T E C H N I C A L

R E P O R T

0347

BOOTSTRAPPING MODIFIED GOODNESS-OF-FIT STATISTICS WITH ESTIMATED PARAMETERS

Paul JANSSEN, Jan SWANEPOEL
and Noël VERAVERBEKE

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

BOOTSTRAPPING MODIFIED GOODNESS-OF-FIT STATISTICS WITH ESTIMATED PARAMETERS

Paul JANSSEN[1], Jan SWANEPOEL[2]
and Noël VERAVERBEKE[1]

[1] Limburgs Universitair Centrum, Belgium
[2] Potchefstroom University, South Africa

Abstract. Goodness-of-fit tests are proposed for testing a composite null hypothesis that is a general parametric family of distribution functions. They are distribution-free under the null hypothesis and have a limiting normal distribution, under the null and the alternative hypothesis. To avoid the estimation of the asymptotic variance under the alternative hypothesis, we propose consistent bootstrap estimators.

Key Words and Phrases: bootstrap, consistency, estimated parameters, goodness-of-fit, U-statistics.

Mathematics Subject Classification (1991): 62G09, 62G10.

Corresponding author: Noël Veraverbeke, Limburgs Universitair Centrum, Universitaire Campus, B-3590 Diepenbeek (noel.veraverbeke@luc.ac.be)

1 Introduction

In this paper we consider, for a random sample from F, the goodness-of-fit hypothesis $H_0 : F = G$, where G is a given general parametric family of distribution functions, containing unknown parameters $\boldsymbol{\theta}$ (e.g. location, scale, ...) which have to be estimated. It is well known that, in the situation where $\boldsymbol{\theta}$ is known, the classical test statistics typically take the form of a degenerate U- or V-statistic and that the limiting null distribution is that of a (possibly infinite) sum of weighted chi-squared variables. Finding the weights is not easy since they are the eigenvalues of some operator equation and they can only be found in some special cases. We avoid this problem by using some slight modification of the empirical distribution function in the construction of our test statistic. A similar idea has been used in Ahmad (1993, 1996) and Ahmad and Alwasel (1999). This leads to test statistics which have a limiting normal distribution with the usual $n^{1/2}$ standardization both under the null and the alternative hypothesis (Section 3). The first main objective of our paper is to provide conditions under which the replacement of the unknown $\boldsymbol{\theta}$ by a suitable estimator $\widehat{\boldsymbol{\theta}}$ keeps this asymptotic normality in force (Section 4). It turns out that the statistic with estimated nuisance parameter has the same limit distribution under H_0. The question of replacing the unknown $\boldsymbol{\theta}$ by an estimator has not been dealt with in the above references. The problem has been considered by De Wet and Randles (1987) in the unmodified case and our result provides an alternative to their paper. Our approach has the advantage of also providing the limit behavior under the alternative hypothesis $H_1 : F \neq G$. The second main result of our paper is to establish the validity of bootstap approximations; this provides a way to avoid the estimation of the complicated and unknown variance parameter in the asymptotic distribution under the alternative hypothesis H_1 (Sections 5 and 6). Our proposed resampling scheme is nonparametric and works under H_0 and H_1. This is more general than the parametric bootstrap in a recent paper of Jiménez-Gamero et al. (2003), who only prove consistency under H_0. We begin, in Section 2, with a useful characterization for the equality of two continuous distribution functions.

2 Characterization

Assume that F and G are continuous distribution functions. The problem of testing the hypothesis $H_0 : F = G$ versus $H_1 : F \neq G$ is often based on the L_2-distance $\int (F - G)^2 dG$ or more generally on the L_{2p}-distance $\int (F - G)^{2p} dG$ for some $p \geq 1$. If $X_1, \dots , X_n$ is a random sample from F, then an obvious test statistic is given by $\int (F_n - G)^{2p} dG$, where F_n is the usual empirical distribution function of the sample. The case $p = 1$ is the well known Cramér-von Mises statistic. Using the binomial expansion and some integration by parts it is easily shown that the L_{2p}-distance statistics can be rewritten and that this leads to the following alternative characterizations of the null hypothesis: we have that $F = G$ if and only if

$$\sum_{i=2}^{2p} \binom{2p}{i} \frac{(-1)^{2p-i}}{2p-i+1} E\left[G^{2p-i+1}(\max(X_1, \dots, X_i))\right] - E\left[G^{2p}(X_1)\right] = \frac{1}{2p+1}$$

(for $p = 1, 2, \dots$).

In the present paper we will only work with the $p = 1$ version of this characterization, which takes the simple form

$$E[G(\max(X_1, X_2))] - E[G^2(X_1)] = \frac{1}{3}. \tag{1}$$

Note that characterization (1) also implicitly appears in Too and Lin (1989) via a totally different approach.

3 Goodness-of-fit: a distribution function with known parameters

Suppose that we have a random sample $X_1, \dots, X_n$ from an unknown continuous distribution function F and that we want to test the composite hypothesis H_0 : $F(\cdot) = G(\cdot\,; \boldsymbol{\theta})$ versus $H_1 : F(\cdot) \neq G(\cdot\,; \boldsymbol{\theta})$, where $G(\cdot\,; \boldsymbol{\theta})$ is a continuous distribution function depending on some parameter $\boldsymbol{\theta}$. Suppose for the moment that $\boldsymbol{\theta}$ is known. A straightforward empirical estimator for the left hand side of (1) is given by

$$U_n(\boldsymbol{\theta}) = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} \max(G(X_i; \boldsymbol{\theta}), G(X_j; \boldsymbol{\theta})) - \frac{1}{n} \sum_{i=1}^{n} G^2(X_i; \boldsymbol{\theta}). \tag{2}$$

