KYBERNETIKA — VOLUME 47 (2011), NUMBER 6, PAGES 814–839

RANDOMIZED GOODNESS OF FIT TESTS

F. Liese and B. Liu

Classical goodness of fit tests are no longer asymptotically distributional free if parameters are estimated. For a parametric model and the maximum likelihood estimator the empirical processes with estimated parameters is asymptotically transformed into a time transformed Brownian bridge by adding an independent Gaussian process that is suitably constructed. This randomization makes the classical tests distributional free. The power under local alternatives is investigated. Computer simulations compare the randomized Cramér–von Mises test with tests specially designed for location-scale families, such as the Shapiro–Wilk and the Shenton–Bowman test for normality and with the Epps–Pulley test for exponentiality.

Keywords: goodness of fit tests with estimated parameters, Kolmogorov–Smirnov test, Cramér–von Mises test, randomization

Classification: 64E17, 62E20

1. INTRODUCTION

Classical goodness of fit tests for the simple null hypothesis $H_0 : F = F_0$ versus the alternative $H_A : F \neq F_0$ are based on suitable functionals of the empirical process $\mathsf{G}_n = \sqrt{n}(\widehat{F}_n - F_0)$ where

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

is the empirical distribution function. If a functional of G_n is invariant under time transformation then the distribution of the resulting test statistic does not depend on F_0 under the null hypothesis. Examples are the Kolmogorov–Smirnov and the Cramér–von Mises statistic. A principal component analysis of the power of the Cramér–von Mises test under local alternatives can be obtained from the Kac–Siegert decomposition of the Brownian bridge.

The case of a simple null hypothesis is atypical in applications. In order to test whether the sample $X_1, \ldots, X_n$ originates from the parametric model $(F_\theta)_{\theta \in \Theta}$ a consistent estimator $\widehat{\theta}_n$ is plugged in F_θ and the Kolmogorov–Smirnov, the Cramér–von Mises statistic or other functionals are applied to the estimated empirical process

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

Durbin [5] and [6] was the first to show that $\widehat{\mathsf{G}}_n$ tends in distribution to some centered Gaussian process Z_θ with a covariance function $K_\theta = cov(Z_\theta(s), Z_\theta(t))$ that depends on the unknown θ. This implies that the classical goodness of fit tests are no longer asymptotically distributional free, the critical values for the tests are unknown and depend on the unknown θ. For a location-scale model $\frac{1}{\sigma} f((t-\mu)/\sigma)$ the covariance function $K_\theta, \theta = (\mu, \sigma)$ satisfies

$$K_\theta(s,t) = K_{0,1}\left(\frac{s-\mu}{\sigma}, \frac{t-\mu}{\sigma}\right).$$

This invariance property implies that the asymptotic quantiles of the Kolmogorov–Smirnov and the Cramér–von Mises statistics depend only on the parent density f but not on the concrete value of $\theta = (\mu, \sigma)$. For special location-scale families the corresponding critical values were obtained by numerical methods and simulations, see [4] and [15] for details and an overview. Using these critical values one easily obtains an asymptotic goodness of fit test.

A principal component analysis of the power of goodness of fit tests with parameters estimated is impossible, in general, as the eigenfunctions and eigenvalues of $K_{0,1}$ can be obtained by numerical methods only, see [15], p. 236.

Outside of the class of location-scale models the asymptotic quantiles of the Cramér–von Mises statistic and the Kolmogorov–Smirnov statistic depend directly on the unknown parameter and the tests are even not asymptotically distributional free if the null hypothesis is a special parametrized family. As a way out bootstrap techniques can be applied. For details and simulation results we refer to [7, 18, 19] and the references therein. The general message is that bootstrap methods work well under the null hypothesis. But there is no systematic method to study the power under local alternatives.

A breakthrough was achieved by Khmaladze [10] who constructed a kernel $\mathcal{K}_\theta(t,\tau)$ that transforms the estimated empirical process $\widehat{\mathsf{G}}_n = \sqrt{n}(\widehat{F}_n - F_{\widehat{\theta}_n})$ into the Wiener process. This transformation makes the classical goodness of fit tests asymptotically distributional free and provides, on the other side, a principal component analysis of the power where the known eigenfunctions and eigenvalues of the covariance function of the Wiener process are employed. The difficulty in applications is that $\mathcal{K}_\theta(t,\tau)$ depends in a convoluted manner on the score function of the model and can be evaluated numerically only. Explicit expressions for $\mathcal{K}_\theta(t,\tau)$ are obtained in [8] for the family of exponential distributions. Even for the family of normal distributions the kernel $\mathcal{K}_\theta(t,\tau)$ can not be evaluated in a closed manner.

To overcome the above mentioned difficulties we use the MLE $\widehat{\theta}_n$ and introduce a randomization of $\widehat{\mathsf{G}}_n$ by adding a suitably constructed centered Gaussian process $R_{n,\theta}$ that is independent of $\widehat{\mathsf{G}}_n$. The background for this new transformation is that the covariance function of $\widehat{\mathsf{G}}_n$ is asymptotically not larger than the covariance function of a time transformed Brownian bridge. One of the main results of this paper is that $\widehat{\mathsf{G}}_n + R_{n,\theta}$ and $\widehat{\mathsf{G}}_n + R_{n,\widehat{\theta}_n}$ tend in distribution to $\mathsf{B}(F_\theta)$, where B denotes the Brownian bridge.

To study the asymptotic power of the tests we consider local alternatives $P_n \notin (P_\theta)_{\theta\in\Theta}$ for which the sequence $\{P_n^{\otimes n}\}$ is contiguous with respect to $\{P_\theta^{\otimes n}\}$ and has the tangent a, where $\theta \in \Theta$ ist the true parameter. We show that $\widehat{\mathsf{G}}_n + R_{n,\widehat{\theta}_n}$ tends under $P_n^{\otimes n}$ in distribution to

$$\mathsf{B}(F_\theta) + \int I_{(-\infty,\cdot)}(s)(\Pi_\theta^\perp a)(s)P_\theta(\mathrm{d}s), \tag{1}$$

where $\Pi_\theta^\perp a$ is the projection of a on the orthogonal complement of the tangent space of the model $(P_\theta)_{\theta\in\Theta}$ at θ. The application of the principal component decomposition of the Brownian bridge to (1) gives an asymptotic approximation for the power of the Cramér–von Mises test if the observed data deviate from the model in the direction a. This means that for a given model we can find the directions for which the Cramér–von Mises test has good power, for which the power is medium and directions for which the power is poor.

In the last section we check the actual size of selected randomized goodness of fit tests by computer simulations. Finally, we compare the power of the randomized Cramér–von Mises test for the family of normal distributions and for the family of exponential distributions with tests that are specially designed for these special location-scale models. The general conclusion is that these tests have a better performance than our new randomized tests but the difference of the power functions is not big. This result is easy to explain. The new principle of making goodness of fit tests distributional free is universal and applicable to any parameterized model. Therefore these new tests will hardly achieve or even extend the power of tests that are tailor made for location-scale models.

