RANDOMIZED GOODNESS OF FIT TESTS

F. Liese and B. Liu[1]

Abstract: Goodness of fit tests based on the empirical distribution function fail to be asymptotically distribution free if parameters are estimated. Asymptotically, the empirical process with estimated parameters is a centered Gaussian process with a covariance that differs from the covariance function of the Brownian bridge. If the maximum likelihood estimator is used then the new covariance function is smaller, in the Loewner semiorder, than the covariance function of the Brownian bridge. Therefore one may transform the empirical process with estimated parameters back to a Brownian bridge by adding an independent process that is suitably constructed. Classical goodness of fit statistics have aftzer this transformation an asymptotic distribution as in the case of known parameters. The power under local alternatives of this new goodness of fit tests is studied and computer simulations compare the new test with their bootstrap versions.
Keywords: Goodness of fit tests with estimated parameters, Kolmogorov-Smirnov test, Cramer-von Mises test, Bootstrap
AMS Classification: 62E17; 62E20

1 Empirical distribution functions with estimated parameters

For independent and uniformly on $[0,1]$ distributed random variables $U_1, U_2, ...$ we denote by

$$\mathsf{U}_n(t) = \frac{1}{\sqrt{n}} \sum\nolimits_{i=1}^{n} (I_{[0,t]}(U_i) - t) \tag{1}$$

the uniform empirical process which is a random element of the Skorokhod space $\mathbb{D}[0,1]$ that we equip with the Skorokhod metric under which $\mathbb{D}[0,1]$ is a complete and separable metric space, see Billingsley (1968). For i.i.d. $X_1, X_2, ...$with common distribution F we denote by

$$\widehat{F}_n(t) = \frac{1}{n} \sum\nolimits_{i=1}^{n} I_{(-\infty,t]}(X_i), \quad \text{and} \quad \mathsf{G}_n = \sqrt{n}(\widehat{F}_n - F).$$

[1]University of Rostock, Department of Mathematics, D-18051 Rostock, Universitaetsplatz 1. e-mail: friedrich.liese@uni-rostock.de, bing.liu@uni-rostock.de

the empirical distribution function and the empirical process, respectively. G_n is a random element of the Skorokhod space $\mathbb{D}[-\infty,\infty]$ that is again equipped with the Skorokhod metric. The Donsker theorem, see p. 97 in Pollard (1984), originally proved for $\mathbb{D}[0,1]$, states in the general situation,

$$\mathcal{L}(\mathsf{G}_n) \Rightarrow \mathcal{L}(\mathsf{B}(F)),$$

where $\Rightarrow$ is the symbol of weak convergence of distributions on the Skorokhod space $\mathbb{D}[-\infty,\infty]$, and B is the Brownian bridge on $[0,1]$. This means that B is a centered continuous Gaussian process with covariance function

$$E(\mathsf{B}(s)\mathsf{B}(t)) = s \wedge t - st.$$

For a given parametric family $(P_\theta)_{\theta\in\Theta}$ of distributions we want to test whether the common distribution of the i.i.d. sample $X_1, ..., X_n$ originates from the model $(P_\theta)_{\theta\in\Theta}$. To this end we set $F_\theta(t) = P_\theta((-\infty,t])$ and compare the empirical distribution function $\widehat{F}_n$ with the estimation obtained by plugging in an estimator $\widehat{\theta}_n$ into F_θ. This leads to the estimated empirical process

$$\widehat{\mathsf{G}}_n = \sqrt{n}(\widehat{F}_n - F_{\widehat{\theta}_n}).$$

The asymptotic distribution of $\widehat{\mathsf{G}}_n$ has been established by many authors starting with Durbin (1973a) and (1973b). The results of different authors differ in the type of regularity conditions that are necessary to make a suitable Taylor expansion, see Shorack and Wellner (1986), Section 5.5. Our approach follows van der Vaart (1998) and Genz and Haeusler (2006). We suppose that Θ is an open subset of $\mathbb{R}^d$ and the sequence of estimators $\widehat{\theta}_n$ is strongly consistent, i.e.

$$\widehat{\theta}_n(X_1, ..., X_n) \to \theta \quad a.s. \tag{2}$$

Moreover we assume that $\widehat{\theta}_n$ admits a first order Taylor expansion in the sense that there exists a measurable function $h_\theta : \mathbb{R} \to \mathbb{R}^d$ such that

$$\sqrt{n}(\widehat{\theta}_n(X_1, ..., X_n) - \theta) = \frac{1}{\sqrt{n}} \sum\nolimits_{i=1}^{n} h_\theta(X_i) + o_P(1) \tag{3}$$

$$E_\theta \left\| h_\theta(X_1) \right\|^2 < \infty, \quad E_\theta h_\theta(X_1) = 0 \tag{4}$$

$$J(\theta) := E_\theta h_\theta(X_1) h_\theta^T(X_1). \tag{5}$$

Here $\|x\|$ denotes the Euclidean norm of the column vector x and the superscript T is the symbol for transposition. We suppose that for every $\theta_0 \in \Theta$ there is a neighborhood $U(\theta_0) \subseteq \Theta$ such that $\theta \longmapsto F_\theta(t), \theta \in U(\theta_0)$ is differentiable and

$$\dot{F}_\theta(t) = (\frac{\partial}{\partial\theta_1} F_\theta(t), ..., \frac{\partial}{\partial\theta_d} F_\theta(t))^T. \tag{6}$$

is a continuous function $\theta \in U(\theta_0), -\infty \leq t \leq \infty$. Furthermore we suppose that

$$\sup_{t\in[-\infty,\infty],\theta\in U(\theta_0)} |F_{\theta+h}(t) - F_\theta(t) - \dot{F}_\theta^T(t)h| = o(\|h\|) \quad \text{as } h \to 0, \tag{7}$$

The next version of Durbin's Theorem was proved in van der Vaart (1998).

Theorem 1 *Under the assumptions (3), (4) and (7) it holds*

$$\sup_t \left|\widehat{\mathsf{G}}_n(t) - n^{-1/2}\sum\nolimits_{i=1}^n (I_{(-\infty,t]}(X_i) - F_\theta(t) - \dot{F}_\theta^T(t)h_\theta(X_i))\right| = o_P(1) \tag{8}$$

and

$$\mathcal{L}(\widehat{\mathsf{G}}_n) \Rightarrow \mathcal{L}(Z),$$

where Z is a centered and continuous Gaussian process with covariance function

$$\begin{aligned} cov(Z(s), Z(t)) &= F_\theta(s \wedge t) - F_\theta(s)F_\theta(t) + \dot{F}_\theta^T(s)J(\theta)\dot{F}_\theta(t) \\ &\quad -\dot{F}_\theta^T(s)Eh_\theta(X_1)I_{(-\infty,t]}(X_1) - \dot{F}_\theta^T(t)Eh_\theta(X_1)I_{(-\infty,s]}(X_1) \end{aligned} \tag{9}$$

for every $s, t \in [-\infty, \infty]$.

We analyze the covariance function in the case if $\widehat{\theta}_n$ is the maximum likelihood estimator (MLE). To this end it is supposed that the family $(P_\theta)_{\theta\in\Theta}$ is dominated by a σ-finite measure and atomless measure μ. Denote by $f_\theta = dP_\theta/d\mu, \theta \in \Theta$, the corresponding densities. As μ is atomless the distribution functions

$$F_\theta(t) = \int I_{(-\infty,t]}(s)f_\theta(s)\mu(ds) \tag{10}$$

are continuous in t. We impose the following conditions on the densities

$$\begin{aligned} & f_\theta(x) = \tfrac{dP_\theta}{d\mu}(x) > 0 \quad \mu - a.s \text{ and } \theta \in \Theta \\ & \theta \longmapsto f_\theta\,(x) \text{ is continuously differentiable for every } x \\ & \int \left\|\dot{f}_\theta(x)\right\|^2 \tfrac{1}{f_\theta(x)}\mu(dx) < \infty \\ & \theta \to I(\theta) = \int \dot{f}_\theta(x)\dot{f}_\theta^T(x)\tfrac{1}{f_\theta(x)}\mu(dx) \text{ is continuous,} \\ & \det(I(\theta)) \neq 0 \text{ for every } \theta \in \Theta \\ & \int I_{(-\infty,t]}(x)\dot{f}_\theta(x)\mu(dx) = \dot{F}_\theta(t), \quad -\infty < t < \infty. \end{aligned} \tag{11}$$

where $\dot{f}_\theta := (\frac{\partial}{\partial\theta_1}f_\theta, ..., \frac{\partial}{\partial\theta_d}f_\theta)^T$. The last condition in (11) means that one may interchange the derivative with respect to θ and the integral with respect to

x in (10). Moreover $I(\theta)$ is the Fisher information matrix. If (11) is fulfilled then under weak additional conditions the MLE $\widehat{\theta}_n$ satisfies (3) with

$$h_\theta = I^{-1}(\theta)\dot{l}_\theta \quad \text{where} \quad \dot{l}_\theta = (\dot{l}_{1,\theta}, ..., \dot{l}_{d,\theta})^T := \frac{1}{f_\theta}\dot{f}_\theta. \tag{12}$$

We consider the covariance function in Theorem 1. It holds

$$\begin{aligned}
& \dot{F}_\theta^T(t) E h_\theta(X_1) I_{(-\infty,t]}(X_1) \\
= \; & \dot{F}_\theta(t)^T \int I_{(-\infty,t]}(x) I^{-1}(\theta) \dot{l}_\theta(x) P_\theta(dx) \\
= \; & \dot{F}_\theta(t)^T \int I_{(-\infty,t]}(x) I^{-1}(\theta) \frac{\dot{f}_\theta(x)}{f_\theta(x)} f_\theta(x) \mu(dx) \\
= \; & \dot{F}_\theta(t)^T I^{-1}(\theta) \int I_{(-\infty,t]}(x) \dot{f}_\theta(x) \mu(dx) \\
= \; & \dot{F}_\theta(t)^T I^{-1}(\theta) \dot{F}_\theta(t),
\end{aligned}$$

where the last equality follows from the last condition in (11). Hence the covariance function in (9) turns into

$$cov(Z(s), Z(t)) = F_\theta(s \wedge t) - F_\theta(s)F_\theta(t) - \dot{F}_\theta^T(t) I^{-1}(\theta) \dot{F}_\theta(s). \tag{13}$$

Corollary 2 *Suppose that the conditions in (11) are satisfied, and the MLE $\widehat{\theta}_n$ satisfies (3) with $h_\theta = I^{-1}(\theta)\dot{l}_\theta$. If (7) holds then*

$$\mathcal{L}(\widehat{\mathsf{G}}_n) \Rightarrow \mathcal{L}(Z),$$

where Z is a centered and continuous Gaussian process whose covariance function is given in (13).