Both terms in (2) are in fact U-statistics with a bounded kernel. Under H_0, it easily follows that $E(U_n(\boldsymbol{\theta})) = 1/3$ and that the kernel is degenerate. So $n^{1/2}(U_n(\boldsymbol{\theta}) - 1/3)$ does not have a limiting normal distribution and the use of the correctly normalized statistic $n(U_n(\boldsymbol{\theta}) - 1/3)$ leads to the problem of finding the eigenvalues. In order to rectify this we consider instead of (2) the following class of modified estimators

$$U_{nc}(\boldsymbol{\theta}) = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} \max(G(X_i; \boldsymbol{\theta}), G(X_j; \boldsymbol{\theta})) - \frac{1}{n} \sum_{i=1}^{n} c_{in} G^2(X_i; \boldsymbol{\theta}) \tag{3}$$

where $\{c_{in}; 1 \leq i \leq n,\ n = 1, 2, \dots\}$ is a triangular array of real numbers, satisfying

$$\begin{array}{ll} \text{(i)} & \max_{1\leq i\leq n} |c_{in}| \leq K, \text{ some constant} \\ \text{(ii)} & \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} c_{in} = 1 \\ \text{(iii)} & \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} c_{in}^2 = c^2 > 1. \end{array} \tag{4}$$

A typical example (see Ahmad (1993)) is: $c_{in} = 1 + \gamma$ if i is odd and $c_{in} = 1 - \gamma$ if i is even, with $0 < \gamma \leq 1$; in this case $c^2 = 1 + \gamma^2$.
The classical theory of U-statistics (see e.g. Serfling (1980)) can be applied to (3), to obtain

$$U_{nc}(\boldsymbol{\theta}) = \mu_{nc}(\boldsymbol{\theta}) + \frac{1}{n}\sum_{i=1}^{n} g_{in}(X_i; \boldsymbol{\theta}) + o_P(n^{-1/2})$$

where (integration is over $I\!R$)

$$\mu_{nc}(\boldsymbol{\theta}) = E(U_{nc}(\boldsymbol{\theta})) = 2\int G(x;\boldsymbol{\theta})F(x)dF(x) - \left\{\frac{1}{n}\sum_{i=1}^{n} c_{in}\right\}\int G^2(x;\boldsymbol{\theta})dF(x) \tag{5}$$

$$\begin{aligned} g_{in}(X_i;\boldsymbol{\theta}) &= 2\left(1 - \int_{X_i}^{\infty} F(x)dG(x;\boldsymbol{\theta}) - 2\int G(x;\boldsymbol{\theta})F(x)dF(x)\right) \\ &\quad -c_{in}\left(G^2(X_i;\boldsymbol{\theta}) - \int G^2(x;\boldsymbol{\theta})dF(x)\right). \end{aligned}$$

It follows that, as $n \to \infty$,

$$n^{1/2}(U_{nc}(\boldsymbol{\theta}) - \mu_{nc}(\boldsymbol{\theta})) \xrightarrow{d} N(0; \sigma^2(\boldsymbol{\theta})) \tag{6}$$

where

$$\begin{aligned} \sigma^2(\boldsymbol{\theta}) &= -4 + 16\int G(x;\boldsymbol{\theta})F(x)dF(x) - 16\left(\int G(x;\boldsymbol{\theta})F(x)dF(x)\right)^2 \\ &\quad +4E\left[\left(\int_{X_1}^{\infty} F(x)dG(x;\boldsymbol{\theta})\right)^2\right] + c^2\left[E(G^4(X_1;\boldsymbol{\theta})) - (EG^2(X_1;\boldsymbol{\theta}))^2\right] \\ &\quad +4E\left[\left(\int_{X_1}^{\infty} F(x)dG(x;\boldsymbol{\theta})\right)(G^2(X_1;\boldsymbol{\theta}) - EG^2(X_1;\boldsymbol{\theta}))\right] \end{aligned} \tag{7}$$

Note 1. (1) Under H_0, we have that $\mu_{nc}(\boldsymbol{\theta}) = \mu_{nc} = \frac{2}{3} - \frac{1}{3}\left\{\frac{1}{n}\sum_{i=1}^{n} c_{in}\right\}$ and $\sigma^2(\boldsymbol{\theta}) = \sigma_0^2 = \frac{4}{45}(c^2 - 1)$. Hence the statistic is asymptotically distribution-free under H_0.

Note 2. In (6) we may replace the exact mean $\mu_{nc}(\boldsymbol{\theta})$ by its asymptotic value, provided that the c_{in} satisfy the condition $n^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n} c_{in} - 1\right) \to 0$. This is satisfied by the specific example above.

4 Goodness-of-fit: a distribution function with estimated parameters

Suppose from now on that the nuisance parameter $\boldsymbol{\theta}$ in $G(\cdot\,;\boldsymbol{\theta})$ is unknown and that it can be estimated by an estimator $\widehat{\boldsymbol{\theta}}$, based on the random sample $X_1, \ldots, X_n$ from F. Let $U_{nc}(\boldsymbol{t})$ and $\mu_{nc}(\boldsymbol{t})$ denote the expressions in (3) and (5) respectively, but with $\boldsymbol{\theta}$ replaced by some general variable $\boldsymbol{t}$.
To obtain the asymptotic distribution of the test statistic $U_{nc}(\widehat{\boldsymbol{\theta}})$ (given in Theorem 2), the following property is a key tool.

Theorem 1. Assume

(i) $n^{1/2}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) = O_P(1)$, as $n \to \infty$

(ii) there exists a neighborhood $K(\boldsymbol{\theta})$ of $\boldsymbol{\theta}$ and a constant $C > 0$ such that, if $\boldsymbol{t} \in K(\boldsymbol{\theta})$ and $D(\boldsymbol{t}, d)$ is a sphere, centered at $\boldsymbol{t}$ with radius d, satisfying $D(\boldsymbol{t}, d) \subset K(\boldsymbol{\theta})$, we have

$$E\left[\sup_{\boldsymbol{t}' \in D(\boldsymbol{t},d)} |G(X_1;\boldsymbol{t}') - G(X_1;\boldsymbol{t})|\right] \le Cd.$$

Then, as $n \to \infty$,

$$n^{1/2}\left[U_{nc}(\widehat{\boldsymbol{\theta}}) - \mu_{nc}(\widehat{\boldsymbol{\theta}}) - U_{nc}(\boldsymbol{\theta}) + \mu_{nc}(\boldsymbol{\theta})\right] \xrightarrow{P} 0.$$