2. EMPIRICAL DISTRIBUTION FUNCTIONS WITH ESTIMATED PARAMETERS

For i.i.d. $X_1, X_2, \ldots$ with common distribution function F we denote by

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

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 equipped with the uniform metric and the σ-algebra $\mathfrak{B}$ generated by the balls or, equivalently, by the coordinate projections. The Donsker theorem, see [13], p. 97, originally proved for $\mathbb{D}[0,1]$, states

$$\mathcal{L}(\mathsf{G}_n) \Rightarrow \mathcal{L}(\mathsf{B}(F)), \tag{2}$$

where $\Rightarrow$ is the symbol for weak convergence, i. e. $E\varphi(\mathsf{G}_n) \to E\varphi(\mathsf{B}(F))$ for every bounded and continuous function that is $\mathfrak{B}$-measurable. B denotes the Brownian bridge on $[0,1]$ that is a centered continuous Gaussian process with covariance function

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

Classical goodness of fit tests for the simple null hypothesis $H_0 : F = F_0$ versus the alternative $H_A : F \neq F_0$ are based on functionals of G_n. Introduce the Kolmogorov–Smirnov and the Cramér–von Mises statistic, respectively, by

$$\mathsf{K}_n = \sup\nolimits_t |\mathsf{G}_n(t)| \quad \text{and} \quad \mathsf{C}_n = \int \mathsf{G}_n^2(t)\,\mathrm{d}\widehat{F}_n(t).$$

If F is continuous then the relation (2) implies

$$\mathcal{L}(\mathsf{K}_n) \quad \Rightarrow \quad \mathcal{L}\left(\sup_{-\infty\le t\le\infty} |B(F(t))|\right) = \mathcal{L}\left(\sup_{0\le s\le 1} |B(s)|\right) = \mathcal{K}, \tag{3}$$

$$\mathcal{L}(\mathsf{C}_n) \quad \Rightarrow \quad \mathcal{L}\left(\int B^2(F(t))\,\mathrm{d}F(t)\right) = \mathcal{L}\left(\int_0^1 B^2(s)\,\mathrm{d}s\right) = \mathcal{C}. \tag{4}$$

The distributions $\mathcal{K}$ and $\mathcal{C}$ are well known and have the distribution functions

$$\begin{aligned}\mathcal{K}([0,x]) &= 1 - 2\sum\nolimits_{k=1}^{\infty}(-1)^{k+1}\exp\{-2k^2x^2\},\\ \mathcal{C}([0,x]) &= P\left(\sum\nolimits_{k=1}^{\infty}\frac{1}{(k\pi)^2}Z_k^2 \le x\right),\end{aligned}$$

where the $Z_1, Z_2, \ldots$ 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 Cramér–von Mises distribution $\mathcal{C}$, respectively. Based on the statistics K_n and C_n the Kolmogorov–Smirnov test and the Cramér–von Mises test are defined by

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

These tests are asymptotic level α tests by construction. Unfortunately, in statistical applications the case of a simple null hypothesis is unusual. Instead, one is often faced with the problem to test whether the common distribution of the i.i.d. sample $X_1, \ldots, X_n$ originates from the model $(P_\theta)_{\theta\in\Theta}, \Theta \subseteq \mathbb{R}_d$, or not. To this end one sets $F_\theta(t) = P_\theta((-\infty, t])$ and compares 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 [5] and [6]. The results of different authors differ in the type of regularity conditions that are necessary to make a suitable Taylor expansion, see [15], Section 5.5. Our approach follows [7] and [20]. 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

$$\sqrt{n}(\widehat{\theta}_n(X_1, \ldots, X_n) - \theta) = \frac{1}{\sqrt{n}}\sum\nolimits_{i=1}^{n} h_\theta(X_i) + o_P(1), \tag{5}$$

$$\begin{aligned}&E_\theta\, \|h_\theta(X_1)\|^2 < \infty, \quad E_\theta h_\theta(X_1) = 0, \qquad\qquad (6)\\ &J(\theta) := E_\theta h_\theta(X_1) h_\theta^T(X_1).\end{aligned}$$

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 for every $t \in [-\infty, \infty]$ the function $\theta \longmapsto F_\theta(t)$, $\theta \in U(\theta_0)$ is differentiable and set

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

Furthermore, we suppose that

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

The next version of Durbin's Theorem was proved in [7] and [20].

Theorem 2.1. Under the assumptions (5), (6) and (7) it holds

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

and $\mathcal{L}(\widehat{\mathsf{G}}_n) \Rightarrow \mathcal{L}(Z_\theta)$, where Z_θ is a centered and continuous Gaussian process with covariance function

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

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 μ that is atomless. Denote by $f_\theta = \mathrm{d}P_\theta / \mathrm{d}\mu, \theta \in \Theta$, the corresponding densities. As μ is atomless the distribution functions

$$F_\theta(t) = \int I_{(-\infty,t]}(s) f_\theta(s) \mu(\mathrm{d}s), \quad \theta \in \Theta, \tag{10}$$

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

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

where $\dot{f}_\theta := (\frac{\partial}{\partial \theta_1} f_\theta, \ldots, \frac{\partial}{\partial \theta_d} f_\theta)^T$. The last condition in (11) means that one may interchange the derivative with respect to θ and the integral with respect to s 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 (5) with

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

Moreover, the function $h_\theta = I^{-1}(\theta)\dot{l}_\theta$ satisfies (6) so that especially

$$\int \dot{l}_\theta \,\mathrm{d}P_\theta = 0, \tag{13}$$

see e. g. Theorem 1.117 and Proposition 1.110 in [11].

We calculate the covariance function in Theorem 2.1 if $\widehat{\theta}_n$ is the MLE. It holds

$$\begin{aligned} & \dot{F}_\theta^T(t) E h_\theta(X_1) I_{(-\infty,t]}(X_1) \\ = \;& \dot{F}_\theta^T(t) \int I_{(-\infty,t]}(x) I^{-1}(\theta) \frac{\dot{f}_\theta(x)}{f_\theta(x)} f_\theta(x)\mu(\mathrm{d}x) \\ = \;& \dot{F}_\theta^T(t) I^{-1}(\theta) \int I_{(-\infty,t]}(x) \dot{f}_\theta(x)\mu(\mathrm{d}x) \\ = \;& \dot{F}_\theta^T(t) I^{-1}(\theta)\dot{F}_\theta(t), \end{aligned} \tag{14}$$

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

$$cov(Z_\theta(s), Z_\theta(t)) = F_\theta(s \wedge t) - F_\theta(s)F_\theta(t) - \dot{F}_\theta^T(t) I^{-1}(\theta)\dot{F}_\theta(s). \tag{15}$$

Corollary 2.2. Suppose that the conditions in (11) are satisfied and the MLE $\widehat{\theta}_n$ satisfies (5) with $h_\theta = I^{-1}(\theta)\dot{l}_\theta$. If (7) holds and the MLE $\widehat{\theta}_n$ is used to construct the estimated empirical process $\widehat{\mathsf{G}}_n$ then $\mathcal{L}(\widehat{\mathsf{G}}_n) \Rightarrow \mathcal{L}(Z_\theta)$, where Z_θ is a centered and continuous Gaussian process whose covariance function is given in (15).

Example 2.3. Let $\mathsf{N}(\mu, \sigma^2)$ be the normal distribution with unknown expectation μ and variance σ^2. Let Φ and φ be the distributions function and the density of $\mathsf{N}(0,1)$, respectively. Then $P_\theta = \mathsf{N}(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2), \mu \in \mathbb{R}$, $\sigma^2 > 0$, $F_\theta(t) = \Phi\left(\frac{t-\mu}{\sigma}\right)$ and

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

All assumptions in Corollary 2.2 are fulfilled. The MLE and the empirical process with estimated parameters are

$$\begin{aligned} \widehat{\theta}_n &= (\overline{X}_n, S_n^2)^T, \quad \text{and} \quad S_n^2 = \frac{1}{n}\sum\nolimits_{i=1}^n (X_i - \overline{X}_n)^2 \\ \widehat{\mathsf{G}}_n(t) &= \sqrt{n}\left(\widehat{F}_n(t) - \Phi\left(\frac{t-\overline{X}_n}{S_n}\right)\right). \end{aligned}$$

Finally,

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

is the inverse of the Fisher information matrix and the covariance function in (15) is given by

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

Example 2.4. Let $G_\lambda, \lambda > 0$, be the family of exponential distributions with parameter $\lambda > 0$

$$G_\lambda(t) = I_{[0,\infty)}(t)(1 - \exp\{-\lambda t\}).$$

The empirical process with estimated parameters is

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

and the covariance function in (15) is given by

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

3. RANDOMIZED EMPIRICAL PROCESS

In this section we introduce a new transformation to make goodness of fit tests with estimated parameters asymptotically distributional free. Our starting point is the structure of the asymptotic covariance function $cov(Z_\theta(s), Z_\theta(t))$ in (15) which is the difference of the positive semidefinite function $F_\theta(s \wedge t) - F_\theta(s)F_\theta(t)$ being the covariance function of $\mathsf{B}(F_\theta)$ and the positive semidefinite function $\dot{F}_\theta^T(t)I^{-1}(\theta)\dot{F}_\theta(s)$. Let $I^{-1/2}(\theta)$ be the symmetric and positive definite matrix with $I^{-1/2}(\theta)I^{-1/2}(\theta) = I^{-1}(\theta)$ and suppose that V is a d dimensional random vector with i.i.d. standard normal components. The stochastic process

$$R_\theta(t) = \dot{F}_\theta^T(t)I^{-1/2}(\theta)V \tag{16}$$

has the covariance function

$$cov(R_\theta(s), R_\theta(t)) = \dot{F}_\theta^T(t)I^{-1}(\theta)\dot{F}_\theta(s).$$