It is easy to see that covariance matrix $cov(Z(t_i), Z(t_j)), i, j = 1, ..., n$ is not larger in the Loewner semiorder of matrices than the covariance matrix $F_\theta(t_i \wedge t_j) - F_\theta(t_i)F_\theta(t_j)$, $i, j = 1, ..., n$. This means that we can eliminate the additional term $\dot{F}_\theta^T(t) I^{-1}(\theta) \dot{F}_\theta(s)$ on the right hand side of (13) by adding a suitable process to $\widehat{\mathsf{G}}_n$. Subsequently $I^{-1/2}(\theta)$ stands for the positive definite symmetric matrix with $I^{-1/2}(\theta) I^{-1/2}(\theta) = I^{-1}(\theta)$. We set for standard normal random vectors V_n that are independent of $X_1, X_2, ...$ and a consistent estimator $\widetilde{\theta}_n$

$$\begin{aligned}
\mathsf{R}_n(\theta) &= \widehat{\mathsf{G}}_n + \dot{F}_\theta^T I^{-1/2}(\theta) V_n & (14) \\
\mathsf{R}_n(\widetilde{\theta}_n) &= \widehat{\mathsf{G}}_n + \dot{F}_{\widetilde{\theta}_n}^T I^{-1/2}(\widetilde{\theta}_n) V_n. & (15)
\end{aligned}$$

Theorem 3 *Suppose the conditions in Corollary 2 are fulfilled, V_n are standard normal random vectors that are independent of $X_1, X_2, ...$ and $\widetilde{\theta}_n$ is a consistent estimator. Then*

$$\mathcal{L}(\mathsf{R}_n(\theta)) \Rightarrow \mathcal{L}(\mathsf{B}(F_\theta)).$$

and

$$\mathcal{L}(\mathsf{R}_n(\widetilde{\theta}_n)) \Rightarrow \mathcal{L}(\mathsf{B}(F_\theta)).$$

Proof. Suppose that Z is a centered Gaussian process that is independent of V and has the covariance function in (13) whereas V is any standard normal random vector that is independent of Z. Then the convergence

$$\mathcal{L}(\widehat{\mathsf{G}}_n + \dot{F}_\theta^T I^{-1/2}(\theta)V_n) \Longrightarrow \mathcal{L}(Z + \dot{F}_\theta^t I^{-1/2}(\theta)V)$$

is obvious. It remains to calculate the covariance function of the centered Gaussian process $Z + \dot{F}_\theta^t I^{-1/2}(\theta)V$. The independence of Z and V yields

$$\begin{aligned} & cov(Z(s) + \dot{F}_\theta^T(s)I^{-1/2}(\theta)V, Z(t) + \dot{F}_\theta^T(t)I^{-1/2}(\theta)V) \\ = \; & cov(Z(s), Z(t)) + cov(\dot{F}_\theta^T(s)I^{-1/2}(\theta)V), \dot{F}_\theta^T(t)I^{-1/2}(\theta)V) \\ = \; & cov(Z(s), Z(t)) + E(\dot{F}_\theta^T(s)I^{-1/2}(\theta)V))(\dot{F}_\theta^T(t)I^{-1/2}(\theta)V)^T \\ = \; & cov(Z(s), Z(t)) + \dot{F}_\theta^T(s)I^{-1}(\theta)\dot{F}_\theta(t) \\ = \; & F_\theta(s \wedge t) - F_\theta(s)F_\theta(t) \end{aligned}$$

where the last equality follows from (13). The second statement follows from Slutsky's lemma. ■

Technically it is sometimes more appropriate to deal with the interpolated empirical distribution function. More precisely, let $\widetilde{F}_n(t)$ be any piecewise linear continuous function that satisfies

$$\widehat{F}_n(t-0) = \frac{1}{n}\sum\nolimits_{i=1}^n I_{(-\infty,t)}(X_i) \le \widetilde{F}_n(t) \le \widehat{F}_n(t) = \frac{1}{n}\sum\nolimits_{i=1}^n I_{(-\infty,t]}(X_i).$$

Set $\widetilde{\mathsf{G}}_n = \sqrt{n}(\widetilde{F}_n - F_{\widehat{\theta}_n})$. Then

$$\sup_t |\widehat{F}_n(t) - \widetilde{F}_n(t)| \le \frac{1}{\sqrt{n}} \quad \text{and} \quad \sup_t |\widetilde{\mathsf{G}}_n - \widehat{\mathsf{G}}_n| \le \frac{1}{\sqrt{n}}.$$

It is clear that $\widetilde{\mathsf{G}}_n$ may be considered as a random element of $\mathbb{C}[-\infty, \infty]$ that we equip with the sup-norm and the σ-algebra of Borel sets. The mapping $T(f) = \sup_t |f(t)|$ is continuous on $\mathbb{C}[-\infty, \infty]$ and the continuous mapping theorem yields that under the assumptions of Theorem 3

$$\mathcal{L}(\sup_t |\widetilde{\mathsf{G}}_n(t) + \dot{F}_{\widetilde{\theta}_n}^T(t)I^{-1/2}(\widetilde{\theta}_n)V|) \Longrightarrow \mathcal{L}(\sup_t |\mathsf{B}(F_{\theta_0}(t))|) = \mathcal{L}(\sup_s |\mathsf{B}(s)|) =: \mathcal{K}$$

where $\mathcal{K}$ is the Kolmogorov distribution with the distribution function

$$\mathcal{K}([0,x]) = 1 - 2\sum\nolimits_{k=1}^{\infty}(-1)^{k+1}\exp\{-2k^2x^2\}. \tag{16}$$

For $\mathsf{R}_n(\widetilde{\theta}_n)$ in (15) we introduce the *randomized Kolmogorov-Smirnov statistic* by

$$\mathsf{K}_n = \sup_t |\widehat{\mathsf{G}}_n(t) + \dot{F}^T_{\widetilde{\theta}_n}(t)I^{-1/2}(\widetilde{\theta}_n)V_n| \tag{17}$$

Then under the assumptions of Theorem 3

$$\mathcal{L}(\mathsf{K}_n) \Rightarrow \mathcal{K} \tag{18}$$

A similar statement holds for statistics of the Cramer von Mises type. Denote by $X_{n:1} \leq \cdots \leq X_{n:n}$ the order statistic and introduce the *randomized Cramer von Mises statistic* by

$$\mathsf{C}_n := \sum\nolimits_{i=1}^{n}((\frac{i}{n} - F_{\widehat{\theta}_n}(X_{n:i}) + \frac{1}{\sqrt{n}}\dot{F}^T_{\widetilde{\theta}_n}(X_{n:i})I^{-1/2}(\widetilde{\theta}_n)V_n)^2. \tag{19}$$

To study the asymptotic behavior of C_n we note that for any compact $K \subseteq \mathbb{C}[-\infty,\infty]$ and i.i.d. $X_1, X_2, ...$ with common distribution F_θ the Glivenko-Cantelli Theorem gives

$$\sup_{f\in K}\left|\int f d\widehat{F}_n - \int f dP_\theta\right| \to 0 \quad a.s. \tag{20}$$

The continuity of F_θ yields $\widehat{F}_n(X_{n:i}) = i/n$ a.s. and

$$\begin{aligned} \mathsf{C}_n &= \sum\nolimits_{i=1}^{n}((\frac{i}{n} - F_{\widehat{\theta}_n}(X_{n:i}) + \frac{1}{\sqrt{n}}\dot{F}^T_{\widetilde{\theta}_n}(X_{n:i})I^{-1/2}(\widetilde{\theta}_n)V_n)^2 \qquad (21) \\ &= \int(\sqrt{n}(\widehat{F}_n(t) - F_{\widehat{\theta}_n}(t)) + \dot{F}^T_{\widetilde{\theta}_n}(t)I^{-1/2}(\widetilde{\theta}_n)V_n)^2 d\widehat{F}_n(t). \end{aligned}$$

The tightness of the distributions of the sequence $\sqrt{n}(\widehat{F}_n - F_\theta - \dot{F}^T_{\widetilde{\theta}_n} I^{-1/2}(\widetilde{\theta}_n)V_n)$ and (20) give

$$\begin{aligned} \mathsf{C}_n &= \int(\sqrt{n}(\widehat{F}_n(t) - F_{\widehat{\theta}_n}(t)) + \dot{F}^T_{\widetilde{\theta}_n}(t)I^{-1/2}(\widetilde{\theta}_n)V_n)^2 P_\theta(dt) + o_P(1) \\ &= \int(\widetilde{\mathsf{G}}_n(t) - \dot{F}^T_{\widetilde{\theta}_n}(t)I^{-1/2}(\widetilde{\theta}_n)V_n)^2 P_\theta(dt) + o_P(1) \qquad (22) \end{aligned}$$

Since $f \mapsto \int f^2 dP_\theta$ is a continuous function on $\mathbb{C}[-\infty,\infty]$ we get from Theorem 3

$$\mathcal{L}(\mathsf{C}_n) \Rightarrow \mathcal{L}(\int_0^1 \mathsf{B}^2(s)ds) =: \mathcal{C}, \tag{23}$$

where we have used the fact that the continuity of F_θ implies

$$\int \mathsf{B}^2(F_\theta)dP_\theta = \int_0^1 \mathsf{B}^2(s)ds.$$

It is well known, see e.g. Shorack and Wellner (1986). p. 215, that

$$\mathcal{C}([0,x]) = P(\sum\nolimits_{k=1}^{\infty} \frac{1}{(k\pi)^2} Z_k^2 \leq x), \tag{24}$$

where the $Z_1, Z_2, ...$ are i.i.d. standard normal. We denote by $k_{1-\alpha}$ and $c_{1-\alpha}$ the $1-\alpha$ quantile of the Kolmogorov distribution $\mathcal{K}$ and the Cramer-von Mises distribution $\mathcal{C}$, respectively. Based on the tests statistic K_n we introduce the *randomized Kolmogorov-Smirnov test* and the *randomized Cramer-von Mises test* by

$$\varphi_{\mathsf{K}_n} = I_{[k_{1-\alpha},\infty)}(\mathsf{K}_n) \quad \text{and} \quad \varphi_{\mathsf{C}_n} = I_{[c_{1-\alpha},\infty)}(\mathsf{C}_n). \tag{25}$$

The next statement is a simple consequence of Theorem 3 and the relations (23) and (18).