Proof. We apply the result of Randles (1982) on U-statistics with estimated parameters. It is easily verified that the presence of the weights $\{c_{in}\}$ causes no complication, due to their uniform boundedness property in (4). We only have to verify that $G(\max(x_1; x_2); \boldsymbol{t})$ and $G^2(x_1; \boldsymbol{t})$ satisfy his condition (2.4), since these functions are bounded (see Lemma 2.6 in Randles (1982)). Since $|G^2(x_1; \boldsymbol{t'}) - G^2(x_1; \boldsymbol{t})| \leq 2|G(x_1; \boldsymbol{t'}) - G(x_1; \boldsymbol{t})|$ and also $E[\sup_{\boldsymbol{t'} \in D(\boldsymbol{t},d)} |G(\max(X_1, X_2); \boldsymbol{t'}) - G(\max(X_1, X_2); \boldsymbol{t})|] = 2E[\sup_{\boldsymbol{t'} \in D(\boldsymbol{t},d)} |G(X_1; \boldsymbol{t'}) - G(X_1; \boldsymbol{t})| F(X_1)]$, his condition (2.4) is guaranteed by our condition (ii). This proves Theorem 1. ■

Now write

$$\begin{aligned} & n^{1/2}(U_{nc}(\widehat{\boldsymbol{\theta}}) - \mu_{nc}(\boldsymbol{\theta})) \\ & = n^{1/2}(U_{nc}(\boldsymbol{\theta}) - \mu_{nc}(\boldsymbol{\theta}) + \mu_{nc}(\widehat{\boldsymbol{\theta}}) - \mu_{nc}(\boldsymbol{\theta})) \\ & + n^{1/2}(U_{nc}(\widehat{\boldsymbol{\theta}}) - U_{nc}(\boldsymbol{\theta}) - \mu_{nc}(\widehat{\boldsymbol{\theta}}) + \mu_{nc}(\boldsymbol{\theta})). \end{aligned} \tag{8}$$

The above theorem gives conditions under which the second term in the right hand side of (8) is $o_P(1)$. The asymptotic normality of $n^{1/2}(U_{nc}(\widehat{\boldsymbol{\theta}}) - \mu_{nc}(\boldsymbol{\theta}))$ will follow if we can show the asymptotic normality of the first term in the right hand side of (8).

Assume that $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ is a k-dimensional parameter, which has been estimated by $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \dots, \widehat{\theta}_k)$. Typically, we can establish an i.i.d. representation for each component of $\widehat{\boldsymbol{\theta}}$, in the following sense

$$\widehat{\theta}_j - \theta_j = \frac{1}{n} \sum_{i=1}^{n} k_j(X_i) + o_P(n^{-1/2}) \tag{9}$$

where $E(k_j(X_1)) = 0$. The asymptotic normality of $n^{1/2}(U_{nc}(\boldsymbol{\theta}) - \mu_{nc}(\boldsymbol{\theta}) + \mu_{nc}(\widehat{\boldsymbol{\theta}}) - \mu_{nc}(\boldsymbol{\theta}))$ will then follow by applying the central limit theorem to

$$n^{-1/2} \sum_{i=1}^{n} [g_{in}(X_i; \boldsymbol{\theta}) + \sum_{j=1}^{k} k_j(X_i) \mu_{nc}^{(j)}(\boldsymbol{\theta})]$$

where $\mu_{nc}^{(j)}(\boldsymbol{\theta})$ is the partial derivative of $\mu_{nc}(\boldsymbol{t})$ with respect to the j-th variable and evaluated at $\boldsymbol{t} = \boldsymbol{\theta}$:

$$\mu_{nc}^{(j)}(\boldsymbol{\theta}) = 2 \int G^{(j)}(x; \boldsymbol{\theta}) F(x) dF(x) - \left\{ \frac{1}{n} \sum_{i=1}^{n} c_{in} \right\} \int 2G(x; \boldsymbol{\theta}) G^{(j)}(x; \boldsymbol{\theta}) dF(x)$$

for $j = 1, \dots, k$ and where $G^{(j)}(x; \boldsymbol{\theta})$ is the partial derivative of G with respect to the j-th component of $\boldsymbol{\theta}$, and evaluated at $\boldsymbol{\theta}$ (we are assuming that these derivatives exist). We are now able to establish the asymptotic distribution of $U_{nc}(\widehat{\boldsymbol{\theta}})$.

Theorem 2. Assume that the components of the estimator $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \dots, \widehat{\theta}_k)$ satisfy (9) with $E(k_j(X_1)) = 0$ and $\mathrm{Var}(k_j(X_1)) > 0$ for $j = 1, \dots, k$.
Assume that $G(x; \boldsymbol{\theta})$ is a continuous distribution function for which condition (ii) of Theorem 1 is satisfied and which has bounded partial derivatives $G^{(j)}$ in a neighborhood of $\boldsymbol{\theta}$.
Then, as $n \to \infty$,

$$n^{1/2}(U_{nc}(\widehat{\boldsymbol{\theta}}) - \mu_{nc}(\boldsymbol{\theta})) \xrightarrow{d} N(0; \tau^2(\boldsymbol{\theta}))$$

where

$$\begin{aligned} \tau^2(\boldsymbol{\theta}) &= \sigma^2(\boldsymbol{\theta}) - 2\sum_{j=1}^{k} \mu_c^{(j)}(\boldsymbol{\theta}) E\left[\left(2\int_{X_1}^{\infty} F(x)dG(x;\boldsymbol{\theta}) + G^2(X_1;\boldsymbol{\theta})\right) k_j(X_1)\right] \\ &+ \sum_{j=1}^{k}\sum_{j'=1}^{k} \mu_c^{(j)}(\boldsymbol{\theta})\mu_c^{(j')}(\boldsymbol{\theta}) E[k_j(X_1)k_{j'}(X_1)] \end{aligned} \tag{10}$$

and $\mu_c^{(j)}(\boldsymbol{\theta}) = \lim_{n\to\infty} \mu_{nc}^{(j)}(\boldsymbol{\theta}) = 2\int G^{(j)}(x;\boldsymbol{\theta})(F(x) - G(x;\boldsymbol{\theta}))dF(x)$.
Under H_0, we have that $\mu_{nc}(\boldsymbol{\theta}) = \frac{2}{3} - \frac{1}{3}\left\{\frac{1}{n}\sum_{i=1}^{n} c_{in}\right\} = \mu_{nc}$ and $\tau^2(\boldsymbol{\theta}) = \sigma^2(\boldsymbol{\theta}) = \sigma_0^2$.