If V is independent of Z_θ then

$$\begin{aligned} cov(Z_\theta(s) + R_\theta(s), Z_\theta(t) + R_\theta(t)) &= cov(Z_\theta(s), Z_\theta(t)) + cov(R_\theta(s), R_\theta(t)) \\ &= F_\theta(s \wedge t) - F_\theta(s)F_\theta(t) \\ &= cov(\mathsf{B}(F_\theta(s), \mathsf{B}(F_\theta(t)). \end{aligned} \tag{17}$$

This simple relation is the basic idea for our randomization technique. Suppose that V_n is standard normal and independent of $X_1, \ldots, X_n$. By adding

$$R_{n,\theta}(t) = \dot{F}_\theta^T(t) I^{-1/2}(\theta) V_n \tag{18}$$

to the estimated empirical process $\widehat{\mathsf{G}}_n$ we transform this process asymptotically to a time transformed Brownian bridge. The processes $R_{n,\theta}$ still depend on the unknown parameter θ. As this dependence appears as a 'factor' we may replace θ by an consistent estimator. We call $\widehat{\mathsf{G}}_n + R_{n,\widehat{\theta}_n}$ the *randomized estimated empirical process.* Note that $\widehat{\mathsf{G}}_n + R_{n,\widehat{\theta}_n}$ is a function of the random vector $X_1, \ldots, X_n, V_n$ that has the distribution $P_\theta^{\otimes n} \otimes \mathsf{N}(0, I)$.

Theorem 3.1. Suppose the conditions in Corollary 2.2 are fulfilled and V_n are standard normal random vectors that are independent of $X_1, \ldots, X_n$. Then

$$\mathcal{L}(\widehat{\mathsf{G}}_n + R_{n,\theta} | P_\theta^{\otimes n} \otimes \mathsf{N}(0, I)) \Rightarrow \mathcal{L}(\mathsf{B}(F_\theta))$$

and

$$\mathcal{L}(\widehat{\mathsf{G}}_n + R_{n,\widehat{\theta}_n} | P_\theta^{\otimes n} \otimes \mathsf{N}(0, I)) \Rightarrow \mathcal{L}(\mathsf{B}(F_\theta)).$$

Proof. The convergence

$$\mathcal{L}(\widehat{\mathsf{G}}_n(\theta) + R_{n,\theta} | P_\theta^{\otimes n} \otimes \mathsf{N}(0, I)) \Rightarrow \mathcal{L}(Z_\theta + R_\theta)$$

follows from the independence of $\widehat{\mathsf{G}}_n$ and $R_{n,\theta}$, the continuous mapping theorem, and Corollary 2.2. Therefore the first statement follows from (17). The second statement follows from Slutsky's lemma, the first condition in (7) and the continuity of $I(\theta)$ which implies

$$\sup_t |R_{n,\widehat{\theta}_n}(t) - R_\theta(t)| \xrightarrow{P_\theta^{\otimes n} \otimes \mathsf{N}(0,I)} 0. \tag{19}$$

□

We introduce the *randomized Kolmogorov–Smirnov statistic* by

$$\mathsf{RK}_n = \sup_t |\widehat{\mathsf{G}}_n(t) + R_{n,\widehat{\theta}_n}(t)|. \tag{20}$$

To introduce a Cramér–von Mises type statistic we denote by $X_{n:1} \leq \cdots \leq X_{n:n}$ the order statistic and define the *randomized Cramér–von Mises statistic* by

$$\mathsf{RC}_n := \sum\nolimits_{i=1}^n \left(\frac{i}{n} - F_{\widehat{\theta}_n}(X_{n:i}) + \frac{1}{\sqrt{n}} R_{n,\widehat{\theta}_n}(X_{n:i}) \right)^2. \tag{21}$$

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

$$\begin{aligned} \mathsf{RC}_n &= \sum\nolimits_{i=1}^n \left(\frac{i}{n} - F_{\widehat{\theta}_n}(X_{n:i}) + \frac{1}{\sqrt{n}} R_{n,\widehat{\theta}_n}(X_{n:i}) \right)^2 \\ &= \int \left(\sqrt{n}(\widehat{F}_n(t) - F_{\widehat{\theta}_n}(t)) + \dot{F}_{\widehat{\theta}_n}^T(t) I^{-1/2}(\widehat{\theta}_n) V_n \right)^2. \end{aligned} \tag{22}$$

Based on the test statistics RK_n and RC_n we introduce the *randomized Kolmogorov–Smirnov test* and the *randomized Cramér–von Mises test* by

$$\varphi_{\mathsf{RK}_n} = I_{[k_{1-\alpha},\infty)}(\mathsf{RK}_n) \quad \text{and} \quad \varphi_{\mathsf{RC}_n} = I_{[c_{1-\alpha},\infty)}(\mathsf{RC}_n). \tag{23}$$

Proposition 3.2. Under the assumptions of Theorem 3.1 it holds

$$\begin{aligned}\mathcal{L}(\mathsf{RK}_n) \quad &\Rightarrow \quad \mathcal{L}\left(\sup_{0\le s\le 1}|B(s)|\right) = \mathcal{K} \quad \text{and} \\ \mathcal{L}(\mathsf{RC}_n) \quad &\Rightarrow \quad \mathcal{L}\left(\int_0^1 B^2(s)\,\mathrm{d}s\right) = \mathcal{C}.\end{aligned}$$

The tests φ_{RK_n} and φ_{RC_n} are asymptotic level α-tests for testing

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

Proof. As the sup-norm is continuous on $\mathbb{D}[-\infty,\infty]$ and measurable with respect to the σ-algebra generated by the projections we get the first statement from Theorem 3.1 and the equality in (3). The second statement follows from Theorem 3.1, Lemma 6.6 in the Appendix and the equality in (4). The fact that φ_{RK_n} and φ_{RC_n} are asymptotic level α-tests follows from the weak convergence of $\mathcal{L}(\mathsf{RK}_n)$ and $\mathcal{L}(\mathsf{RC}_n)$ combined with the continuity of $\mathcal{K}$ and $\mathcal{C}$. □

4. POWER OF THE RANDOMIZED CRAMER-VON MISES TEST UNDER LOCAL ALTERNATIVES

In order to investigate the asymptotic power of the goodness of fit tests constructed in the previous section we use concepts from LeCam's asymptotic decision theory. In the first part of this section we collect some of these results for any parametric model and apply these facts to the goodness of fit problem in the second part.

Given a measurable space $(\mathcal{X},\mathfrak{A})$ and a distribution P on $(\mathcal{X},\mathfrak{A})$ we denote by $L_2(P)$ be the space of all P-square integrable functions $a : \mathcal{X} \to \mathbb{R}$. For the null hypothesis $H_0 : P$ we study the power of tests for alternatives $H_A : P_n$ for which the sequence $\{P_n^{\otimes n}\}$ is contiguous with respect to the sequence $\{P^{\otimes n}\}$ ($P_n^{\otimes n} \lhd P^{\otimes n}$), i. e.

$$\lim_{n\to\infty} P^{\otimes n}(A_n) = 0 \quad \Rightarrow \quad \lim_{n\to\infty} P_n^{\otimes n}(A_n) = 0, \tag{24}$$

which is equivalent with the uniform integrability of likelihood ratios $\mathrm{d}P_n^{\otimes n}/\mathrm{d}P^{\otimes n}$ with respect to $P^{\otimes n}$, see e. g. Theorem 6.26 in [11].

To construct a sequence $\{P_n\}$ that satisfies (24) we suppose $P_n \ll P$, set

$$g_n = \frac{\mathrm{d}P_n}{\mathrm{d}P} \quad \text{and} \quad a_n = 2\sqrt{n}(\sqrt{g_n} - 1) \tag{25}$$

and assume that the sequence $a_n \in L_2(P)$ converges to $a \in L_2(P)$, i. e.

$$\lim_{n\to\infty} \int (a_n(x) - a(x))^2 P(\mathrm{d}x) = 0. \tag{26}$$

Then (24) is satisfied in view Lemma 6.2 in the Appendix.