Proposition 4 *Under the assumptions of Theorem 3 the tests φ_{K_n} and φ_{C_n} are asymptotic level $\alpha-$tests for testing*

$$\mathsf{H}_0 : \mathcal{L}(X_1) \in \{P_\theta, \theta \in \Theta\} \quad \textit{versus} \quad \mathsf{H}_A : \mathcal{L}(X_1) \notin \{P_\theta, \theta \in \Theta\}.$$

2 Power under local alternatives

Suppose Q_n and Q are distributions on $(\mathcal{X}, \mathfrak{A})$, assume $Q_n \ll Q$ and put $g_n = dQ_n/dQ$,

$$a_n = 2\sqrt{n}(\sqrt{g_n} - 1). \tag{26}$$

Then

$$\begin{aligned} \int (1 + \frac{1}{2\sqrt{n}} a_n)^2 dQ &= 1 \\ \int a_n dQ &= -\frac{1}{4\sqrt{n}} \int a_n^2 dQ. \end{aligned} \tag{27}$$

Suppose there is some $a \in L_2(Q)$ with

$$\lim_{n\to\infty} \int (a_n(x) - a(x))^2 Q(dx) = 0. \tag{28}$$

Then (27) yields.

$$\int adQ = 0. \tag{29}$$

Denote by $X_1, ..., X_n$ the projections of $\mathcal{X}^n$ onto $\mathcal{X}$. The next statement is a special case of Theorem 3 in Shorack and Wellner (1986). p. 154. Under the assumption (28) the sequence $Q_n^{\otimes n}$ satisfies the LAN-condition in the sense that

$$\begin{aligned}\ln L_n &= \frac{1}{\sqrt{n}} \sum\nolimits_{i=1}^{n} a(X_i) - \frac{1}{2} \int_0^1 a^2(x) Q(dx) + o_{Q^{\otimes n}}(1). \quad (30)\\ L_n &= \frac{dQ_n^{\otimes n}}{dQ^{\otimes n}}(X_1, ..., X_n).\end{aligned}$$

We study the asymptotic behavior of the linear statistics

$$T_n = \frac{1}{\sqrt{n}} \sum\nolimits_{i=1}^{n} b_n(X_i), \tag{31}$$

where $b_n, b \in L_2^0(Q)$ with

$$\lim_{n\to\infty} \int (b_n - b)^2 dQ = 0.$$

The representations (30), (31) and the central limit theorem yield

$$\mathcal{L}((T_n, \ln L_n)^T | Q^{\otimes n}) \Rightarrow \mathsf{N}\left(\left(\begin{array}{c} 0 \\ -\frac{1}{2}\int_0^1 a^2 dQ \end{array} \right), \left(\begin{array}{cc} \int_0^1 b^2 dQ & \int abdQ \\ \int abdQ & \int_0^1 a^2 dQ \end{array} \right) \right).$$

From here and LeCam's third Lemma , see Theorem 4, in Shorack and Wellner (1986), p. 154, we get the asymptotic distribution of T_n under $Q_n^{\otimes n}$

$$\mathcal{L}(T_n | Q_n^{\otimes n}) \Rightarrow \mathsf{N}(\int abdQ, \int b^2 dQ). \tag{32}$$

Let us return to the parametric model $(P_\theta)_{\theta\in\Theta}$, set $Q = P_\theta$ for any fixed θ and suppose that there is a sequence $Q_n \ll Q$ such that (28) is satisfied. Our aim is to study the randomized Kolmogorov-Smirnov and Cramer-von Mises statistics under $Q_n^{\otimes n}$. For any real numbers a_j it holds under the assumptions of Theorem 3

$$\begin{aligned}&\sum\nolimits_{j=1}^{n} a_j \widehat{\mathsf{G}}_n(t_j)\\ &= \frac{1}{\sqrt{n}} \sum\nolimits_{i=1}^{n} (\sum\nolimits_{j=1}^{n} a_j (I_{(-\infty,t_j]}(X_i) - F_\theta(t_j) - \dot{F}_\theta^T(t_j) I^{-1}(\theta) \dot{l}_\theta(X_i)) + o_P(1)\\ &= \frac{1}{\sqrt{n}} \sum\nolimits_{i=1}^{n} b(X_i) + o_P(1),\end{aligned}$$

where

$$\begin{aligned} b(s) &= \sum_{j=1}^{n} a_j \Lambda_\theta(t_j, s) \\ \Lambda_\theta(t,s) &= I_{(-\infty,t]}(s) - F_\theta(t) - \dot{F}_\theta^T(t) I^{-1}(\theta) \dot{l}_\theta(s). \end{aligned} \tag{33}$$

An application of the central limit theorem to

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (a(X_i), b(X_i))^T$$

and (32) provide the fidi-convergence of $\widehat{\mathsf{G}}_n$ under $Q_n^{\otimes n}$ to the process $Z + \Lambda_\theta a$ where Z is a Gaussian process with covariance function (13) and

$$(\Lambda_\theta a)(t) = \int \Lambda_\theta(t,s) a(s) P_\theta(ds). \tag{34}$$

The LAN-property (30) and LeCam's first lemma imply the contiguity of $Q_n^{\otimes n}$ with respect to $Q^{\otimes n}$, see Shorack and Wellner (1986), p.157. We get that the sequence of distributions of $\widehat{\mathsf{G}}_n$ is tight under $Q_n^{\otimes n}$ as well, where tightness is to be understood with respect to the modulus of continuity of the Skorokhod space, see Pollard (1984), p.131. Using the independence of $\widehat{\mathsf{G}}_n$ and $\dot{F}_\theta^T I^{-1/2}(\theta) V_n$ and the tightness of $\dot{F}_\theta^T I^{-1/2}(\theta) V_n$ we arrive at the following result.

Theorem 5 *If the assumptions in Theorem 3 are fulfilled and the condition (28) holds then*

$$\mathcal{L}(\widehat{\mathsf{G}}_n + \dot{F}_{\widetilde{\theta}_n}^T I^{-1/2}(\widetilde{\theta}_n) V_n | Q_n^{\otimes n}) \Longrightarrow \mathcal{L}(\mathsf{B}(F_\theta) + \Lambda_\theta a),$$

where $\Lambda_\theta a$ is defined in (34).

If the distribution of the local alternative Q_n comes from inside the model, i.e. $Q_n = P_{\theta + h/\sqrt{n}}$, then by a result of Hajek the conditions in (11) imply that the model $(P_\theta)_{\theta \in \Theta}$ is L_2-differentiable with derivative $\dot{l}_\theta$, see Bickel et al. (1993), p.13. Then by Shorack and Wellner (1986), p. 157 it follows that (30) is satisfied with $a = h^T \dot{l}_\theta$. Furthermore,

$$\begin{aligned} (\Lambda_\theta h^T \dot{l}_\theta)(t) &= \int [I_{(-\infty,t]}(s) - F_\theta(t) - \dot{F}_\theta^T(t) I^{-1}(\theta) \dot{l}_\theta(s)] h^T \dot{l}_\theta(s) P_\theta(ds) \\ &= h^T \dot{F}_\theta(t) - h^T \dot{F}_\theta(t) = 0, \end{aligned}$$

where the second equality follows from

$$\begin{aligned} \int \dot{l}_\theta(s) P_\theta(ds) &= 0 \\ \int (I_{(-\infty,t]}(s) \dot{l}_\theta^T(s) P_\theta(ds) &= \dot{F}_\theta^T(t) \end{aligned}$$

which are consequences of (11). Thus we have obtained the following result.

Proposition 6 *Suppose the assumptions in Theorem 3 are fulfilled and* $Q_n = P_{\theta+h/\sqrt{n}}$ *then*

$$\mathcal{L}(\widehat{\mathsf{G}}_n + \dot{F}^T_{\widetilde{\theta}_n} I^{-1/2}(\widetilde{\theta}_n)V_n | Q_n^{\otimes n}) \Longrightarrow \mathcal{L}(\mathsf{B}(F_\theta)).$$

We see that asymptotic level α tests that are based on $\mathsf{R}_n(\widetilde{\theta}_n)$ in (15) have only the power α as long as the local alternatives come from inside of the model, i.e. $Q_n \in \{P_\theta, \theta \in \Theta\}$.

