1)$ is that for all $\varepsilon > 0$, $P^*(\|\mathbf{R}_n^*\| \geq \varepsilon) \to 0$ a.s. as $n \to \infty$. Assumptions (B.1) and (P), or (B.1) and (N) hold, in general, for maximum likelihood estimators, M–estimators and L–statistics.

Define the parametric bootstrap version of $\widehat{T}_{n,\beta}$ in (3.1) as,

$$\widehat{T}_{n,\beta}^* = \|Z_n^*\|^2, \text{ with } Z_n^*(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n g(\widehat{Y}_j^*, \hat{\vartheta}_{2n}^*; t), \tag{4.1}$$

where $\widehat{Y}_j^* = h(X_j^*, \hat{\vartheta}_{1n}^*)$. Theorem 4.1 below shows that the conditional distribution of $\widehat{T}_{n,\beta}^*$, given $X_1, X_2, \ldots, X_n$, approximates consistently the null distribution of $\widehat{T}_{n,\beta}$.

Theorem 4.1 *Let $X_1, X_2, \ldots, X_n$ be i.i.d. observations with DF, $F \in \mathcal{F}_\Theta$. Assume conditions (B.1)–(B.4) and (P) hold. Suppose that, as $n \to \infty$, $\hat{\vartheta}_n \to \vartheta_0$, P_{ϑ_0}–a.s., and $\hat{\vartheta}_n^* - \hat{\vartheta}_n \to 0$, P^*–a.s. Then, as $n \to \infty$,*

$$\sup_{x \in \mathbf{R}} \left| P^*(\widehat{T}_{n,\beta}^* \le x) - P(\widehat{T}_{n,\beta} \le x) \right| \to 0, \text{ a.s.}$$

Proof. The proof proceeds in exactly the same way as that of Theorem 3.1. Applying the standing assumptions, the expression in (3.8) now becomes

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} g(Y_j^*, \hat{\vartheta}_{2n}; t) + \Big(\mathrm{E}_* \Big[q(Y_1^*, \hat{\vartheta}_{2n}; t) \Big] \Big)^T \sqrt{n}(\hat{\vartheta}_n^* - \hat{\vartheta}_n), \tag{4.2}$$

where $Y_j^* = h(X_j^*, \hat{\vartheta}_{1n})$, and E_* denotes the expectation with respect to P^*. The next step is to replace $\sqrt{n}(\hat{\vartheta}_n^* - \hat{\vartheta}_n)$ with the asymptotically non-negligible part of the right–hand side of the equation in assumption (P). Hence, $W_j^0(t)$ in (3.4) is now replaced by

$$W_j^{0*}(t) = g(Y_j^*, \hat{\vartheta}_{2n}; t) + \Big(\mathrm{E}_* \Big[q(Y_1^*, \hat{\vartheta}_{2n}; t) \Big] \Big)^T \mathbf{I}(X_j^*; \hat{\vartheta}_n). \tag{4.3}$$

Now, it is a matter of simple algebra using the stated assumptions, Fubini's theorem and Lebesgue's Dominated Convergence Theorem to show that

$$\mathrm{E}_* \Big[\big\| W_1^{0*} \big\|^2 \Big] \to \mathrm{E}_{\vartheta_0} \Big[\big\| W_1^0 \big\|^2 \Big] < \infty, \text{ a.s. as } n \to \infty, \tag{4.4}$$

with $W_1^0(\cdot)$ as defined in (3.4).

As in the proof of Theorem 3.1, it follows from (4.2)–(4.4), the Hilbert Space CLT and the Continuous Mapping Theorem (assumption (B.4) implies Lindeberg's condition) that along almost all sample sequences $X_1, X_2, \ldots$, given $(X_1, X_2, \ldots, X_n)$, $\|W_n^*\|^2 \xrightarrow{\mathcal{D}^*} \|\mathcal{W}\|^2$ as $n \to \infty$, where $W_n^*(t) = n^{-1/2} \sum_{j=1}^n W_j^{0*}(t)$ and $\mathcal{W}(\cdot)$ as defined in Theorem 3.1. Here $\mathcal{D}^*$ denotes convergence in distribution with respect to P^*. Hence,

$$\widehat{T}_{n,\beta}^* \xrightarrow{\mathcal{D}^*} \|\mathcal{W}\|^2, \ P_{\vartheta_0} - \text{a.s. as } n \to \infty. \tag{4.5}$$

The proof of the theorem now follows from (3.6), (4.5) and Polyá's Theorem. ■

Remark 4.1 If conditions (B.1)–(B.4) and (N) hold, then by repeating the steps of Theorem 4.1, it follows directly that Theorem 4.1 also holds for the nonparametric bootstrap version of $\widehat{T}_{n,\beta}$, defined by $\widetilde{T}_{n,\beta}^* = \|\widetilde{Z}_n^*\|^2$, with $\widetilde{Z}_n^*(t) = Z_n^*(t) - Z_n(t)$, where $Z_n(t)$ and $Z_n^*(t)$ are defined in (3.1) and (4.1) respectively.