For $r_n = a_n - a$ the definition of a_n yields

$$\sqrt{g_n} - 1 = \frac{1}{2\sqrt{n}} a + \frac{1}{2\sqrt{n}} r_n,$$

where $\int r_n^2 \, \mathrm{d}P \to 0$, if (26) is satisfied. Therefore we call a the *tangent* of the sequence $\{P_n\}$. According to Lemma 6.1 in the Appendix the set of all possible tangents is just the space

$$L_2^0(P) = \left\{ a : a \in L_2(P), \int a \, \mathrm{d}P = 0 \right\}.$$

For a bounded $a \in L_2^0(P)$ the construction of a sequence $\{P_n\}$ with tangent a is easy. Indeed, $g_n = 1 + \frac{1}{\sqrt{n}} a$ is a probability density for all sufficiently large n, so that

$$P_n(A) = \int_A \left(1 + \frac{1}{\sqrt{n}} a \right) \mathrm{d}P \tag{27}$$

becomes a probability measure. Using a first order Taylor expansion for $\sqrt{1+x}$ one can easily verify that a_n in (25) with $g_n = (1 + \frac{1}{\sqrt{n}} a)$ satisfies (26).

Now we consider a parametric model $\mathcal{P} = (P_\theta)_{\theta \in \Theta}, \Theta \subseteq \mathbb{R}_d$, that satisfies the condition (11). Suppose that θ is an interior point of Θ. For $u \in \mathbb{R}_d$ with $\theta + u \in \Theta$ we set

$$\begin{aligned} L_\theta(u) &= \frac{\mathrm{d}P_{\theta+u}}{\mathrm{d}P_\theta} = \frac{f_{\theta+u}}{f_\theta} \\ \dot{l}_\theta &= \left(\frac{1}{f_\theta} \frac{\partial f_\theta}{\partial \theta_1}, \ldots, \frac{1}{f_\theta} \frac{\partial f_\theta}{\partial \theta_d} \right)^T. \end{aligned}$$

It is well-known, see e. g. Theorem 1.117 in [11], that the condition (11) implies

$$\int \left(\sqrt{L_\theta(u)} - 1 - \frac{1}{2} u^T \dot{l}_\theta \right)^2 \mathrm{d}P_\theta = o(\|u\|^2). \tag{28}$$

Fix $h \in \mathbb{R}_d$ and put $u = h/\sqrt{n}$. For

$$P = P_\theta, \quad P_n = P_{\theta + h/\sqrt{n}},$$

we get $g_n = L_\theta(h/\sqrt{n})$. The relation (28) shows that (26) is satisfied with

$$a_n = 2\sqrt{n}(\sqrt{g_n} - 1), \quad a = h^T \dot{l}_\theta.$$

Consequently, the sequence $\{P_n\}$ has the tangent $h^T \dot{l}_\theta$ which belongs to $L_2^0(P_\theta)$ in view of (13). We introduce the *tangent space* $\mathcal{T}(\theta)$ of the model $\mathcal{P}$ at θ by

$$\mathcal{T}(\theta) = \{ h^T \dot{l}_\theta, h \in \mathbb{R}_d \}$$

and see that $\mathcal{T}(\theta) \subseteq L_2^0(P_\theta)$ is the set of all tangents of sequences $P_n = P_{\theta+h/\sqrt{n}}, h \in \mathbb{R}_d$, that originates from the model. Note that according to Lemma 6.1 in the

Appendix the larger space $L_2^0(P_\theta)$ is the space of tangents of all sequences that do not necessarily originate from the model $\mathcal{P}$.

Let $\Pi_\theta a$ denote the projection of $a \in L_2^0(P_\theta)$ on $\mathcal{T}(\theta)$. It will turn out that the asymptotic power of the goodness of fit tests under local alternatives P_n depend on the projection $\Pi_\theta^\perp a$ of the tangent a of the sequence $\{P_n\}$ on the orthogonal complement $\mathcal{T}^\perp(\theta)$ of the tangent space. Therefore we need explicit expressions for $\Pi_\theta a$ and $\Pi_\theta^\perp a$.

Lemma 4.1. Suppose that the condition (11) is fulfilled. Then

$$\Pi_\theta a = c^T(\theta) I^{-1}(\theta)\dot{l}_\theta \quad \text{and} \quad \Pi_\theta^\perp a = a - c^T(\theta) I^{-1}(\theta)\dot{l}_\theta, \tag{29}$$

where $c(\theta) = \int (\dot{l}_\theta a)\, \mathrm{d}P_\theta$.

P r o o f. $\Pi_\theta a$ is characterized by the conditions $\Pi_\theta a \in \mathcal{T}(\theta)$ and $a - \Pi_\theta a \perp \mathcal{T}(\theta)$ or $a - \Pi_\theta a \perp h^T \dot{l}_\theta$ for every h. The first condition is $c^T(\theta) I^{-1}(\theta)\dot{l}_\theta \in \mathcal{T}(\theta)$ and is clear from the definition of $\mathcal{T}(\theta)$. To establish the second condition we note that $I(\theta) = \int \dot{l}_\theta \dot{l}_\theta^T \, \mathrm{d}P_\theta$ implies for every $h \in \mathbb{R}_d$

$$\int h^T \dot{l}_\theta (a - \dot{l}_\theta^T I^{-1}(\theta) c(\theta))\, \mathrm{d}P_\theta = h^T c(\theta) - h^T I(\theta) I^{-1}(\theta) c(\theta) = 0.$$

□

Subsequently we need the asymptotic behavior of the linear statistics

$$T_n = \frac{1}{\sqrt{n}} \sum\nolimits_{i=1}^n b_n(X_i) + o_{P^{\otimes n}}(1). \tag{30}$$

Proposition 4.2. Suppose that $X_1, \ldots, X_n$ are i.i.d. with common distribution $P_n \ll P$ and assume that the condition (26) is fulfilled with $g_n = \mathrm{d}P_n/\mathrm{d}P$ and a_n from (25). If $b_n \in L_2^0(P)$ satisfies

$$\lim_{n\to\infty} \int (b_n - b)^2 \, \mathrm{d}P = 0. \tag{31}$$

for some $b \in L_2(P)$ then

$$\mathcal{L}(T_n | P_n^{\otimes n}) \Rightarrow \mathsf{N}\left(\int ab \, \mathrm{d}P, \int b^2 \, \mathrm{d}P\right). \tag{32}$$

P r o o f. The condition (31) and the lemma of Slutzky implies that we may replace b_n with b in (30) which together with the representations (51) in Lemma 6.2 in the Appendix yields

$$\begin{pmatrix} T_n \\ \ln L_n \end{pmatrix} = \frac{1}{\sqrt{n}} \sum\nolimits_{i=1}^n \begin{pmatrix} b(X_i) \\ a(X_i) \end{pmatrix} + \begin{pmatrix} 0 \\ -\frac{1}{2}\int_0^1 a^2(x) P(\mathrm{d}x) \end{pmatrix} + o_{P^{\otimes n}}(1).$$

The central limit theorem provides

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

From here and LeCam's third lemma, see Theorem 4 in [15], p. 154, we get (32). □

Now we apply (32) to study $\widehat{\mathsf{G}}_n + R_{n,\widehat{\theta}_n}$ under $P_n^{\otimes n} \otimes \mathcal{L}(V_n)$ where $\mathcal{L}(V_n) = \mathsf{N}(0, I)$ is the d-dimensional standard normal distribution which is the common distribution of the V_n that have been used in (18) to construct $R_{n,\theta}$.