Example 1. Test for exponentiality.
$G(x;\theta) = 1 - e^{-x/\theta}$; $k = 1$; $\widehat{\theta} = \overline{X}$, the sample mean; $k_1(X_1) = X_1 - \theta$; $G^{(1)}(x;\theta) = -\frac{x}{\theta^2}e^{-x/\theta}$.

Example 2. Test in location-scale family.
$G(x;\boldsymbol{\theta}) = G_0\left(\frac{x-\mu}{\sigma}\right)$ (G_0 known, with density g_0); $k = 2$; $\boldsymbol{\theta} = (\mu, \sigma)$; $\widehat{\boldsymbol{\theta}} = (\overline{X}, S)$, where S^2 is the sample variance; $k_1(X_1) = X_1 - \mu$; $k_2(X_1) = \frac{1}{2\sigma}\left[(X_1-\mu)^2 - \sigma^2\right]$; $G^{(1)}(x;\boldsymbol{\theta}) = -\frac{1}{\sigma}g_0\left(\frac{x-\mu}{\sigma}\right)$; $G^{(2)}(x;\boldsymbol{\theta}) = \frac{\mu - x}{\sigma^2}g_0\left(\frac{x-\mu}{\sigma}\right)$.

5 Bootstrap approximation

If $F \neq G$, then the asymptotic variances $\sigma^2(\boldsymbol{\theta})$ and $\tau^2(\boldsymbol{\theta})$ in (7) and (10) cannot be calculated because the expressions depend on the unknown F and on $\boldsymbol{\theta}$. Therefore

we now establish the validity of bootstrap approximations as alternatives to the normal approximations. In this section we first deal with the case of a known $\boldsymbol{\theta}$ and show uniform consistency for the bootstrap estimator for the distribution of $n^{1/2}\left(U_{nc}(\boldsymbol{\theta})-\mu_{nc}(\boldsymbol{\theta})\right)$. In the next section we will deal with the analogous problem with estimated $\boldsymbol{\theta}$.
Let $(c_{in}^*, X_i^*) \equiv (c_{in}, X_i)^*$, $i=1,\dots,n$, be a random sample with replacement from the set of pairs $\{(c_{1n}, X_1),\dots,(c_{nn}, X_n)\}$ giving equal probability n^{-1} to each pair; and use as notation $Y_{in}(\boldsymbol{t}) = c_{in}G^2(X_i;\boldsymbol{t})$ and $Y_{in}^*(\boldsymbol{t}) = c_{in}^*G^2(X_i^*;\boldsymbol{t})$. Based on this resample we define the bootstrapped version of $U_{nc}(\boldsymbol{t})$ as

$$U_{nc}^*(\boldsymbol{t}) = \binom{n}{2}^{-1} \sum_{1\le i<j\le n} \max(G(X_i^*;\boldsymbol{t}), G(X_j^*;\boldsymbol{t})) - \frac{1}{n}\sum_{i=1}^n Y_{in}^*(\boldsymbol{t}).$$

The following theorem shows the uniform strong consistency of the bootstrap estimator $P^*(n^{1/2}(U_{nc}^*(\boldsymbol{\theta}) - U_{nc}(\boldsymbol{\theta})) \le x)$. We use the notation P^* and E^* for the conditional probability and expectation given (c_{in}, X_i), $i=1,\dots,n$.

Theorem 3. As $n\to\infty$,

$$\sup_{x\in\mathbb{R}} \left| P^*\left(n^{1/2}(U_{nc}^*(\boldsymbol{\theta}) - U_{nc}(\boldsymbol{\theta})) \le x\right) - P\left(n^{1/2}\left(U_{nc}(\boldsymbol{\theta}) - \mu_{nc}(\boldsymbol{\theta})\right) \le x\right)\right| \to 0 \text{ a.s.}$$

Proof. Write

$$\begin{aligned}
&U_{nc}^*(\boldsymbol{\theta}) - U_{nc}(\boldsymbol{\theta}) \\
&= \left\{ \binom{n}{2}^{-1} \sum_{1\le i<j\le n} \max(G(X_i^*;\boldsymbol{\theta}), G(X_j^*;\boldsymbol{\theta})) - \binom{n}{2}^{-1} \sum_{1\le i<j\le n} \max(G(X_i;\boldsymbol{\theta}), G(X_j;\boldsymbol{\theta})) \right\} \\
&- \left\{ \frac{1}{n}\sum_{i=1}^n Y_{in}^*(\boldsymbol{\theta}) - \frac{1}{n}\sum_{i=1}^n Y_{in}(\boldsymbol{\theta}) \right\} \\
&= \left(V_n^*(\boldsymbol{\theta}) - V_n(\boldsymbol{\theta})\right) - \left(W_{nc}^*(\boldsymbol{\theta}) - W_{nc}(\boldsymbol{\theta})\right). \qquad (11)
\end{aligned}$$

From the bootstrap theory for U-statistics (see e.g. Bickel and Freedman (1981), Janssen (1997), we have

$$V_n^*(\boldsymbol{\theta}) - V_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^n \psi_1(X_i^*) - \frac{1}{n}\sum_{i=1}^n \psi_1(X_i) + R_n^*$$

where $\psi_1(x) = 2\left(1 - \int_x^{\infty} F(y)dG(y;\boldsymbol{\theta}) - 2\int G(x;\boldsymbol{\theta})F(x)dF(x)\right)$ and $R_n^* = o_{P^*}(n^{-1/2})$ a.s. (meaning that, for all $\varepsilon > 0$, $P^*(n^{1/2}|R_n^*| > \varepsilon) \to 0$ a.s.).
Hence,

$$U_{nc}^*(\boldsymbol{\theta}) - U_{nc}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^n g_{in}^*(\boldsymbol{\theta}) - \frac{1}{n}\sum_{i=1}^n g_{in}(\boldsymbol{\theta}) + R_n^*$$