We recall to the Kac-Siegert decomposition of the Brownian bridge $\mathsf{B}(t), 0 \leq t \leq 1$ which has the covariance function $K(s,t) = s \wedge t - st$. The eigenvalues and the normalized eigenfunctions of this kernel are

$$\lambda_k = \frac{1}{(k\pi)^2} \quad \text{and} \quad \varphi_k(t) = \sqrt{2}\sin k\pi t, \quad k = 1, 2, ...$$

The Kac-Siegert decomposition of the Brownian bridge reads

$$\mathsf{B}(t) = \sqrt{2}\sum\nolimits_{k=1}^{\infty} Z_k \frac{\sin k\pi t}{k\pi}.$$

where $Z_i = k\pi \int_0^1 \mathsf{B}(t)\varphi_k(t)dt$ and the $Z_1, Z_2, ...$ are i.i.d. standard normal. As the system of eigenfunctions $\{\varphi_k\}$ is complete we may expand any square integrable function c in a Fourier series

$$c(t) = \sum\nolimits_{k=1}^{\infty} c_k \sin k\pi t \quad c_k = \sqrt{2}\int_0^1 c(s)\sin k\pi s ds.$$

This yields

$$\mathcal{L}(\int_0^1 (\mathsf{B}(t) + c(t))^2 dt) = \mathcal{L}(\sum\nolimits_{k=1}^{\infty} \frac{1}{(k\pi)^2}(Z_k + d_k)^2)$$

where $d_k = k\pi c_k$. From here we get a similar statement for the process $\mathsf{B}(F_\theta)$. Indeed, if F_θ is continuous then the mapping $\varphi \mapsto \varphi(F_\theta)$ is a isometry between $L_2[0,1]$ and $L_2(P_\theta)$ which gives

$$\mathcal{L}(\int (\mathsf{B}(F_\theta) + b)^2 dP_\theta) = \mathcal{L}(\sum\nolimits_{k=1}^{\infty} \frac{1}{(k\pi)^2}(Z_k + d_k)^2),$$
$$Z_k = k\pi \int \mathsf{B}(F_\theta)\varphi_k(F_\theta)dP_\theta \quad \text{and} \quad d_k = k\pi \int \varphi_k(F_\theta) b dP_\theta.$$

The next result follows from Theorem 5.

Theorem 7 *Suppose the assumptions in Theorem 5 are fulfilled. Then the asymptotic power of the randomized Cramer-von Mises test* φ_{C_n} *in (25) under the local alternative* $Q_n^{\otimes n}$ *is given by*

$$\begin{aligned}\lim_{n\to\infty} Q_n^{\otimes n}(\varphi_{\mathsf{C}_n} &= 1) = P(\sum\nolimits_{k=1}^{n} \frac{1}{(k\pi)^2}(Z_k + b_k)^2 > c_{1-\alpha}) \\ b_k &= k\pi\sqrt{2}\int \sin(k\pi F_\theta(t))(\Lambda_\theta a)(t)P_\theta(dt).\end{aligned}$$

3 Bootstrap of the Randomized Process

Bootstrap approximations to the estimated empirical process have been considered by several authors, see e.g. Stute et al. (1993), Genz and Haeusler (2006), van der Vaart and Wellner (1996) and the references therein. These authors mainly used the bootstrap approximation for cases where the quantiles of the asymptotic distribution of the test statistic are not available. Although in our case the limit distribution is known we use the bootstrap to check whether the bootstrap approximation improves the asymptotic.

We now suppose that the condition (3) for the MLE holds not only for any fixed θ but for every convergent sequence $\theta_n \to \theta$, i.e. we suppose that the MLE $\widehat{\theta}_n$ satisfies

$$\begin{aligned}\sqrt{n}(\widehat{\theta}_n(X_{n,1},...,X_{n,n}) - \theta_n) &= \frac{1}{\sqrt{n}}\sum\nolimits_{i=1}^{n} I^{-1}(\theta_n)\dot{l}_{\theta_n}(X_{n,i}) + o_P(1) \quad (35) \\ \mathcal{L}(X_{n,i}) &= F_{\theta_n}, \quad \theta_n \to \theta.\end{aligned}$$

Furthermore we assume that $F_\theta(t)$ and $\dot{F}_\theta(t)$ are continuous functions of $(\theta, t) \in \Theta \times [-\infty, \infty]$ and

$$\sup_{t\in[-\infty,\infty],\theta\in U(\theta_0)} |F_{\theta+h}(t) - F_\theta(t) - \dot{F}_\theta^T(t)h| = o(\|h\|) \quad \text{as } h \to 0. \tag{36}$$

The following lemma is contained in the one or other form in all papers dealing with bootstrap of empirical processes with estimated parameters.

Lemma 8 *Under the assumptions (35) and (36) it holds*

$$\sup_t \left|\widehat{\mathsf{G}}_n(t) - (\sqrt{n}(\widehat{F}_n(t) - F_{\theta_n}(t)) - \dot{F}_{\theta_n}^T(t)\sqrt{n}(\widehat{\theta}_n - \theta_n)\right| = o_P(1) \quad \text{as } n \to \infty. \tag{37}$$

where $\widehat{\mathsf{G}}_n = \sqrt{n}(\widehat{F}_n - F_{\widehat{\theta}_n}(t))$ *and*

$$\widehat{F}_n(t) = \frac{1}{n}\sum\nolimits_{i=1}^{n} I_{(-\infty,t]}(X_{n,i})$$

Proof. It holds

$$\widehat{\mathsf{G}}_n(t) = \sqrt{n}(\widehat{F}_n(t) - F_{\theta_n}(t)) - \sqrt{n}(F_{\widehat{\theta}_n}(t) - F_{\theta_n}(t)).$$

For U_n in (1) the processes $\mathsf{U}_n(F_{\theta_n})$ and $\sqrt{n}(\widehat{F}_n(t) - F_{\theta_n}(t))$ have the same distributions. As F_{θ_n} converges uniformly to F_θ the tightness of $\mathsf{U}_n(F_{\theta_n})$ follows from the tightness of U_n. The tightness of $\dot{F}^T_{\theta_n}(t)\sqrt{n}(\widehat{\theta}_n - \theta_n)$ follows from the continuity of $(\theta, t) \mapsto \dot{F}_\theta(t)$ and (35). The condition (36) together with $\sqrt{n}(\widehat{\theta}_n - \theta_n) = O_P(1)$ yield that $\sqrt{n}(F_{\widehat{\theta}_n} - F_{\theta_n})$ is also tight. Hence it remains to prove that for every fixed t

$$\sqrt{n}(F_{\widehat{\theta}_n}(t) - F_{\theta_n}(t)) - \dot{F}^T_{\theta_n}(t)\sqrt{n}(\widehat{\theta}_n - \theta_n) = o_P(1).$$

But this follows from (36). ∎

We denote by $X_{n,1}, ..., X_{n,n}$ i.i.d. random variables with distribution P_{θ_n} and by $X_1, ..., X_n$ i.i.d. random variables with distribution P_θ. A special construction for the $X_{n,i}$ and X_i is given by

$$X_{n,i} = F^{-1}_{\theta_n}(U_i) \quad \text{and} \quad X_i = F^{-1}_\theta(U_i), i = 1, ..., n. \tag{38}$$

Lemma 9 *Suppose that (35), (36) and the conditions in Theorem 3 are fulfilled and it holds*

$$E\left\|\dot{l}_{\theta_n}(X_{n,1}) - \dot{l}_\theta(X_1)\right\|^2 \to 0 \tag{39}$$

for the special construction (38) as $\theta_n \to \theta$. If the standard normal V_n is independent of $X_{n,1}, ..., X_{n,n}$ then for every sequence $\theta_n \to \theta$ it holds

$$\mathcal{L}(\widehat{\mathsf{G}}_n + \dot{F}^T_{\widehat{\theta}_n} I^{-1/2}(\widehat{\theta}_n)V_n | P^{\otimes n}_{\theta_n}) \Longrightarrow \mathcal{L}(\mathsf{B}(F_\theta))$$

Proof. Let V be a standard normal random vector. The continuity of $I^{-1/2}(\theta)$ and $\dot{F}_\theta$ implies

$$\mathcal{L}(\dot{F}^T_{\widehat{\theta}_n} I^{-1/2}(\widehat{\theta}_n)V_n) \Longrightarrow \mathcal{L}(\dot{F}^T_\theta I^{-1/2}(\theta)V).$$

As $\widehat{\mathsf{G}}_n$ and V_n are independent and it suffices to deal with $\widehat{\mathsf{G}}_n(t)$ or in view of Lemma 8 with the processes

$$\begin{aligned} Z_n(t) &= Z_{n,1}(t) + Z_{n,2}(t) \\ &= \sqrt{n}(\widehat{F}_n(t) - F_{\theta_n}(t)) - \dot{F}^T_{\theta_n}(t)\sqrt{n}(\widehat{\theta}_n - \theta_n). \end{aligned}$$

We have already proved the tightness of the processes $Z_{n,1}, Z_{n,2}$ in the proof of Lemma 8. It remains to establish the fidi-convergence. Note that (35) and (39) imply for the special construction of $X_{n,i}$ and X_i in (38)

$$\sqrt{n}(\widehat{\theta}_n(X_{n,1},...,X_{n,n})-\theta_n)-\sqrt{n}(\widehat{\theta}_n(X_1,...,X_n)-\theta)=o_P(1).$$

Furthermore

$$\begin{aligned}&E\left(\frac{1}{\sqrt{n}}\sum\nolimits_{i=1}^n(I_{(-\infty,t]}(X_{n,i})-F_{\theta_n}(t))-\frac{1}{\sqrt{n}}\sum\nolimits_{i=1}^n(I_{(-\infty,t]}(X_i)-F_\theta(t))\right)^2\\ &\le E((I_{(-\infty,t]}(X_{n,1})-F_{\theta_n}(t))-(I_{(-\infty,t]}(X_1)-F_\theta(t)))^2\\ &\le 2E((I_{(-\infty,t]}(F_{\theta_n}^{-1}(U_1))-I_{(-\infty,t]}(F_\theta^{-1}(U_1)))^2+2(F_{\theta_n}(t)-F_\theta(t))^2\\ &=2E((I_{(-\infty,F_{\theta_n}(t)]}(U_1)-I_{(-\infty,F_\theta(t)]}(U_1))^2+2(F_{\theta_n}(t)-F_\theta(t))^2\to 0\end{aligned}$$

by the continuity of $\theta\mapsto F_\theta(t)$. By Slutsky's lemma it remains to investigate the finite dimensional distributions of

$$\frac{1}{\sqrt{n}}\sum\nolimits_{i=1}^n(I_{(-\infty,t]}(X_i)-F_\theta(t))-\dot{F}_\theta^T(t)\sqrt{n}(\widehat{\theta}_n(X_1,...,X_n)-\theta)$$

which has been done in the proof of Theorem 1. ■

Suppose that $X_1,...,X_n$ are i.i.d. with common distribution P_θ. For every n and every realization $x_1,...,x_n$ we denote by $X_{n,1}^*,...,X_{n,n}^*$ i.i.d. random variables with common distribution $P^{\otimes n}_{\widehat{\theta}_n(x_1,...,x_n)}$. We call

$$\widehat{\theta}_n^*=\widehat{\theta}_n(X_{n,1},...,X_{n,n})$$

the bootstrap estimator and introduce the bootstrapped empirical process with estimated parameters by

$$\widehat{\mathsf{G}}_n^*(t)=\frac{1}{\sqrt{n}}\sum\nolimits_{i=1}^n(I_{(-\infty,t]}(X_{n,i}^*)-F_{\widehat{\theta}_n^*}(t)).$$

Suppose that V_n is standard normal and independent of $X_{n,1}^*,...,X_{n,n}^*,X_1,...,X_n$.