Theorem 4.3. Suppose the assumptions in Theorem 3.1 are fulfilled. If $P_n \ll P_\theta$ and the conditions (25) and (26) are satisfied with $P = P_\theta$ then

$$\mathcal{L}(\widehat{\mathsf{G}}_n + R_{n,\widehat{\theta}_n} | P_n^{\otimes n} \otimes \mathsf{N}(0, I)) \Rightarrow \mathcal{L}(\mathsf{B}(F_\theta) + \Lambda_\theta a),$$

where $R_{n,\widehat{\theta}_n} = \dot{F}_{\widehat{\theta}_n}^T I^{-1/2}(\widehat{\theta}_n) V_n$ and the shift term $\Lambda_\theta a$ is defined by

$$(\Lambda_\theta a)(t) = \int \Lambda_\theta(t, s) a(s) P_\theta(\mathrm{d}s) \tag{33}$$

$$\Lambda_\theta(t, s) = I_{(-\infty,t]}(s) - F_\theta(t) - \dot{F}_\theta^T(t) I^{-1}(\theta) \dot{l}_\theta(s). \tag{34}$$

Corollary 4.4. Under the assumptions in Theorem 3.1 it holds

$$\mathcal{L}(\widehat{\mathsf{G}}_n + R_{n,\widehat{\theta}_n} | P_n^{\otimes n} \otimes \mathsf{N}(0, I)) \Rightarrow \mathcal{L}\left(\mathsf{B}(F_\theta) + \int I_{(-\infty,\cdot]}(s)(\Pi_\theta^\perp a)(s) P_\theta(\mathrm{d}s)\right),$$

where $\Pi_\theta^\perp a$ is the projection of the tangent a of the sequence $\{P_n\}$ on the orthogonal complement of the tangent space $\mathcal{T}(\theta)$ of the model $(P_\theta)_{\theta\in\Theta}$ at θ.

P r o o f. First of all we note that the Lemmas 6.2 and 6.3 in the Appendix provide

$$P_n^{\otimes n} \otimes \mathsf{N}(0, I) \lhd P_\theta^{\otimes n} \otimes \mathsf{N}(0, I),$$

which implies in conjunction with (19) that we may replace $R_{n,\widehat{\theta}_n}$ with $R_{n,\theta}$ in the statement to be established. It holds

$$\begin{aligned} \mathcal{L}((\widehat{\mathsf{G}}_n, R_{n,\theta}) | P_n^{\otimes n} \otimes \mathsf{N}(0, I)) &= \mathcal{L}(\widehat{\mathsf{G}}_n | P_n^{\otimes n}) \otimes \mathcal{L}(R_{n,\theta} | \mathsf{N}(0, I)) \qquad (35) \\ &= \mathcal{L}(\widehat{\mathsf{G}}_n | P_n^{\otimes n}) \otimes \mathcal{L}(R_\theta | \mathsf{N}(0, I)) \end{aligned}$$

as $\mathcal{L}(V_n) = \mathsf{N}(0, I) = \mathcal{L}(V)$ and by the definition of R_θ in (16). Hence it remains to study the marginal distributions $\mathcal{L}(\widehat{\mathsf{G}}_n | P_n^{\otimes n})$ as $n \to \infty$. To prove the convergence of the finite dimensional distributions we apply the Cramér–Wold device and consider $\sum_{j=1}^n a_j \widehat{\mathsf{G}}_n(t_j)$ for fixed real numbers a_i. Then by (8), with $h_\theta = I^{-1}(\theta)\dot{l}_\theta$,

$$\sum\nolimits_{j=1}^n a_j \widehat{\mathsf{G}}_n(t_j) = \frac{1}{\sqrt{n}} \sum\nolimits_{i=1}^n b(X_i) + o_{P^{\otimes n}}(1),$$

where $b(s) = \sum_{j=1}^{n} a_j \Lambda_\theta(t_j, s)$. The application of Proposition 4.2 with $b_n = b$ yields

$$\mathcal{L}(\sum\nolimits_{j=1}^{n} a_j \widehat{\mathsf{G}}_n(t_j) | P_n^{\otimes n}) \Rightarrow \mathsf{N}\left(\int ab \, \mathrm{d}P_\theta, \int b^2 \, \mathrm{d}P_\theta\right). \tag{36}$$

It holds

$$\int ab \, \mathrm{d}P_\theta = \sum\nolimits_{j=1}^{n} a_j \int a(s)\Lambda_\theta(t_j, s) P_\theta(\mathrm{d}s). \tag{37}$$

To calculate $\int b^2 \, \mathrm{d}P_\theta$ we note that the conditions in (11) and consequently (13) imply for every u and v

$$\begin{aligned}
&\int \dot{F}_\theta^T(v) I^{-1}(\theta) \dot{l}_\theta(s) P_\theta(\mathrm{d}s) = 0, \\
&\int I_{(-\infty,u]}(s) \dot{F}_\theta^T(v) I^{-1}(\theta) \dot{l}_\theta(s) P_\theta(\mathrm{d}s) = \dot{F}_\theta^T(v) I^{-1}(\theta) \dot{F}_\theta(u), \\
&\int \dot{F}_\theta^T(u) I^{-1}(\theta) \dot{l}_\theta(s) \dot{F}_\theta^T(v) I^{-1}(\theta) \dot{l}_\theta(s) P_\theta(\mathrm{d}s) \\
= &\int \dot{F}_\theta^T(u) I^{-1}(\theta) \dot{l}_\theta(s) \dot{l}_\theta^T(s) I^{-1}(\theta) \dot{F}_\theta(v) P_\theta(\mathrm{d}s) = \dot{F}_\theta^T(u) I^{-1}(\theta) \dot{F}_\theta(v).
\end{aligned}$$

Furthermore,

$$\int [I_{(-\infty,u]}(s) - F_\theta(u)][I_{(-\infty,v]}(s) - F_\theta(v)] P_\theta(\mathrm{d}s) = F_\theta(u \wedge v) - F_\theta(u) \wedge F_\theta(v).$$

Combining these results we get

$$\int \Lambda_\theta(u, s) \Lambda_\theta(v, s) P_\theta(\mathrm{d}s) = cov(Z_\theta(u), Z_\theta(v)),$$

where Z_θ is the centered Gaussian process with covariance function (15). As Z_θ and $Z_\theta + \Lambda_\theta a$ have the same covariance function we get

$$\int b^2 \, \mathrm{d}P_\theta = \sum\nolimits_{i,j=1}^{n} a_i a_j cov(Z_\theta(t_i) + (\Lambda_\theta a)(t_i), Z_\theta(t_j) + (\Lambda_\theta a)(t_j)). \tag{38}$$

From (36), (37) and (38) we get the convergence of the finite dimensional distributions of $\widehat{\mathsf{G}}_n$ under $P_n^{\otimes n}$ to the process $Z_\theta + \Lambda_\theta a$. As $P_n^{\otimes n} \lhd P_\theta^{\otimes n}$ in view of Lemma 6.2 in the Appendix we may apply apply Lemma 6.5 in the Appendix to get

$$\mathcal{L}(\widehat{\mathsf{G}}_n | P_n^{\otimes n}) \Rightarrow \mathcal{L}(Z_\theta + \Lambda_\theta a).$$

Combining this result with (35) we get

$$\mathcal{L}((\widehat{\mathsf{G}}_n + R_{n,\theta}) | P_n^{\otimes n} \otimes \mathsf{N}(0, I)) \Rightarrow \mathcal{L}(Z_\theta + \Lambda_\theta a + R_\theta),$$

where Z_θ and R_θ are independent. To complete the proof of the theorem we use (17) to see that $Z_\theta + \Lambda_\theta a + R_\theta$ has the covariance function $cov(\mathsf{B}(F_\theta(s)), \mathsf{B}(F_\theta(t)))$.