with $g_{in}^*(\boldsymbol{\theta}) = \psi_1(X_i^*) - Y_{in}^*(\boldsymbol{\theta})$ and $g_{in}(\boldsymbol{\theta}) = \psi_1(X_i) - Y_{in}(\boldsymbol{\theta})$.
From (6), with Φ the standard normal distribution function, we have that $\sup_{x\in\mathbb{R}}\left|P\left(n^{1/2}(U_{nc}(\boldsymbol{\theta}) - \mu_{nc}(\boldsymbol{\theta})) \le x\right) - \Phi\left(\frac{x}{\sigma(\boldsymbol{\theta})}\right)\right| \to 0$. Now, for each $\varepsilon > 0$,

$$\begin{aligned}
&\sup_{x\in\mathbb{R}}\left|P^*\left(n^{1/2}(U_{nc}^*(\boldsymbol{\theta}) - U_{nc}(\boldsymbol{\theta})) \le x\right) - \Phi\left(\frac{x}{\sigma(\boldsymbol{\theta})}\right)\right| \\
&\le \sup_{x\in\mathbb{R}}\left|P^*\left(n^{1/2}\left(\frac{1}{n}\sum_{i=1}^n g_{in}^*(\boldsymbol{\theta}) - \frac{1}{n}\sum_{i=1}^n g_{in}(\boldsymbol{\theta})\right) \le x\right) - \Phi\left(\frac{x}{\sigma(\boldsymbol{\theta})}\right)\right| \\
&+O(\varepsilon) + P^*\left(n^{1/2}|R_n^*| > \varepsilon\right). \qquad (12)
\end{aligned}$$

The first term in the right hand side of (12) can be rewritten and bounded above by

$$\sup_{x\in\mathbb{R}}\left|P^*\left(\frac{1}{\sqrt{n}S_n}\sum_{i=1}^n (g_{in}^*(\boldsymbol{\theta}) - \overline{g}_n) \le x\right) - \Phi(x)\right| + \sup_{x\in\mathbb{R}}\left|\Phi\left(\frac{x}{S_n}\right) - \Phi\left(\frac{x}{\sigma(\boldsymbol{\theta})}\right)\right| \quad (13)$$

where $\overline{g}_n = \frac{1}{n}\sum_{i=1}^n g_{in}(\boldsymbol{\theta})$ and $S_n^2 = \frac{1}{n}\sum_{i=1}^n (g_{in}(\boldsymbol{\theta}) - \overline{g}_n)^2$.

The second term in (13) tends to 0 almost surely since $S_n^2 \to \sigma^2(\boldsymbol{\theta})$ a.s. Indeed,

$$\begin{aligned}
S_n^2 &= \frac{1}{n}\sum_{i=1}^n \left[\psi_1(X_i) - \frac{1}{n}\sum_{j=1}^n \psi_1(X_j)\right]^2 \\
&+ \frac{1}{n}\sum_{i=1}^n \left[Y_{in}(\boldsymbol{\theta}) - \frac{1}{n}\sum_{j=1}^n Y_{jn}(\boldsymbol{\theta})\right]^2
\end{aligned}$$

$$- \quad \frac{2}{n}\sum_{i=1}^{n}\psi_1(X_i)\left[Y_{in}(\boldsymbol{\theta}) - \frac{1}{n}\sum_{j=1}^{n}Y_{jn}(\boldsymbol{\theta})\right]$$

$$\overset{a.s.}{\to} \quad E(\psi_1^2(X_1)) + c^2[E(G^4(X_1;\boldsymbol{\theta}) - (EG^2(X_1;\boldsymbol{\theta}))^2]$$

$$-2E[\psi_1(X_1)G^2(X_1;\boldsymbol{\theta})] = \sigma^2(\boldsymbol{\theta}), \text{ as in (7).}$$

Here we applied the result of Choi and Sung (1987) which, since the c_{in} are uniformly bounded, implies that $\frac{1}{n}\sum_{i=1}^{n} c_{in}Z_i \to E(Z_1)$ a.s. and $\frac{1}{n}\sum_{i=1}^{n} c_{in}^2 Z_i \to c^2 E(Z_1)$ a.s., for any sequence of i.i.d. random variables $Z_1, \dots, Z_n$ with $E|Z_1| < \infty$. For the first term in (13), we apply the Berry-Esseen bound (see e.g. Chung (1974)) to find that it is bounded above by a constant times Γ_n, where

$$\Gamma_n = \sum_{i=1}^{n} E^* \left|\frac{g_{in}^*(\boldsymbol{\theta}) - \overline{g}_n}{\sqrt{n}S_n}\right|^3 = \frac{n^{-3/2}}{S_n^3}\sum_{i=1}^{n}|g_{in}(\boldsymbol{\theta}) - \overline{g}_n|^3$$

$$\leq 4\frac{n^{-3/2}}{S_n^3}\left\{\sum_{i=1}^{n}|g_{in}(\boldsymbol{\theta})|^3 + n\,|\overline{g}_n|^3\right\}.$$

Now,

$$n^{-3/2}\sum_{i=1}^{n}|g_{in}(\boldsymbol{\theta})|^3 \leq 4n^{-3/2}\sum_{i=1}^{n}|\psi_1(X_i)|^3 + 4K^3 n^{-3/2}\sum_{i=1}^{n}G^6(X_i;\boldsymbol{\theta})$$

$$n^{-3/2}n\,|\overline{g}_n|^3 \leq 4n^{-7/2}\sum_{i=1}^{n}|\psi_1(X_i)|^3 + 4K^3 n^{-7/2}\sum_{i=1}^{n}G^6(X_i;\boldsymbol{\theta}).$$

All these terms tend to 0 a.s. by application of the law of large numbers. This, together with the fact that $S_n^2 \to \sigma^2(\boldsymbol{\theta})$ a.s., makes that $\Gamma_n \to 0$ a.s.

5 Bootstrap approximation in the case of estimated parameter

In this section we prove consistency of the bootstrap estimator for the distribution of the test statistic in which the nuisance parameter has been estimated.

Theorem 4. Assume condition (ii) of Theorem 1 and moreover

(1) the components of the estimator $\widehat{\boldsymbol{\theta}} = (\widehat{\theta}_1, \ldots, \widehat{\theta}_k)$ satisfy (9) with $E(k_j(X_1)) = 0$ and $\mathrm{Var}(k_j(X_1)) > 0$ for $j = 1, \ldots, k$.