Theorem 10 *If the conditions of Lemma 9 are fulfilled and the MLE $\widehat{\theta}_n$ is strongly consistent then*

$$\mathcal{L}(\widehat{\mathsf{G}}_n^*+\dot{F}^T_{\widehat{\theta}_n^*}I^{-1/2}(\widehat{\theta}_n^*)V_n|P^{\otimes n}_{\widehat{\theta}_n})\Longrightarrow\mathcal{L}(\mathsf{B}(F_\theta))\quad a.s.\ as\ n\to\infty.$$

Proof. Set $\mathbb{N}=\{1,2,...\}$, fix a set $A\in\mathfrak{B}^{\otimes\mathbb{N}}$ with $P_\theta^{\otimes\mathbb{N}}$-probability one so that $\widehat{\theta}_n$ converges pointwise to θ on A. The application of Lemma 9 for every sequence $(x_1,x_2,...)\in A$ yields the statement. ■

Let $T_n(X_1, ..., X_n)$ be a sequence of statistics and assume that

$$\mathcal{L}(T_n(X_1, ..., X_n)|P_\theta^{\otimes n}) \Rightarrow Q_\theta. \tag{40}$$

Suppose that the distribution function $G_\theta(t) = Q_\theta((-\infty, t])$ is continuous in (θ, t). To construct a bootstrap approximation to the quantiles of the distribution of the test statistic we generate conditionally i.i.d. $X_{n,1}^*, ..., X_{n,n}^*$ with common distribution $P_{\widehat{\theta}_n}$. Assume that the bootstrap is consistent, i.e.

$$\int \varphi(T_n(X_{n,1}^*, ..., X_{n,n}^*))dP_{\widehat{\theta}_n}^{\otimes n} \to^{P_\theta^{\otimes n}} \int \varphi Q_\theta, \quad n \to \infty \tag{41}$$

for every bounded and continuous function φ and suppose that

$$X_{n,1}^*, ..., X_{n,n}^*, X_{n,n+1}^*, ..., X_{n,2n}^*, ..., X_{n,(B-1)n+1}^*, ..., X_{n,B\cdot n}^*$$

are conditionally i.i.d. with common distribution $P_{\widehat{\theta}_n}$. Applying T_n to the B independent blocks each with length n we get

$$T_{n,i}^* = T_n(X_{n,(i-1)n+1}^*, ..., X_{n,in}^*), \quad i = 1, ..., B.$$

Put

$$G_{n,B}^*(t) := \frac{1}{B} \sum\nolimits_{i=1}^B I_{(-\infty,t]}(T_{n,i}^*).$$

To study the convergence of $G_{n,B}^*(t)$ we set

$$G_n(t|\widehat{\theta}_n) = P_{\widehat{\theta}_n}^{\otimes n}(T_{n,1}^* \leq t).$$

Then

$$E(G_{n,B}^*(t) - G_\theta(t))^2 \leq 2(E(G_{n,B}^*(t) - G_n(t|\widehat{\theta}_n))^2 + E(G_n(t|\widehat{\theta}_n) - G_\theta(t))^2).$$

The $T_{n,i}^*, i = 1, ..., B$ are i.i.d. conditionally on $\widehat{\theta}_n$. This gives

$$E(G_{n,B}^*(t) - G_n(t|\widehat{\theta}_n))^2|\widehat{\theta}_n) = \frac{1}{B}E((I_{(-\infty,t]}(T_{n,1}^*) - G_n(t|\widehat{\theta}_n))^2|\widehat{\theta}_n) \leq \frac{1}{B},$$

so that the left hand term tends stochastically to zero as $B \to \infty$. The second term tends to zero as $n \to \infty$ in view of (41) and the continuity of $G_\theta(t)$. Hence $G_{n,B}^*(t)$ tends stochastically to $G_\theta(t)$ for every fixed t. The continuity of $G_\theta(t)$ implies for $n, B \to \infty$

$$\sup_t \left|G_{n,B}^*(t) - G_\theta(t)\right| \to^P 0.$$

If G_θ is strictly increasing we may conclude

$$(G_{n,B}^*)^{-1}(\alpha) \to^P G_\theta^{-1}(\alpha) \tag{42}$$

for every $0 < \alpha < 1$. We use the statistics T_n to construct a sequence of tests.

$$\varphi_{T_n} = I_{[G_\theta^{-1}(1-\alpha),\infty)}(T_n).$$

The φ_{T_n} are asymptotic level α-tests in the sense that $\lim_{n\to\infty} E\varphi_{T_n} = \alpha$. The bootstrap version of the test φ_{T_n} is then given by

$$\varphi_{T_n}^* = I_{[(G_{n,B}^*)^{-1}(1-\alpha),\infty)}(T_n).$$

It follows from (42) that under conditions (40) and (41) the bootstrap test $\varphi_{T_n}^*$ is consistent in the sense that

$$\lim_{n,B\to\infty} E\varphi_{T_n}^* = \lim_{n\to\infty} E\varphi_{T_n} = \alpha.$$

Let us now specialize the statistics T_n. We use the randomized Kolmogorov-Smirnov statistic K_n and the randomized Cramer von Mises statistic C_n

$$\begin{aligned}
\mathsf{K}_n &= \sup_t |\widehat{\mathsf{G}}_n(t) + \dot{F}_{\widehat{\theta}_n}^T(t) I^{-1/2}(\widehat{\theta}_n) V_n| \\
\mathsf{C}_n &= \sum\nolimits_{i=1}^n ((\frac{i}{n} - F_{\widehat{\theta}_n}(X_{n:i}) + \frac{1}{\sqrt{n}} \dot{F}_{\widehat{\theta}_n}^T(X_{n:i}) I^{-1/2}(\widehat{\theta}_n) V_n)^2.
\end{aligned}$$

The conditions (40) and (41) for these statistics follow from Theorem 1 and Theorem 10 under the conditions formulated there. Thus we have obtained the following result.

Theorem 11 *Under the assumptions of the Theorems 1 and 10 the bootstrap tests φ_{K_n} and φ_{C_n} in (25) based on the randomized Kolmogorov-Smirnov statistic K_n and the randomized Cramer von Mises statistic C_n are consistent.*

4 Examples

4.1 Normal distribution

4.1.1 Normal distribution with unknown μ

Let $\mathsf{N}(\mu, \sigma^2)$ be the normal distribution with expectation μ and variance σ^2. Denote by Φ and $\varphi(x)$ be the distribution function and density function, respectively, of $\mathsf{N}(0,1)$. As σ^2 is known we may assume $\sigma^2 = 1$ without loss of generality. The Gaussian location model is then given by $P_\theta = \mathsf{N}(\mu, 1)$, $\theta = \mu \in \mathbb{R}$ and

$$F_\mu(t) = \Phi(t-\mu) \quad \text{and} \quad \dot{F}_\mu(t) = -\varphi(t-\mu).$$

The condition (7) is obviously satisfied. The MLE of μ is $\overline{X}_n$ and the conditions (3), (4) and (5) are clear with $h_\theta(x) = x$ and $J(\mu) = I(\mu) = 1$. Furthermore, the regularity conditions (11) are fulfilled. Thus we get that the empirical process with estimated parameters

$$\widehat{\mathsf{G}}_n(t) = \sqrt{n}(\widehat{F}_n(t) - \Phi(t - \overline{X}_n))$$

satisfies $\mathcal{L}(\widehat{\mathsf{G}}_n) \Rightarrow \mathcal{L}(Z_\mu)$ where Z_μ is a centered Gaussian process where the covariance function is according to (13) given by

$$cov(Z_\mu((s), Z_\mu(t)) = \Phi((s \wedge t) - \mu) - \Phi(s-\mu)\Phi(t-\mu) - \varphi(s-\mu)\varphi(t-\mu).$$

We set

$$\mathsf{M}_n = n \int \left(\hat{F}_n(t) - \Phi\left(t - \overline{X}_n\right)\right)^2 \mathsf{N}(\mu, 1)(dt)$$

and consider two versions of the Cramer-von Mises statistic

$$\begin{aligned} \mathsf{M}_{n,1} &= n \int \left(\hat{F}_n(t) - \Phi\left(t - \overline{X}_n\right)\right)^2 \mathsf{N}(\overline{X}_n, 1)(dt), \\ \mathsf{M}_{n,2} &= n \int \left(\hat{F}_n(t) - \Phi\left(t - \overline{X}_n\right)\right)^2 d\hat{F}_n(t). \end{aligned} \tag{43}$$