To prove the Corollary we note that $\int a\,\mathrm{d}P_\theta = 0$ by Lemma 6.1 in the Appendix. This implies with $c(\theta)$ from Lemma 4.1

$$\int \Lambda_\theta(t,s)a(s)P_\theta(\mathrm{d}s) = \int [I_{(-\infty,t]}(s) - \dot{F}_\theta^T(t)I^{-1}(\theta)\dot{l}_\theta(s)]a(s)P_\theta(\mathrm{d}s)$$
$$= \int I_{(-\infty,t]}(s)a(s)P_\theta(\mathrm{d}s) - \dot{F}_\theta^T(t)I^{-1}(\theta)c(\theta).$$

From the last condition in (11) we get that the right hand term is

$$\int I_{(-\infty,t]}(s)[a(s) - c^T(\theta)I^{-1}(\theta)\dot{l}_\theta(s)]P_\theta(\mathrm{d}s)$$
$$= \int I_{(-\infty,t]}(s)(\Pi_\theta^\perp a)(s)P_\theta(\mathrm{d}s)$$

where the equality follows from Lemma 4.1. The proof is complete. □

Subsequently we need the Kac–Siegert decomposition of the Brownian bridge $\mathsf{B}(t), 0 \le t \le 1$, which has the covariance function $K(s,t) = s \wedge t - st$. The eigenvalues and the normalized eigenfunctions of K are

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

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

$$Z_k = k\pi \int_0^1 \mathsf{B}(t)\varphi_k(t)\,\mathrm{d}t$$

are i.i.d. standard normal. The system of eigenfunctions $\{\varphi_k\}$ is complete in $L_2[0,1]$. If F_θ is continuous then the mapping $\varphi \mapsto \varphi(F_\theta)$ is an isometry between $L_2[0,1]$ and $L_2(P_\theta)$ which implies that $\{\varphi_k(F_\theta)\}$ is a complete orthonormal system in $L_2(P_\theta)$. This yields for every function $C \in L_2(P_\theta)$

$$\mathcal{L}\left(\int (\mathsf{B}(F_\theta) + C)^2\,\mathrm{d}P_\theta\right) = \mathcal{L}\left(\sum\nolimits_{k=1}^{\infty} \frac{1}{(k\pi)^2}(Z_k + k\pi C_k)^2\right),$$
$$Z_k = k\pi \int \mathsf{B}(F_\theta)\varphi_k(F_\theta)\,\mathrm{d}P_\theta \quad \text{and} \quad C_k = \sqrt{2}\int C\sin(k\pi F_\theta)\,\mathrm{d}P_\theta. \qquad (39)$$

Theorem 4.5. Suppose that the assumptions in Theorem 4.3 are fulfilled where $a \in L_2^0(P_\theta)$ is the tangent of the sequence $\{P_n\}$. Then the asymptotic power of the randomized Cramér–von Mises test φ_{RC_n} in (23) under the local alternative $P_n^{\otimes n} \otimes \mathsf{N}(0,I)$ is given by

$$\lim_{n\to\infty} \int \varphi_{\mathsf{RC}_n}\,\mathrm{d}(P_n^{\otimes n} \otimes \mathsf{N}(0,I)) = P\left(\sum\nolimits_{k=1}^{\infty} \frac{1}{(k\pi)^2}(Z_k + d_k)^2 > c_{1-\alpha}\right),$$

where

$$d_k = \sqrt{2} \int (\Pi_\theta^\perp a) \cos(k\pi F_\theta) \, \mathrm{d}P_\theta. \tag{40}$$

P r o o f. The relation (22) gives

$$\mathsf{RC}_n = \int (\widehat{\mathsf{G}}_n + R_{n,\widehat{\theta}_n})^2 \, \mathrm{d}\widehat{F}_n.$$

Hence by Corollary 4.4 and Lemma 6.6 in the Appendix

$$\begin{aligned} \mathcal{L}(\mathsf{RC}_n) &\Rightarrow \mathcal{L}(\int (\mathsf{B}(F_\theta) + C)^2 \, \mathrm{d}P_\theta), \quad \text{where} \\ C(t) &= \int I_{(-\infty,t]}(s)(\Pi_\theta^\perp a)(s) P_\theta(\mathrm{d}s). \end{aligned}$$

It remains to calculate C_k in (39). Changing the integration with respect to s and t we obtain

$$C_k = \sqrt{2} \int \left(\int I_{(-\infty,t]}(s) \sin(k\pi F_\theta(t)) P_\theta(\mathrm{d}t) \right) (\Pi_\theta^\perp a)(s) P_\theta(\mathrm{d}s).$$

As F_θ is continuous we get

$$\begin{aligned} &\int I_{(-\infty,t]}(s) \sin(k\pi F_\theta(t)) P_\theta(\mathrm{d}t) = \int I_{(-\infty,F_\theta(t)]}(F_\theta(s)) \sin(k\pi F_\theta(t)) P_\theta(\mathrm{d}t) \\ = &\int_{F_\theta(s)}^1 \sin(k\pi u) \, \mathrm{d}u = \frac{1}{k\pi}(\cos(k\pi F_\theta(s)) - \cos(k\pi)). \end{aligned}$$

From (13), (29), and $a \in L_2^0(P_\theta)$ we conclude $\Pi_\theta^\perp a \in L_2^0(P_\theta)$ which yields

$$\begin{aligned} C_k &= \sqrt{2} \int \frac{1}{k\pi}(\cos(k\pi F_\theta(s)) - \cos(k\pi))(\Pi_\theta^\perp a)(s) P_\theta(\mathrm{d}s) \\ &= \frac{1}{k\pi}\sqrt{2} \int \cos(k\pi F_\theta(s))(\Pi_\theta^\perp a)(s) P_\theta(\mathrm{d}s) = \frac{1}{k\pi} d_k. \end{aligned}$$

□

Theorem 4.5 can be used to give a local approximation of the power of the Cramér–von Mises test if the data have the distribution P_n and the sequence $\{P_n\}$ has the tangent a. We consider the Gaussian location model. Then $F_\mu(t) = \Phi(t - \mu)$, $P_\mu = \mathsf{N}(\mu, 1)$ and

$$\dot{l}_\mu(t) = t - \mu \quad \text{and} \quad I(\mu) = \int \dot{l}_\mu^2 \, \mathrm{d}P_\mu = 1.$$

Without loss of generality we assume $\mu = 0$ and set $\varphi = \Phi'$. Then in view of (29) the projection of any $a \in L_2^0(P_\mu)$ on the tangent space of the model and its orthogonal complement at $\mu = 0$ are given by

$$(\Pi_0 a)(t) = \left(\int a(s) s \varphi(s) \, \mathrm{d}s \right) \cdot t \quad \text{and} \quad \Pi_0^\perp a = a - \Pi_0 a. \tag{41}$$

We will calculate the asymptotic power of the randomized Cramér–von Mises test along a sequence of distributions which have the pregiven tangents

$$a_{l,h} = h\varphi_l(\Phi) \quad \text{where} \quad \varphi_l(s) = \sqrt{2}\cos(l\pi s). \tag{42}$$

The additional parameter h controls the distance from the null hypothesis. As φ_l is bounded we may use the construction in (27). For $a(t) = a_{l,h} = h\varphi_l(\Phi)$ the alternatives in (27) are given by

$$P_{l,n,h}(A) = \int_A (1 + \frac{h}{\sqrt{n}}\varphi_l(\Phi))\varphi(t)\,\mathrm{d}t, \quad -1/\sqrt{2} < \frac{h}{\sqrt{n}} < 1/\sqrt{2} \tag{43}$$

and have the Lebesgue densities for $l = 1, 2, ..$

$$p_{l,\eta}(t) = (1 + \eta\varphi_l(\Phi))\varphi(t), \quad -1/\sqrt{2} < \eta = \frac{h}{\sqrt{n}} < 1/\sqrt{2}.$$

If $l = 2m$ is even then $\Phi(-t) = 1 - \Phi(t)$ yields $p_{2m,\eta}(-t) = p_{2m,\eta}(t)$. For odd $l = 2m + 1$ we get

$$\begin{aligned} p_{2m+1,\eta}(-t) &= (1 + \eta\sqrt{2}\cos((2m + 1)\pi(1 - \Phi(t))))\varphi(t) \\ &= (1 - \eta\sqrt{2}\cos((2m + 1)\pi\Phi(t)))\varphi(t) = p_{2m+1,-\eta}(t). \end{aligned}$$

The course of the densities $p_{l,\eta}$ is plotted for special l and η in the following pictures.