(2) the components of the estimator $\widehat{\boldsymbol{\theta}}^* = (\widehat{\theta}_1^*, \ldots, \widehat{\theta}_k^*)$ satisfy

$$\widehat{\theta}j^* - \widehat{\theta}_j = \frac{1}{n}\sum_{i=1}^n k_j(X_i^*) - \frac{1}{n}\sum_{i=1}^n k_j(X_i) + r_{nj}^*$$

where for each $j = 1, \ldots, k$ and all $\varepsilon > 0 : P^*(n^{1/2}|r_{nj}^*| \geq \varepsilon) \xrightarrow{P} 0$.

(3) the partial derivatives $G^{(j)}$ are continuous at $\boldsymbol{\theta}$ and not all zero at $\boldsymbol{\theta}$.

Then, as $n \to \infty$,

$$\sup_{x \in I\!R} \left| P^* \left(n^{1/2}(U_{nc}^*(\widehat{\boldsymbol{\theta}}^*) - U_{nc}(\widehat{\boldsymbol{\theta}})) \leq x \right) - P \left(n^{1/2}(U_{nc}(\widehat{\boldsymbol{\theta}}) - \mu_{nc}(\boldsymbol{\theta})) \leq x \right) \right| \xrightarrow{P} 0.$$

Proof. From the proof of Theorem 1 and the definition of $\mu_{nc}^{(j)}(\theta)$ it follows that

$$U_{nc}(\widehat{\boldsymbol{\theta}}) - \mu_{nc}(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^n \left[g_{in}(X_i; \boldsymbol{\theta}) + \sum_{j=1}^k k_j(X_i)\mu_{nc}^{(j)}(\boldsymbol{\theta}) \right] + o_P(n^{-1/2}). \tag{14}$$

Following ideas in Liu, Singh and Lo (1989) it is therefore sufficient to show that

$$U_{nc}^*(\widehat{\boldsymbol{\theta}}^*) - U_{nc}(\widehat{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{i=1}^n \widetilde{g}_{in}^*(\boldsymbol{\theta}) - \frac{1}{n}\sum_{i=1}^n \widetilde{g}_{in}(\boldsymbol{\theta}) + \widetilde{R}_n^* \tag{15}$$

with $\widetilde{g}_{in}(\boldsymbol{\theta}) = g_{in}(\boldsymbol{\theta}) + \sum_{j=1}^k k_j(X_i)\mu_{nc}^{(j)}(\boldsymbol{\theta})$ and $\widetilde{g}_{in}^*(\boldsymbol{\theta}) = g_{in}^*(\boldsymbol{\theta}) + \sum_{j=1}^k k_j(X_i^*)\mu_{nc}^{(j)}(\boldsymbol{\theta})$, and where, for each $\varepsilon > 0$, $P^*(n^{1/2}|\widetilde{R}_n^*| \geq \varepsilon) \xrightarrow{P} 0$. To obtain the representation (15) we consider the following decomposition

$$\begin{aligned} &U_{nc}^*(\widehat{\boldsymbol{\theta}}^*) - U_{nc}(\widehat{\boldsymbol{\theta}}) \\ &= (U_{nc}^*(\boldsymbol{\theta}) - U_{nc}(\boldsymbol{\theta})) + (\mu_{nc}(\widehat{\boldsymbol{\theta}}^*) - \mu_{nc}(\widehat{\boldsymbol{\theta}})) + R_{n1}^* + R_{n2}^* \end{aligned} \tag{16}$$

with $R_{n1}^* = R_{n11}^* - R_{n12}^*$, $R_{n2}^* = R_{n21}^* - R_{n22}^*$ and

$$\begin{aligned} R_{n11}^* &= (V_n^*(\widehat{\boldsymbol{\theta}}^*) - V_n(\widehat{\boldsymbol{\theta}}^*)) - (V_n^*(\boldsymbol{\theta}) - V_n(\boldsymbol{\theta})) \\ R_{n12}^* &= (W_{nc}^*(\widehat{\boldsymbol{\theta}}^*) - W_{nc}(\widehat{\boldsymbol{\theta}}^*)) - (W_{nc}^*(\boldsymbol{\theta}) - W_{nc}(\boldsymbol{\theta})) \\ R_{n21}^* &= (V_n(\widehat{\boldsymbol{\theta}}^*) - \mu_1(\widehat{\boldsymbol{\theta}}^*)) - (V_n(\widehat{\boldsymbol{\theta}}) - \mu_1(\widehat{\boldsymbol{\theta}})) \\ R_{n22}^* &= (W_{nc}(\widehat{\boldsymbol{\theta}}^*) - \mu_{nc2}(\widehat{\boldsymbol{\theta}}^*)) - (W_{nc}(\widehat{\boldsymbol{\theta}}) - \mu_{nc2}(\widehat{\boldsymbol{\theta}})). \end{aligned}$$

Here V_n^*, V_n, W_{nc}^*, W_{nc} are the four functions appearing in the decomposition $U_{nc}^* - U_{nc} = (V_n^* - V_n) - (W_{nc}^* - W_{nc})$ in (11). Also μ_1 and μ_{nc2} are the two terms in the decomposition $\mu_{nc} = \mu_1 - \mu_{nc2}$ in (5). The i.i.d. sums in the right hand side of (15) are easily obtained from the i.i.d. representations for $U_{nc}^*(\boldsymbol{\theta}) - U_{nc}(\boldsymbol{\theta})$ and $\mu_{nc}(\widehat{\boldsymbol{\theta}}^*) - \mu_{nc}(\widehat{\boldsymbol{\theta}})$ in (16).