By similar considerations that have led us to (22) one can see that

$$\mathsf{M}_{n,1} = \mathsf{M}_{n,2} + o_P(1) = \mathsf{M}_n + o_P(1).$$

Finally, by the continuous mapping theorem

$$\mathcal{L}(\mathsf{M}_n) \Rightarrow \mathcal{L}(\int Z_\mu^2(t)\mathsf{N}(\mu, 1)(dt)) = \mathcal{L}(\int Z_0^2(t)\mathsf{N}(0, 1)(dt))$$

and therefore

$$\mathcal{L}(\mathsf{M}_{n,i}) \Rightarrow \mathcal{L}(\int Z_0^2(t)\mathsf{N}(0, 1)(dt)), i = 1, 2.$$

Thus we get the well known result, that for a location model the asymptotic distribution of the Cramer- von Mises statistic of the empirical process with estimated parameters does not depend on the special value of θ. But the distribution does depend on the parent distribution that generates the model. A similar statement holds for the Kolmogorov-Smirnov statistic

$$\mathsf{S}_n = \sqrt{n} \sup_t |\hat{F}_n(t) - \Phi\left(t - \overline{X}_n\right)|. \tag{44}$$

It holds

$$\mathcal{L}(\mathsf{S}_n) \Rightarrow \mathcal{L}(\sup_t |Z_\mu(t)|) = \mathcal{L}(\sup_t |Z_0(t)|).$$

The process $\mathsf{R}_n(\widetilde{\theta}_n)$ in (15) is given by

$$\mathsf{R}_n(\widetilde{\theta}_n) = \sqrt{n}(\widehat{F}_n(t) - \Phi\left(t - \overline{X}_n\right)) - \varphi\left(s - \overline{X}_n\right) V_n,$$

where we used $\widetilde{\theta}_n = \overline{X}_n$ and V_n is standard normal and independent of $X_1, ..., X_n$. Then we get from Theorem 3 that

$$\mathcal{L}(\sqrt{n}(\widehat{F}_n - \Phi\left(\cdot - \overline{X}_n\right)) - \varphi\left(\cdot - \overline{X}_n\right) V_n) \Rightarrow \mathcal{L}(\mathsf{B}(\Phi\left(\cdot - \mu\right))). \tag{45}$$

In the special case under consideration the randomized Cramer von Mises Statistic in (19) and the randomized Kolmogorov-Smirnov statistic in (17), are given by

$$\mathsf{C}_n = \sum\nolimits_{i=1}^n (\frac{i}{n} - \Phi\left(X_{n:i} - \overline{X}_n\right) - \frac{1}{\sqrt{n}}\varphi\left(X_{n:i} - \overline{X}_n\right) V_n)^2 \tag{46}$$

$$\mathsf{K}_n = \sup_t \left|\sqrt{n}(\widehat{F}_n(t) - \Phi\left(t - \overline{X}_n\right)) - \varphi\left(t - \overline{X}_n\right) V_n\right|. \tag{47}$$

Using the notations in (16) and (24) we get from (45)

$$\mathcal{L}(\mathsf{C}_n) \Rightarrow \mathcal{C} \quad \text{and} \quad \mathcal{L}(\mathsf{K}_n) \Rightarrow \mathcal{K}.$$

4.1.2 Normal distribution with unknown μ and σ^2

Now we allow the variance to be unknown and arrive at a location-scale model generated by the standard normal distribution. Hence $P_\theta = \mathsf{N}(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2), \mu \in \mathbb{R}$, $\sigma^2 > 0$. Then $F_\theta(t) = \Phi\left(\frac{t-\mu}{\sigma}\right)$ and

$$\begin{aligned}\dot{F}_\theta(t) &= \left(\frac{\partial}{\partial\mu}\Phi\left(\frac{t-\mu}{\sigma}\right),)\frac{\partial}{\partial\sigma^2}\Phi\left(\frac{t-\mu}{\sigma}\right)\right)^T \\ &= \left(-\frac{1}{\sigma}\varphi\left(\frac{t-\mu}{\sigma}\right), -\frac{(t-\mu)}{2\sigma^3}\varphi\left(\frac{t-\mu}{\sigma}\right)\right)^T\end{aligned}$$

and the condition (7) is again satisfied. The MLE of $\theta = (\mu, \sigma^2)$ is

$$\widehat{\theta}_n = (\overline{X}_n, S_n^2)^T, \quad \text{and} \quad S_n^2 = \frac{1}{n}\sum\nolimits_{i=1}^{n}(X_i - \overline{X}_n)^2.$$

As $\sqrt{n}(\overline{X}_n - \mu)^2 = o_P(1)$ it follows

$$\sqrt{n}\begin{pmatrix} \overline{X}_n - \mu \\ S_n^2 - \sigma^2 \end{pmatrix} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} X_i - \mu \\ (X_i - \mu)^2 - \sigma^2 \end{pmatrix} + o_P(1).$$

and

$$\mathcal{L}\left(\sqrt{n}\begin{pmatrix} \overline{X}_n - \mu \\ S_n^2 - \sigma^2 \end{pmatrix}\right) \Rightarrow \mathsf{N}(0, \Sigma)$$

where

$$\sum = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{pmatrix} = I^{-1}(\theta).$$

Hence the conditions (3), (4) and (5) are satisfied. Furthermore, the regularity conditions (11) are fulfilled. Thus the empirical process with estimated parameters

$$\widehat{\mathsf{G}}_n(t) = \sqrt{n}\left(\widehat{F}_n(t) - \Phi\left(\frac{t - \overline{X}_n}{S_n}\right)\right)$$

satisfies $\mathcal{L}(\widehat{\mathsf{G}}_n) \Rightarrow \mathcal{L}(Z_{(\mu,\sigma^2)})$ where $Z_{(\mu,\sigma^2)}$ is a centered Gaussian process where the covariance function is, according to (13), given by

$$cov(Z_{(\mu,\sigma^2)}((s), Z_{(\mu,\sigma^2)}(t)) = \Phi\left(\frac{(s \wedge t) - \mu}{\sigma}\right) - \Phi\left(\frac{s - \mu}{\sigma}\right)\Phi\left(\frac{t - \mu}{\sigma}\right)$$
$$-\varphi\left(\frac{s - \mu}{\sigma}\right)\varphi\left(\frac{t - \mu}{\sigma}\right) - \frac{(s - \mu)(t - \mu)}{2\sigma^2}\varphi\left(\frac{s - \mu}{\sigma}\right)\varphi\left(\frac{t - \mu}{\sigma}\right).$$

We set

$$\mathsf{M}_n = n\int\left(\hat{F}_n(t) - \Phi\left(\frac{t - \overline{X}_n}{S_n}\right)\right)^2 \mathsf{N}(\mu, \sigma^2)(dt)$$

and consider two versions of the Cramer-von Mises statistic

$$\begin{aligned} \mathsf{M}_{n,1} &= n\int\left(\hat{F}_n(t) - \Phi\left(\frac{t - \overline{X}_n}{S_n}\right)\right)^2 \mathsf{N}(\overline{X}_n, S_n)(dt), \qquad (48) \\ \mathsf{M}_{n,2} &= n\int\left(\hat{F}_n(t) - \Phi\left(\frac{t - \overline{X}_n}{S_n}\right)\right)^2 d\hat{F}_n(t). \end{aligned}$$

Again one can see that

$$\mathsf{M}_{n,1} = \mathsf{M}_{n,2} + o_P(1) = \mathsf{M}_n + o_P(1).$$

Finally, by the continuous mapping theorem

$$\mathcal{L}(\mathsf{M}_n) \Rightarrow \mathcal{L}(\int Z^2_{(\mu,\sigma^2)}(t)\mathsf{N}(\mu,\sigma^2)(dt)) = \mathcal{L}(\int Z^2_{(0,1)}(t)\mathsf{N}(0,1)(dt)),$$

and therefore

$$\mathcal{L}(\mathsf{M}_{n,i}) \Rightarrow \mathcal{L}(\int Z^2_{(0,1)}(t)\mathsf{N}(0,1)(dt)), \quad i = 1,2. \tag{49}$$

where $Z_{(0,1)}$ is the centered Gaussian process with covariance function

$$cov(Z_{(0,1)}((s), Z_{(0,1)}(t)) = \Phi(s \wedge t) - \Phi(s)\Phi(t) - \varphi(s)\varphi(t) - \frac{1}{2}st\varphi(s)\varphi(t).$$

Denote by

$$\mathsf{S}_n = \sqrt{n}\sup_t \left| \hat{F}_n(t) - \Phi\left(\frac{t - \overline{X}_n}{S_n}\right)\right|. \tag{50}$$

the Kolmogorov-Smirnov statistic. It holds

$$\mathcal{L}(\mathsf{S}_n) \Rightarrow \mathcal{L}(\sup_t |Z_{(\mu,\sigma^2)}(t)|) = \mathcal{L}(\sup_t |Z_{(0,1)}(t)|).$$

The randomized process $\mathsf{R}_n(\widetilde{\theta}_n)$ in (15) is given by

$$\mathsf{R}_n(\widetilde{\theta}_n) = \sqrt{n}\left(\widehat{F}_n(t) - \Phi\left(\frac{t-\overline{X}_n}{S_n}\right)\right) - \frac{1}{\sqrt{n}}\varphi\left(\frac{t-\overline{X}_n}{S_n}\right)V_{1,n}$$
$$- \frac{(t-\overline{X}_n)}{\sqrt{2n}S_n}\varphi\left(\frac{t-\overline{X}_n}{S_n}\right)V_{2,n},$$

where we used $\widetilde{\theta}_n = (\overline{X}_n, S_n^2)$ and $V_{1,n}, V_{2,n}$ are independent, standard normal and independent of $X_1, ..., X_n$. We get from Theorem 3 that

$$\mathcal{L}\left(\widehat{\mathsf{G}}_n + \dot{F}^T_{\widetilde{\theta}_n} I^{-1/2}(\widetilde{\theta}_n)V_n\right) \Rightarrow \mathcal{L}\left(\mathsf{B}(\Phi((\cdot - \mu)/\sigma)\right).$$