Null Hypothesis	Alternative Hypothesis	
	$p_{1,0.5}; p_{1,-0.5}$	dashed
$p_{l,0}$ solid $l = 1, 2, 3, 4$	$p_{2,0.5}; p_{2,-0.5}$	dotted
	$p_{3,0.5}; p_{3,-0.5}$	dotdash
	$p_{4,0.5}; p_{4,-0.5}$	longdash

Now we study the power of the Cramér–von Mises test under the sequence of alternatives $P_{l,n,h}$ in (43) that have the tangents $a_{l,h}$ in (42). It follows from (40) and (41) with $\dot{l}_0(t) = t$ and $a = a_{l,h}$ that for fixed $l = 1, 2, \ldots$ and $\theta = \mu = 0$

$$(\Pi_0^{\perp} a_{l,h})(t) = h\varphi_l(\Phi(t)) - ht(\int \varphi_l(\Phi(s))s\,\mathrm{d}\Phi(s)).$$

Set

$$\gamma_l = \int s\varphi_l(\Phi(s))\,\mathrm{d}\Phi(s) = \int_0^1 \Phi^{-1}(t)\varphi_l(t)\,\mathrm{d}t. \tag{44}$$

Then d_k in (40) turns into

$$d_{k,l} = h\int (\varphi_l(\Phi(t)) - \gamma_l t)\varphi_k(\Phi(t))\,\mathrm{d}\Phi(t).$$

As the $\varphi_l(\Phi)$ form an orthonormal system in $L_2(N(0,1))$ we get

$$d_{k,l} = h(\delta_{k,l} - \gamma_k\gamma_l), \tag{45}$$

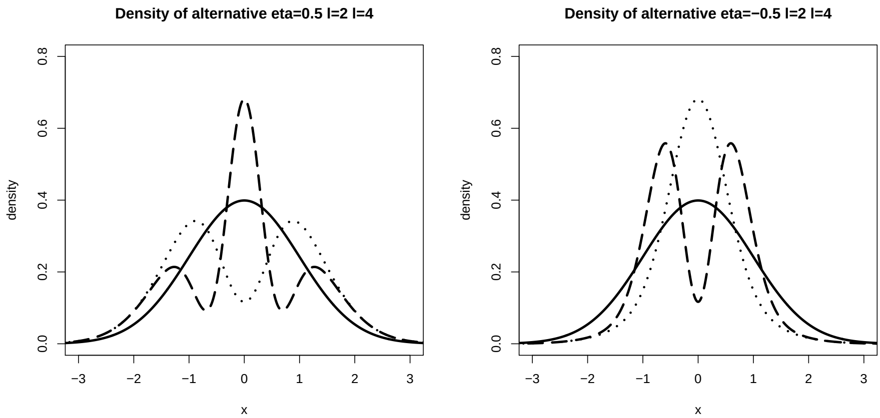


Fig. 1

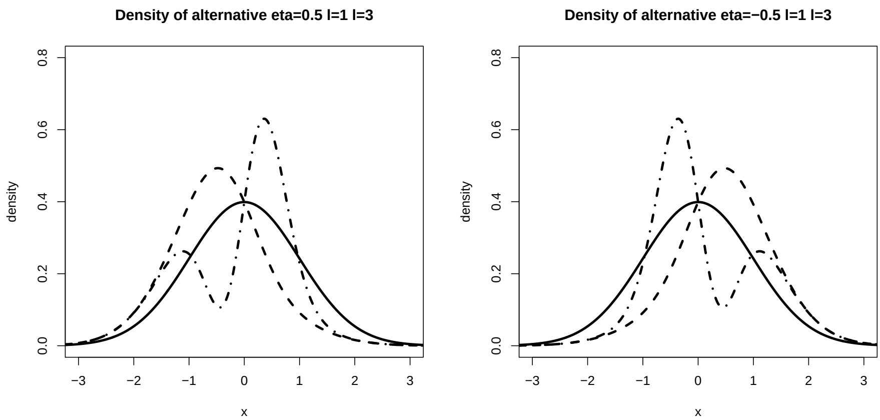


Fig. 2

where $\delta_{k,l}$ is the Kronecker symbol. The symmetry of the standard normal distribution implies $\Phi^{-1}(0.5+t) = -\Phi^{-1}(0.5-t)$ for $0 \leq t \leq 0.5$ and

$$\int_0^1 \Phi^{-1}(t) g(t)\, \mathrm{d}t = 0$$

for every function g which is symmetric with respect to 0.5. Hence by (42) and (44)

$$\gamma_{2m} = \int_0^1 \Phi^{-1}(t) \varphi_{2m}(\Phi(t))\, \mathrm{d}t = 0. \tag{46}$$

Using (45) and (46) we get the statistics $S_l = \sum_{k=1}^{\infty} (k\pi)^{-2}(Z_k + hd_{k,l})^2$ that appear in Theorem 4.5 if for $l = 1, \ldots, 4$ the local alternatives with the tangents in (42) are considered

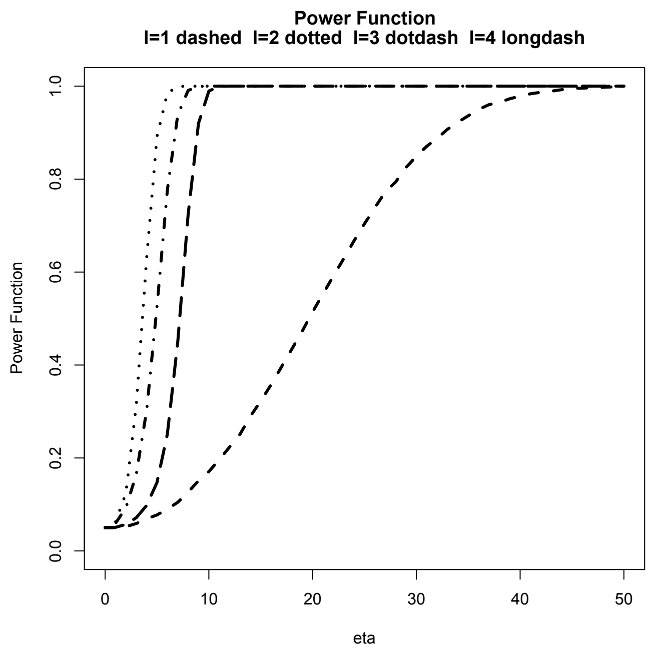
$$
\begin{aligned}
S_1 &= \frac{1}{\pi^2}(Z_1 + h(1-\gamma_1^2))^2 + \sum\nolimits_{m=1}^{\infty} \frac{1}{(2m\pi)^2} Z_{2m}^2 \\
&\quad + \sum\nolimits_{m=1}^{\infty} \frac{1}{((2m+1)\pi)^2} (Z_{2m+1} - h\gamma_1\gamma_{2m+1})^2, \\
S_2 &= \frac{1}{\pi^2} Z_1^2 + \frac{1}{4\pi^2}(Z_2 + h)^2 + \sum\nolimits_{k=3}^{\infty} \frac{1}{(k\pi)^2} Z_k^2, \\
S_3 &= \frac{1}{\pi^2}(Z_1 - h\gamma_1\gamma_3))^2 + \frac{1}{4\pi^2} Z_2^2 + \frac{1}{9\pi^2}(Z_3 + h(1-\gamma_3^2))^2 \\
&\quad + \sum\nolimits_{m=2}^{\infty} \frac{1}{(2m\pi)^2} Z_{2m}^2 + \sum\nolimits_{m=2}^{\infty} \frac{1}{((2m+1)\pi)^2} (Z_{2m+1} - h\gamma_3\gamma_{2m+1})^2, \\
S_4 &= \frac{1}{\pi^2} Z_1^2 + \frac{1}{4\pi^2} Z_2^2 + \frac{1}{9\pi^2} Z_3^2 + \frac{1}{16\pi^2}(Z_4 + h)^2 + \sum\nolimits_{k=5}^{\infty} \frac{1}{(k\pi)^2} Z_k^2.
\end{aligned}
$$

The asymptotic power $\pi_l(h)$ of φ_{RC_n} for observations with distribution $P_{l,n,h}$ from (43) is given by

$$
\pi_l(h) = P\left(S_l > c_{1-\alpha}\right), \tag{47}
$$

where $c_{1-\alpha}$ is the $1-\alpha$ quantil of the Cramér–von Mises distribution in (4).

The subsequent pictures show the power functions $\pi_1(h), \ldots, \pi_4(h)$.