To show that the contributions of R_{n11}^* and R_{n21}^* are negligible we can apply Lemma 1 in Janssen and Veraverbeke (1992). It therefore remains to prove that

$$P^*(n^{1/2}|R_{n12}^*| \geq \varepsilon) \xrightarrow{P} 0 \tag{17}$$

and

$$P^*(n^{1/2}|R_{n22}^*| \geq \varepsilon) \xrightarrow{P} 0. \tag{18}$$

To establish (17) note that, with

$$Q_{nc}^*(\boldsymbol{s}) = n^{1/2}\frac{1}{n}\left\{\left[\sum_{i=1}^n Y_{in}^*\left(\boldsymbol{\theta} + \frac{\boldsymbol{s}}{\sqrt{n}}\right) - \sum_{i=1}^n Y_{in}\left(\boldsymbol{\theta} + \frac{\boldsymbol{s}}{\sqrt{n}}\right)\right]\right.$$
$$\left. - \left[\sum_{i=1}^n Y_{in}(\boldsymbol{\theta}) - \sum_{i=1}^n Y_{in}(\boldsymbol{\theta})\right]\right\},$$

we have $n^{1/2}R_{n12}^* = Q_{nc}^*(n^{1/2}(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}))$. With $D(\mathbf{0}, d)$ a sphere with radius d centered at zero we have for each $\varepsilon > 0$

$$\begin{aligned} &P^*(n^{1/2}|R_{n12}^*| \geq \varepsilon) \\ &\leq P^*\left(\sup_{\boldsymbol{s}\in D(\mathbf{0},d)} |Q_{nc}^*(\boldsymbol{s})| \geq \varepsilon\right) + P^*\left(n^{1/2}(\widehat{\boldsymbol{\theta}}^* - \boldsymbol{\theta}) \notin D(\mathbf{0}, d)\right). \end{aligned} \tag{19}$$

The second term in the right hand side of (19) is $o_P(1)$ (see Janssen and Veraverbeke (1992), p.1598 for details). To handle the first term in the right hand side of (19) we use the discretization argument explained on p.1598 in Janssen and Veraverbeke (1992), from which we also borrow the notation. This leads to

$$Q_{nc}^*(\boldsymbol{s}) = Q_{nc}^*(\delta\boldsymbol{r}) + Q_{r,nc}^*(\boldsymbol{s})$$

with

$$Q_{r,nc}^*(\boldsymbol{s}) = n^{1/2}\frac{1}{n}\left\{\left[\sum_{i=1}^n Y_{in}^*\left(\boldsymbol{\theta} + \frac{\boldsymbol{s}}{\sqrt{n}}\right) - \sum_{i=1}^n Y_{in}\left(\boldsymbol{\theta} + \frac{\boldsymbol{s}}{\sqrt{n}}\right)\right]\right.$$
$$\left. - \left[\sum_{i=1}^n Y_{in}^*\left(\boldsymbol{\theta} + \frac{\delta\boldsymbol{r}}{\sqrt{n}}\right) - \sum_{i=1}^n Y_{in}\left(\boldsymbol{\theta} + \frac{\delta\boldsymbol{r}}{\sqrt{n}}\right)\right]\right\}.$$

Therefore the first term in the right hand side of (19) tends to zero in probability if we show that

$$P^*\left(\sup_{\boldsymbol{s}\in D(\boldsymbol{v},\delta)}|Q^*_{r,nc}(\boldsymbol{s})|\geq\varepsilon\right)\overset{P}{\to}0 \tag{20}$$

and

$$P^*\left(|Q^*_{nc}(\delta\boldsymbol{r})|\geq\varepsilon\right)\overset{P}{\to}0, \tag{21}$$

where $D(\boldsymbol{v},\delta)$ denotes some sphere with radius δ centered at $\boldsymbol{v}$.
To show (20), rewrite $Q^*_{r,nc}(\boldsymbol{s})$ as follows:

$$Q^*_{r,nc}(\boldsymbol{s})=n^{1/2}\frac{1}{n}\sum_{i=1}^{n}\left\{h_{in}\left(\boldsymbol{\theta}+\frac{\boldsymbol{s}}{\sqrt{n}}\right)-h_{in}\left(\boldsymbol{\theta}+\frac{\delta\boldsymbol{r}}{\sqrt{n}}\right)\right\}$$

where $h_{in}(\boldsymbol{t})=Y_{in}(\boldsymbol{t})-\frac{1}{n}\sum_{j=1}^{n}Y_{jn}(\boldsymbol{t})$.
With $H_{inr}\left(\frac{\delta\boldsymbol{r}}{\sqrt{n}}\right)=\sup_{\boldsymbol{s}\in D(\boldsymbol{v},\delta)}\left|h_{in}\left(\boldsymbol{\theta}+\frac{\boldsymbol{s}}{\sqrt{n}}\right)-h_{in}\left(\boldsymbol{\theta}+\frac{\delta\boldsymbol{r}}{\sqrt{n}}\right)\right|$, we have that

$$\sup_{\boldsymbol{s}\in D(\boldsymbol{v},\delta)}\left|Q^*_{r,nc}(\boldsymbol{s})\right|\leq D_{nc1}+D_{nc2}$$

where

$$\begin{aligned}D_{nc1}&=n^{1/2}\frac{1}{n}\sum_{i=1}^{n}\left[H_{inr}\left(\frac{\delta\boldsymbol{r}}{\sqrt{n}}\right)-E^*H_{inr}\left(\frac{\delta\boldsymbol{r}}{\sqrt{n}}\right)\right]\\D_{nc2}&=n^{1/2}E^*H_{inr}\left(\frac{\delta\boldsymbol{r}}{\sqrt{n}}\right).\end{aligned}$$

Using the uniform boundedness of the c_{in}, together with the law of large numbers and condition (ii) of Theorem 1, we easily find that

$$D_{nc2}\leq 4Kn^{-1/2}E\left[\sup_{\boldsymbol{s}\in D(\boldsymbol{v},\delta)}\left|G^2\left(X_1;\boldsymbol{\theta}+\frac{\boldsymbol{s}}{\sqrt{n}}\right)-G^2\left(X_1;\boldsymbol{\theta}+\frac{\delta\boldsymbol{r}}{\sqrt{n}}\right)\right|\right]\leq 8KC\delta$$

and this can be made arbitrary small by appropriate choice of δ.
For D_{nc1} we use the inequality $P^*(|D_{nc1}|\geq\varepsilon)\leq\varepsilon^{-2}E^*(D^2_{nc1})=\varepsilon^{-2}E^*\left(H^2_{1nr}\left(\frac{\delta\boldsymbol{r}}{\sqrt{n}}\right)\right)$.
By the uniform boundedness of the c_{in}, we obtain that

$$E^*\left(H^*_{1nr}\left(\frac{\delta\boldsymbol{r}}{\sqrt{n}}\right)\right)\leq 4K^2\left(\frac{1}{n}\sum_{i=1}^{n}\sup_{\boldsymbol{s}\in D(\boldsymbol{v},\delta)}\left|G^2\left(X_i;\boldsymbol{\theta}+\frac{\boldsymbol{s}}{\sqrt{n}}\right)-G^2\left(X_i;\boldsymbol{\theta}+\frac{\delta\boldsymbol{r}}{\sqrt{n}}\right)\right|\right)^2$$

and this tends to zero in probability, by the law of large numbers and condition (ii) of Theorem 1.