Acknowledgements: The first author wishes to sincerely thank Professor N. Henze and Dr. B. Klar for many discussions related to this work. The second author's research was financially supported by the National Research Foundation of South Africa.

References

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). A goodness–of–fit test for the Poisson distribution based on the empirical generating function. *Statist. Probab. Lett.*, *13*, 269–274.

Csörgő, S. (1981). Limit behaviour of the empirical characteristic function. *Ann. Probab.*, *9*, 130–144.

D'Agostino, R., and Stephens, M. (1986). *Goodness-of-fit techniques.* Marcel Dekker, Inc., New York.

Epps, T.W. (2005). Tests for location–scale families based on the empirical characteristic function. *Metrika*, *62*, 99–114.

Epps, T.W. , and Pulley, L.B. (1983). A test for normality based on the empirical characteristic function. *Biometrika*, *70*, 723–726.

Feuerverger, A. (1989). On the empirical saddlepoint approximation. *Biometrika*, *76*, 457–464.

Gürtler, N., and Henze, N. (2000a). Goodness–of–fit tests for the Cauchy distribution based on the empirical characteristic function. *Ann. Inst. Statist. Math.*, *52*, 267–286.

Gürtler, N., and Henze, N. (2000b). Recent and classical goodness–of–fit tests for the Poisson distribution. *J. Statist. Plann. Infer.*, *90*, 207–225.

Heathcote, C.R. (1972). A test of goodness of fit for symmetric random variables. *Austral. J. Statist.*, *14*, 172–181.

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

Henze, N., and Klar, B. (2002). Goodness–of–fit tests for the inverse Gaussian distribution based on the empirical Laplace transform. *Ann. Inst. Statist. Math.*, *54*, 425–444.

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. G. (2002b). Goodness-of-fit tests based on a new characterization of the exponential distribution. *Commun. Statist. – Theory Meth.*, *31*, 1479–1497.

Henze, N., and Meintanis, S. (2004). Goodness–of–fit tests for the gamma distribution based on the empirical Laplace transform. In A. Andonov *et al*. (Eds.), Transactions of the XXIV International Seminar on Stability Problems for Stochastic Models, 277–285.

Luong, A., and Thompson, M.E. (1987). Minimum–distance methods based on quadratic distances for transforms. *Canad. J. Statist.*, *15*, 239–251.

Matsui, M., and Takemura, A. (2005). Empirical characteristic function approach to goodness–of–fit tests for the Cauchy distribution with parameters estimated by MLE or EISE. *Ann. Inst. Statist. Math.*, *57*, 183–199.

Matsui, M., and Takemura, A. (2006). Goodness–of–fit tests for symmertric stable distributions–Empirical characteristic function approach. *Unpublished manuscript*.

Meintanis, S.G. (2004a). A class of omnibus tests for the Laplace distribution based on the empirical characteristic function. *Commun. Statist. – Theory Meth.*, *33*, 925–948.

Meintanis, S.G. (2004b). Goodness–of–fit tests for the logistic distribution based on empirical transforms. *Sankhyā*, *66*, Part 2, 306–326.

Meintanis, S.G. (2004c). Testing for the skew normal distribution based on the empirical moment generating function. In A. Andonov *et al*. (Eds.), Transactions of the XXIV International Seminar on Stability Problems for Stochastic Models, 310–316.

Meintanis, S.G. (2005). Consistent tests for symmetric stability with finite mean based on the empirical characteristic function. *J. Statist. Plann. Inferen.*, *128*, 373–380.

Meintanis, S.G. (2006). Omnibus tests for strictly positive stable laws based on the empirical Laplace transform. *Math. Meth. Statist.*, *14*, 468–478.

Meintanis, S.G. and Bassiakos, Y. (2005). Goodness–of–fit tests for additively closed count models with an application to the generalized Hermite distribution. *Sankhyā*, *67*, 538–552.

Meintanis, S.G. and Iliopoulos, G. (2003). Test of fit for the Rayleigh distribution based on the empirical Laplace transform. *Ann. Inst. Statist. Math.*, *55*, 137–151.

Rémillard, B., and Theodorescu, R. (2000). Inference based on the empirical probability generating function for mixtures of Poisson distributions. *Statistics and Decisions*, *18*, 349–366.

Rueda, R., and O'Reilly, F. (1999). Tests of fit for discrete distributions based on the probability generating function. *Commun. Statist.– Comput. Simul.*, *28*, 259–274.

Rueda, R., Pérez–Abreu, V., and O'Reilly, F. (1991). Goodness of fit tests for the Poisson distribution based on the probability generating function. *Commun. Statist.– Theory Meth.*, *20*, 3093–3110.

van der Vaart, A.W. , and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics.* Springer–Verlag, New York.