In the special case under consideration the randomized Cramer-von Mises statistic in (19) and the randomized Kolmogorov-Smirnov statistic in (17), respectively, are given by

$$\mathsf{C}_n = \sum\nolimits_{i=1}^n \left(\frac{i}{n} - \Phi\left(\left(X_{n:i} - \overline{X}_n\right)/S_n\right)\right. \tag{51}$$
$$\left. - \frac{1}{\sqrt{n}}\varphi\left(\left(X_{n:i} - \overline{X}_n\right)/S_n\right)V_{1,n} - \frac{(X_{n:i} - \overline{X}_n)}{\sqrt{2n}S_n}\varphi\left(\left(X_{n:i} - \overline{X}_n\right)/S_n\right)V_{2,n}\right)^2$$

$$\mathsf{K}_n = \sup_t \left| \sqrt{n}(\widehat{F}_n(t) - \Phi\left(t - \overline{X}_n\right)) - \frac{1}{S_n}\varphi\left(\left(t - \overline{X}_n\right)/S_n\right)V_{1,n}\right. \tag{52}$$
$$\left. - \frac{(X_{n:i} - \overline{X}_n)}{\sqrt{2}S_n}\varphi\left(\left(t - \overline{X}_n\right)/S_n\right)V_{2,n}\right|.$$

Using the notations in (16) and (24) we get from (45)

$$\mathcal{L}(\mathsf{C}_n) \Rightarrow \mathcal{C} \quad \text{and} \quad \mathcal{L}(\mathsf{K}_n) \Rightarrow \mathcal{K}.$$

4.1.3 Exponential distribution with unknown parameter

Finally we apply the general results to the family of exponential distributions with parameter $\lambda > 0$. Let

$$G(t) = I_{[0,\infty)}(t)(1 - \exp\{-t\}) \quad \text{and} \quad g(t) = I_{[0,\infty)}(t)\exp\{-t\},$$

be the distribution function and density function of the standard exponential distribution, respectively. The exponential distribution model is then given by $P_\theta = G(\lambda t)$, $\theta = \lambda \in \mathbb{R}^+$ and we have

$$G_\lambda(t) = I_{[0,\infty)}(t)(1 - \exp\{-\lambda t\}) \quad \text{and} \quad \dot{G}_\lambda(t) = I_{[0,\infty)}(t)(t\exp\{-\lambda t\}).$$

The condition (7) is satisfied. The MLE of λ is $\widehat{\lambda}_n = 1/\overline{X}_n$ and it holds

$$\sqrt{n}(\frac{1}{\overline{X}_n} - \lambda) = -\frac{\lambda}{\overline{X}_n}\frac{1}{\sqrt{n}}\sum\nolimits_{i=1}^{n}(X_i - \frac{1}{\lambda}).$$

As $V(X_i) = \frac{1}{\lambda^2}$ we see that the conditions (3), (4) and (5) are satisfied with $h_\lambda(x) = \lambda^2(1/\lambda - x)$ and $J(\lambda) = I(\lambda) = 1/\lambda^2$. Furthermore, the regularity conditions (11) are fulfilled. Thus we get that the empirical process with estimated parameters

$$\widehat{\mathsf{G}}_n(t) = \sqrt{n}(\widehat{F}_n(t) - G_{1/\overline{X}_n}(t))$$

satisfies $\mathcal{L}(\widehat{\mathsf{G}}_n) \Rightarrow \mathcal{L}(Z_\lambda)$. The process Z_λ is a centered Gaussian process where the covariance function for $0 \le s, t < \infty$ is according to (13) given by

$$\begin{aligned} cov(Z_\lambda(s), Z_\lambda(t)) &= G_\lambda(s \wedge t) - G_\lambda(s)G_\lambda(t) - \lambda^2\dot{G}_\lambda(s)\dot{G}_\lambda(t) \\ &= 1 - \exp\{-\lambda(s \wedge t)\} - (1 - \exp\{-\lambda s\})(1 - \exp\{-\lambda t\}) \\ &\quad -\lambda^2 st\exp\{-\lambda t\}\exp\{-\lambda s\}. \end{aligned}$$

We consider two versions of the Cramer-von Mises statistic

$$\begin{aligned} \mathsf{M}_{n,1} &= n\int_0^\infty \left(\hat{F}_n(t) - G_{1/\overline{X}_n}(t)\right)^2 \frac{1}{\overline{X}_n}\exp\left\{-\frac{t}{\overline{X}_n}\right\}dt \qquad (53) \\ \mathsf{M}_{n,2} &= n\int \left(\hat{F}_n(t) - G_{1/\overline{X}_n}(t)\right)^2 d\hat{F}_n(t) \end{aligned}$$

The same arguments that proved (22) yield

$$\mathcal{L}(\mathsf{M}_{n,i}) \Rightarrow \mathcal{L}(\int_0^\infty Z_1^2(t) \exp\{-t\} dt, \quad i = 1, 2.$$

For the Kolmogorov-Smirnov statistic

$$\mathsf{S}_n = \sqrt{n} \sup_t |\hat{F}_n(t) - G_{1/\overline{X}_n}(t)| \tag{54}$$

a similar statement holds

$$\mathcal{L}(\mathsf{S}_n) \Rightarrow \mathcal{L}(\sup_t |Z_1(t)|) = \mathcal{L}(\sup_t |Z_1(t)|).$$

The randomized Cramer-von Mises statistic in (19) and the randomized Kolmogorov-Smirnov statistic in (17), in the special case under consideration, are given by

$$\begin{aligned} \mathsf{C}_n &= \sum\nolimits_{i=1}^n \Bigg(\frac{i}{n} - (1 - \exp\{-X_{n:i}/\overline{X}_n\}) \\ &\qquad - \frac{V_n}{\sqrt{n}} \frac{X_{n:i}}{\overline{X}_n} \exp\{-X_{n:i}/\overline{X}_n\} \Bigg)^2 \qquad (55) \\ \mathsf{K}_n &= \sup_t \left| \sqrt{n}(\widehat{F}_n(t) - \big(1 - \exp\{-t/\overline{X}_n\}\big) - \frac{tV_n}{\overline{X}_n} \exp\{-t/\overline{X}_n\} \right|. \qquad (56) \end{aligned}$$

We get from (45)

$$\mathcal{L}(\mathsf{C}_n) \Rightarrow \mathcal{C} \quad \text{and} \quad \mathcal{L}(\mathsf{K}_n) \Rightarrow \mathcal{K}.$$

5 Simulations

5.1 Randomized tests

Monte Carlo sampling experiments to check the grade of accuracy of the approximation by the limit distribution of the Cramer- von Mises statistics and the Kolmogorov-Smirnov statistic, respectively, have been carried out by several authors, see e.g. Stephens (1974) and (1976). In this section we check the actual significance level of tests that are based on the randomized statistics in (19) and (17).

The simulation experiment is performed according to the following steps. Let F_θ be a normal or an exponential distribution and let T_n stand for one of the following statistics, where the number in the display refers to the corresponding formula

Randomized Cramer-von Mises statistics	(46), (51), (55)
Randomized Kolmogorov-Smirnov statistic	(47), (52), (56).

(57)

To carry out the simulations we used the programm R. The implemented pseudo random generator is the Mersenne-Twister generator, see Matsumoto and Nishimura (1998).

1. For $n = 20; 50; 100$ we generate $X_1, ..., X_n$ from F_θ.
2. Calculate the MLE $\hat{\theta}_n$.
3. Calculate the values of the statistics T_n from the display (57).
4. Carry out the test

$$\varphi_n = \begin{cases} 1 & , \quad T_n > c_{1-\alpha} \\ 0 & , \quad \text{else} \end{cases} ,$$

   where $c_{1-\alpha}$ is the $1-\alpha$ quantile of T_n. As an approximation of $c_{1-\alpha}$ for different n we used the modified statistics in Table 1, page 239 in Shorack and Wellner (1986). The corresponding values are listed in the tables below in the rows that are named "CM exact" and "KS exact".

5. Repeat the steps 1.-4. N times and estimate the actual confidence level by

$$\hat{\alpha}_{T_n} = \frac{\text{number of rejections of } H_0}{N}.$$

We used $N = 10000$ in our simulations.

5.2 Bootstrap approximation

We have used computer simulations to study the grade of accuracy of the actual distribution of the test statistics by the asymptotic distribution in the previous section. Now we apply bootstrap approximations and check whether there is a significant improvement of the accuracy. The bootstrap simulations are carried out according to the following scheme that corresponds to the approach in Stute et al. (1993).

1. For $n = 20; 50; 100$ we generate pseudo random numbers from F_θ, where again F_θ stands either for the normal or for the exponential distribution.
2. Calculate the MLE $\hat{\theta}_n$ of θ
3. Calculate the values of the statistics T_n from the display (57).

4. Generate $B \cdot n$ i.i.d. random variables

$$\begin{array}{ccc} X^*_{1,1} & \cdots & X^*_{1,n} \\ \vdots & \ddots & \vdots \\ X^*_{B,1} & \cdots & X^*_{B,n} \end{array}$$

   with common distribution function $F_{\hat{\theta}_n}$.

5. Calculate the bootstrap MLE $\hat{\theta}^*_{1,n}, ..., \hat{\theta}^*_{B,n}$

6. Calculate the bootstrap version $T^*_{1,n}, ..., T^*_{B,n}$ of the statistics listed in display (57)

7. Calculate the $1-\alpha$ quantile $c^*_{1-\alpha}$ of the bootstrap empirical distribution function $F^*_B(t) = \frac{1}{B}\sum_{i=1}^B I_{(-\infty,t)}(T^*_{i,n})$

8. Carry out the bootstrap test

$$\varphi^*_n = \begin{cases} 1 & \text{if} \quad T_n > c^*_{1-\alpha} \\ 0 & \text{else} \end{cases}$$

9. Repeat the steps 1.-8. $N = 1000$ times and estimate the actual first kind error probability by

$$\hat{\alpha} = \frac{\text{number of rejections of } H_0}{N}.$$

10. Calculate the arithmetic mean $\overline{c}^*_{1-\alpha}$of the bootstrap quantiles obtained from the $N = 1000$ replications of the steps 1.-8.