To show (21), rewrite $Q^*_{nc}(\delta\boldsymbol{r})$ as follows:

$$Q^*_{nc}(\delta\boldsymbol{r}) = n^{1/2}\frac{1}{n}\sum_{i=1}^{n}\left\{h_{in}\left(\boldsymbol{\theta}+\frac{\delta\boldsymbol{r}}{\sqrt{n}}\right) - h_{in}(\boldsymbol{\theta})\right\}$$

with $h_{in}(\boldsymbol{t})$ as before.

To prove (21) it suffices to establish that $E^*\left[\left(h_{1n}\left(\boldsymbol{\theta}+\frac{\delta\boldsymbol{r}}{\sqrt{n}}\right) - h_{1n}(\boldsymbol{\theta})\right)^2\right] \xrightarrow{P} 0$, as $n \to \infty$. Using the uniform boundedness of the c_{in} we obtain that

$$E^*\left[\left(h_{1n}\left(\boldsymbol{\theta}+\frac{\delta\boldsymbol{r}}{\sqrt{n}}\right) - h_{1n}(\boldsymbol{\theta})\right)^2\right] \leq 2K^2\left\{\frac{1}{n}\sum_{i=1}^{n}\left(G^2\left(X_i;\boldsymbol{\theta}+\frac{\delta\boldsymbol{r}}{\sqrt{n}}\right) - G^2(X_i;\boldsymbol{\theta})\right)^2\right.$$

$$\left. + \left(\frac{1}{n}\sum_{i=1}^{n}\left(G^2\left(X_i;\boldsymbol{\theta}+\frac{\delta\boldsymbol{r}}{\sqrt{n}}\right) - G^2\left(X_i;\boldsymbol{\theta}\right)\right)\right)^2\right\}.$$

By condition (ii) of Theorem 1, this right hand side tends in probability to a quantity which is $O(\delta)+O(\delta^2)$, and this can be made arbitrary small by choice of δ.

To establish (18) note that

$$P^*(n^{1/2} \mid R^*_{n22} \mid \geq \varepsilon)$$

$$\leq P^*\left(n^{1/2}\left|(W^*_{nc}(\widehat{\boldsymbol{\theta}}^*) - \mu_{nc2}(\widehat{\boldsymbol{\theta}}^*)) - (W_{nc}(\boldsymbol{\theta}) - \mu_{nc2}(\boldsymbol{\theta}))\right| \geq \frac{\varepsilon}{2}\right)$$

$$+I\left\{\left|(W_{nc}(\widehat{\boldsymbol{\theta}}) - \mu_{nc2}(\widehat{\boldsymbol{\theta}})) - (W_{nc}(\boldsymbol{\theta}) - \mu_{nc2}(\boldsymbol{\theta})\right| \geq \frac{\varepsilon}{2}\right).\qquad(22)$$

The second term in the right hand side of (22) tends to zero in probability. The first term in the right hand side also tends to zero in probability as follows, from a modification (allowing weights) of the argument used on p.1602 in Janssen and Veraverbeke (1992).

Acknowledgement
This work was supported by the Ministry of the Flemish Community (Project BIL00/28, International Scientific and Technological Cooperation) by the Interuniversity Attraction Poles research network nr P5/24 of the Belgian State (Federal Office for Scientific, Technical and Cultural Affairs) and by the National Research Foundation of South Africa.

References

Ahmad, I.A., 1993. Modification of some goodness of fit statistics to yield asymptotically normal null distribution, Biometrika **80**, 466-472.

Ahmad, I.A., 1996. Modification of some goodness of fit statistics II: two-sample and symmetry testing, Sankhyā Series A **58**, 464-472.

Ahmad, I.A. and Alwasel, I.A., 1999. A goodness-of-fit test for exponentiality based on the memoryless property, J.R. Statist. Soc. B **61**, 681-689.

Bickel, P.J. and Freedman, D.A., 1981. Some asymptotic theory for the bootstrap, Ann. Statist. **9**, 1196-1217.

Choi, B.D. and Sung, S.H., 1987. Almost sure convergenve theorems of weighted sums of random variables, Stochastic Anal. **5**, 365-377.

Chung, K.L., 1974. A course in Probability Theory, 2nd ed., Academic Press, New York.

De Wet, T. and Randles, R.H., 1987. On the effect of substitution parameter estimators in limiting χ^2 U and V statistics, Ann. Statist. **15**, 398-412.

Janssen, P., 1997. Bootstrapping U-statistics, South African Statistical Journal **31**, 185-216.

Janssen, P. and Veraverbeke, N., 1992. Bootstrapping U-statistics with estimated parameters, Commun. Statist. - Theory Meth. **21**, 1585-1603.

Jiménez-Gamero, M.D., Muñoz-García, J. and Pino-Mejías, R., 2003. Bootstrapping parameter estimated degenerate U and V statistics, Statist. Probab. Lett. **61**, 61-70.

Liu, R.Y., Singh, K. and Lo, S.-H., 1989. On a representation related to the bootstrap, Sankhyā, Ser. A, **51**, 168-177.

Randles, R.H., 1982. On the asymptotic normality of statistics with estimated parameters, Ann. Statist. **10**, 462-474.

Serfling, R.J., 1980. Approximations Theorems of Mathematical Statistics, John Wiley, New York.

Too, Y-H. and Lin, G.D., 1989. Characterizations of uniform and exponential distributions, Statist. Probab. Lett. **7**, 357-359