5.3 Results of the computer simulations

Subsequently α(CMR) denotes the actual level of the randomized Cramer-von Mises test. α(BCMR) is the actual level of the bootstrap version of the randomized Cramer-von Mises test. $\mathrm{c}_{n,1-\alpha}$ is the exact critical value for the Cramer- von Mises test. We calculated these values with the help of the modified test statistic in Table 6 p. 149 in Shorack and Wellner (1986). c(CMR) is the actual critical values obtained from computer simulations with 10000 replications for the Cramer-von Mises test. Similarly $c(\text{BCMR}) = \overline{c}^*_{1-\alpha}$ is the arithmetic mean of the bootstrap quantiles from $N = 1000$ replications for the Cramer-von Mises test.

Normal distribution with known σ^2 and unknown μ						
	α	α(CMR)	α(BCMR)	$c_{n,1-\alpha}$	c(CMR)	c(BCMR)
n=20	0.100	0.112	0.117	0.349	0.367	0.377
	0.050	0.058	0.063	0.458	0.492	0.509
	0.010	0.013	0.014	0.726	0.806	0.797
n=50	0.100	0.104	0.106	0.348	0.353	0.357
	0.050	0.051	0.058	0.460	0.465	0.483
	0.010	0.010	0.012	0.736	0.740	0.779
n=100	0.100	0.110	0.103	0.348	0.361	0.353
	0.050	0.054	0.052	0.460	0.473	0.465
	0.010	0.010	0.013	0.740	0.744	0.793
n=1000	0.100	0.096	0.108	0.347	0.351	0.348
	0.050	0.052	0.050	0.461	0.457	0.458
	0.010	0.011	0.011	0.743	0.736	0.768

The next table shows the numerical results if both parameters are unknown.

Normal distribution with unknown μ und σ^2						
	α	α(CMR)	α(BCMR)	c$_{n,1-\alpha}$	c(CMR)	c(BCMR)
n=20	0.100	0.114	0.109	0.349	0.372	0.371
	0.050	0.060	0.060	0.458	0.502	0.495
	0.010	0.013	0.015	0.726	0.783	0.790
n=50	0.100	0.104	0.104	0.348	0.354	0.356
	0.050	0.052	0.049	0.460	0.471	0.473
	0.010	0.010	0.012	0.736	0.755	0.751
n=100	0.100	0.107	0.096	0.348	0.360	0.350
	0.050	0.053	0.045	0.460	0.469	0.467
	0.010	0.010	0.011	0.740	0.751	0.471
n=1000	0.100	0.099	0.093	0.347	0.346	0.347
	0.050	0.047	0.052	0.461	0.453	0.458
	0.010	0.010	0.009	0.743	0.758	0.720

The subsequent table shows the simulation results for the family of exponential distributions with unknown parameter λ.

Exponential distribution with unknown λ						
	α	α(CMR)	α(BCMR)	$c_{n,1-\alpha}$	c(CMR)	c(BCMR)
n=20	0.100	0.108	0.107	0.349	0.359	0.355
	0.050	0.055	0.053	0.458	0.480	0.469
	0.010	0.012	0.010	0.726	0.777	0.741
n=50	0.100	0.104	0.110	0.348	0.353	0.351
	0.050	0.051	0.058	0.460	0.467	0.463
	0.010	0.012	0.011	0.736	0.791	0.734
n=100	0.100	0.105	0.109	0.348	0.355	0.348
	0.050	0.053	0.056	0.460	0.471	0.461
	0.010	0.012	0.013	0.740	0.726	0.730
n=1000	0.100	0.099	0.098	0.347	0.346	0.348
	0.050	0.049	0.048	0.461	0.459	0.458
	0.010	0.009	0.017	0.743	0.731	0.728

Parallel to the inspection of the actual level we compared the critical values of the limit distribution listed in the column $c_{n,1-\alpha}$ with the actual $0.90; 0.95$ and 0.99 quantiles listed in the last two columns.

Now we turn to the Kolmogorov-Smirnov test and use similar abbreviations.

Normal distribution with unknown μ						
	α	α(KS)	α(BKS)	$k_{n,1-\alpha}$	k(KS)	k(BKS)
n=20	0.100	0.079	0.081	1.186	1.179	1.187
	0.050	0.039	0.041	1.135	1.134	1.318
	0.010	0.006	0.012	1.577	1.158	1.572
n=50	0.100	0.089	0.105	1.201	1.199	1.200
	0.050	0.044	0.059	1.332	1.341	1.333
	0.010	0.008	0.007	1.597	1.594	1.591
n=100	0.100	0.090	0.106	1.208	1.203	1.207
	0.050	0.044	0.056	1.240	1.337	1.339
	0.010	0.008	0.012	1.607	1.592	1.596
n=1000	0.100	0,094	0.106	1.219	1.211	1.217
	0.050	0.049	0.055	1.353	1.354	1.349
	0.010	0.010	0.016	1.622	1.622	1.604

Normal distribution with unknown μ und σ^2						
	α	α(KS)	α(BKS)	$\mathrm{k}_{n,1-\alpha}$	k(KS)	k(BKS)
	0.100	0.079	0,092	1.186	1,174	1,181
n=20	0.050	0,037	0,049	1.135	1,306	1,312
	0.010	0,007	0,010	1.577	0,574	1,565
	0.100	0,087	0,085	1.201	1,192	1,196
n=50	0.050	0,043	0,047	1.332	1,330	1,328
	0.010	0,009	0,012	1.597	1,608	1,583
	0.100	0,089	0,104	1.208	1,201	1,205
n=100	0.050	0,045	0,055	1.240	1,338	1,337
	0.010	0,009	0,011	1.607	1,606	1,595
	0.100	0,095	0,105	1.219	1,213	1,215
n=1000	0.050	0,047	0,049	1.353	1,346	1,348
	0.010	0,009	0,014	1.622	1,615	1,605

Exponential distribution with unknown λ						
	α	α(KS)	α(BKS)	$k_{n,1-\alpha}$	k(KS)	k(BKS)
	0.100	0.081	0.105	1.186	1.180	1.184
n=20	0.050	0.039	0.059	1.135	1.313	1.315
	0.010	0.007	0.016	1.577	1.567	1.571
	0.100	0.086	0.087	1.201	1.195	1.199
n=50	0.050	0.042	0.045	1.332	1.327	1.329
	0.010	0.008	0.006	1.597	1.587	1.583
	0.100	0.091	0.103	1.208	1.205	1.204
n=100	0.050	0.045	0.050	1.240	1.341	1.337
	0.010	0.008	0.018	1.607	1.608	1.596
	0.100	0.096	0.086	1.219	1.215	1.216
n=1000	0.050	0.049	0.046	1.353	1.353	1.349
	0.010	0.009	0.014	1.622	1.619	1.604

Again we compared the critical values of the limit distribution in column $k_{n,1-\alpha}$ with the actual $0.90; 0.95$ and 0.99 quantiles listed in the last two columns.

From the above tables we may conclude that the randomized goodness of fit test statistics even for small sample sizes behave very similar as the corresponding goodness of fit test statistics for simple null hypothesis. This is demonstrated by the fact that the actual level of the tests for $n = 20; 50; 100; 100$ are very close to the predetermined α if the critical values are chosen as it has to be done for simple null hypothesis. These critical values are the number $c_{n,1-\alpha}$ and $k_{n,1-\alpha}$, respectively, in the above tables. As to the bootstrap versions one can say that in most cases they correct the test in the right direction and takes the actual level of the test closer to the required one. Sometimes this leads to an overcorrection. Only in few cases the bootstrap modifies the test in the false direction. But altogether we have the result that the bootstrap does not provide substantial improvements. Presumably, bootstrap approximations of tests lead to considerable improvements of the actual level only when the actual level of the test to be bootstrapped differs considerably from the required one.

References

Bickel, P.J., Klaassen, C.A.J., Ritov, Y. and Wellner, J.A. (1993) Efficient and Adaptive Estimation for Semiparametric Models. Johns Hopkins Univ. Press, Baltimore.

Billingsley, P.(1986) Covergence of Probability Measures. Wiley, New York.

Durbin, J. (1973a) Distribution theory for tests based on the sample distribution function. Regional conference Series in Applied Mathematics, Vol. 9, SIAM, Philadelphia, Pennsylvania.

Durbin, J. (1973b) Weak convergence of the sample distribution function when parameters are estimated. Ann. Statist., 1, 279-290.

Genz, M. and Haeusler, E. (2006) Empirical processes with estimated parameters under auxiliary information. Journal of Computional and Applied Mathematics, Vol. 186, 191-216.

Matsumoto, M. and Nishimura, T. (1998) Mersenne-Twister: A 623-dimensionally equidistributed uniform pseudo-random generator. ACM Transactions on Modeling and Computer Simulations, 8 (19), 3-30.

Pettit, A. (1977) Tests for the exponential distribution with censored data using Cramer - von - Mises statistics. Biometrika, 64, 629-632

Pollard,D.(1984) Convergence of Stochastic Processes. Springer - Verlag, New York.

Shorack, G. and Wellner, J.A. (1986) Empirical Processes with Applications to Statistics.

Stute, W., Manteiga, W.G. and Quindimil, M.P.(1993) Bootstrap Based Goodness-Of-Fit-Tests. Metrika(1993) 40: 243-256.

van der Vaart, A.W. and Wellner, J.A. (1996) Weak convergence and Empirical Processes. With Applications to Statstics. 2nd pr. 2000. Springer, New York.

van der Vaart, A.W. (1998) Asymptotic Statistics, Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge Univ. Press.

Stephens, M.A. (1974) EDF statistics for goodness of fit and some comparisons. J. Am. Statist. Assoc., 69, 730-737.

Stephens, M.A. (1976) Goodness of fit for the extreme value distribution. Biometrika, 64, 583-588.