To discuss the power functions that have been generated by simulation we remark at first that the γ_l in (44) are just the scalar product of the tangent $\dot{l}_0$ of the model and the tangent $\varphi_l(\Phi(t))$ of the sequence $P_{l,n,h}$ in (43). As the power function depends on the projection of $\varphi_l(\Phi(t))$ on the orthogonal complement of the tangent space of the model, we can expect a good power only if $|\gamma_l|$ is small. We have already seen that $\gamma_{2m} = 0$. The γ_l for odd l have been numerically evaluated. The first values are

$\gamma_1 = -0.9484$	$\gamma_3 = -0.2407$	$\gamma_5 = -0.1306$	$\gamma_7 = -0.0880$
$\gamma_9 = -0.0657$	$\gamma_{11} = -0.0522$	$\gamma_{13} = -0.0431$	$\gamma_{15} = -0.0366$

To explain the poor slope of the power functions $\pi_1(h)$ we inspect the first leading terms in the statistic S_1

$$\begin{aligned} & \frac{1}{\pi^2}(Z_1 + h(1-\gamma_1^2))^2 + \frac{1}{(3\pi)^2}(Z_3 - h\gamma_1\gamma_3)^2 \\ & + \frac{1}{(5\pi)^2}(Z_5 - h\gamma_1\gamma_5)^2 + \frac{1}{(7\pi)^2}(Z_7 - h\gamma_1\gamma_7)^2 + \dots \\ \approx \; & 0.1013(Z_1 + 0.01h)^2 + 0.0113(Z_3 + 0.2283h)^2 \\ & + 0.0041(Z_5 - 0.1239h)^2 + 0.0021(Z_7 - 0.083h)^2 + \dots \end{aligned}$$

We see that the right hand term only weakly depends on h and produces, therefore, a poor power only. The largest power function is $\pi_2(h)$. The big slope can be explained by inspecting the first terms in S_2 that are given by

$$0.1013Z_1^2 + 0.0253(Z_2 + h)^2 + 0.0113Z_3^2 + \dots$$

We recognize a stronger dependence of S_2 on h compared with S_1.

Summarizing we can say that in the Gaussian location model the Cramér–von Mises tests has poor power for alternatives plotted in the pictures in Figure 2 for $l = 1$ and high power for alternatives plotted in the pictures in Figure 1 for $l = 2$.

5. COMPUTER SIMULATIONS

5.1. Actual sizes of the randomized goodness of fit tests

Monte Carlo sampling experiments to check the accuracy of the approximation by the limit distribution of the Cramér–von Mises statistic and the Kolmogorov–Smirnov statistic, respectively, have been carried out by several authors, see e. g. [16] and [17]. We have checked the actual significance level of tests that are based on the randomized statistic in (21) by computer simulations.

1. Normal distribution with unknown μ and σ^2

We use the notations in Example 2.3 and assume $\mu = 0$ and $\sigma^2 = 1$ without loss of generality. The randomized Cramér–von Mises statistic in (21) is given by

$$\begin{aligned} \mathsf{RC}_n = \sum\nolimits_{i=1}^{n} \Big(& \frac{i}{n} - \Phi\left(\left(X_{n:i} - \overline{X}_n\right)/S_n\right) - \frac{1}{\sqrt{n}}\varphi\left(\left(X_{n:i} - \overline{X}_n\right)/S_n\right) V_{1,n} \\ & - \frac{(X_{n:i} - \overline{X}_n)}{\sqrt{2n}S_n}\varphi\left(\left(X_{n:i} - \overline{X}_n\right)/S_n\right) V_{2,n} \Big)^2 \end{aligned} \qquad (48)$$

2. Exponential distribution with unknown parameter λ

We use the notations in Examples 2.4 and assume $\lambda = 1$ without loss of generality. Then the randomized Cramér–von Mises statistic in (21) is given by

$$\begin{aligned} \mathsf{RC}_n &= \sum\nolimits_{i=1}^{n} \left(\frac{i}{n} - (1 - \exp\{-X_{n:i}/\overline{X}_n\}) \right. \\ &\qquad \left. - \frac{V_n}{\sqrt{n}} \frac{X_{n:i}}{\overline{X}_n} \exp\{-X_{n:i}/\overline{X}_n\} \right)^2 \end{aligned} \tag{49}$$

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 statistics in (48) or (49). To carry out the simulations we used the programm R. The implemented pseudo random generator is the Mersenne-Twister generator, see [12].

1. For $n = 20; 50; 100; 1000$ we generate $X_1, \ldots, X_n$ from F_θ.
2. Calculate the MLE $\hat{\theta}_n$.
3. Calculate the values of the statistics T_n.
4. Carry out the test $\varphi_n = \begin{cases} 1, & T_n > c_{n,1-\alpha} \\ 0, & \text{else} \end{cases}$,

   where the $c_{n,1-\alpha}$ are the $1-\alpha$ quantiles of the classical Cramér–von Mises statistic for sample size n. We have taken $c_{n,1-\alpha}$ from [3].
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. Subsequently α(CMR) denotes the actual level of the randomized Cramér–von Mises test.

α(CMR): Normal distribut., μ,σ^2 unknown				
α	$n = 20$	$n = 50$	$n = 100$	$n = 1000$
0.01	0.013	0.010	0.010	0.010
0.05	0.060	0.052	0.053	0.047
0.1	0.114	0.104	0.107	0.099
α(CMR): Exponential distribut., λ unknown				
0.01	0.012	0.012	0.012	0.009
0.05	0.055	0.051	0.053	0.049
0.1	0.108	0.104	0.105	0.099

From the above table we may conclude that under the null hypothesis the randomized goodness of fit test statistics even for small sample sizes behave very similar as the corresponding goodness of fit test statistics for a simple null hypothesis. This is demonstrated by the fact that the actual levels of the tests for $n = 20; 50; 100; 1000$ are very close to the predetermined α.

5.2. Asymptotic power under special alternatives

Let $X_1, \ldots, X_n$ be i.i.d. with a common normal distribution with unknown μ and σ^2 and $X_{n:1} \leq \ldots \leq X_{n:n}$ be the order statistic. V denotes the covariance matrix of $(X_{n:1}, \ldots, X_{n:n})$ and m is the vector of the expectations of the order statistic of independent standard normal random variables. Set

$$a^T = (a_1, \ldots, a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{1/2}}.$$

The Shapiro–Wilk statistic is defined by

$$\mathsf{W}_n = \frac{\sum_{i=1}^n a_i X_{n:i}}{\sum_{i=1}^n (X_i - \overline{X}_n)^2}.$$

According to [14], for $n > 50$ the statistic W_n can be approximated by

$$\begin{aligned} \mathsf{W}_n^* &= \frac{\sum_{i=1}^n b_i X_{n:i}}{\sum_{i=1}^n (X_i - \overline{X}_n)^2}, \\ b &= (b_1, \ldots, b_n) = \frac{1}{\sqrt{m^T m}} m. \end{aligned}$$

Let now X be a random variable with $EX^8 < \infty$ and introduce the skewness and curtosis by

$$\beta_1 = \frac{1}{\sigma^3} E(X - \mu)^3 \quad \text{and} \quad \beta_2 = \frac{1}{\sigma^4} E(X - \mu)^4 - 3,$$

where $\mu = EX$, $\sigma^2 = V(X)$. If X has a normal distribution then $\beta_1 = \beta_2 = 0$, so that the Bowman–Shenton statistic

$$\begin{aligned} \mathsf{BS}_n &= \frac{n}{6} (\widehat{\beta}_{1,n})^2 + \frac{n}{24} (\widehat{\beta}_{2,n})^2, \qquad (50) \\ \widehat{\beta}_{1,n} &= \frac{1}{S_n^3} \frac{1}{n} \sum\nolimits_{i=1}^n (X_i - \overline{X}_n)^3 \quad \widehat{\beta}_{2,n} = \frac{1}{S_n^4} \frac{1}{n} \sum\nolimits_{i=1}^n (X_i - \overline{X}_n)^4 - 3, \end{aligned}$$

indicates deviations in the skewness and curtosis and can be used for testing normality. Bowman and Shenton [2] proved under the null hypothesis (normality) that $\mathcal{L}(\mathsf{BS}_n) \Rightarrow \chi^2$-distribution with two degrees of freedom. The Bowman–Shenton test rejects the null hypothesis for large values of BS_n. D'Agostino and Stephens [4] and other authors noticed that the asymptotic is poor and proposed transformations of $\widehat{\beta}_{1,n}, \widehat{\beta}_{2,n}$. We do not use this approach. Instead we determine the quantiles of BS_n by simulations.

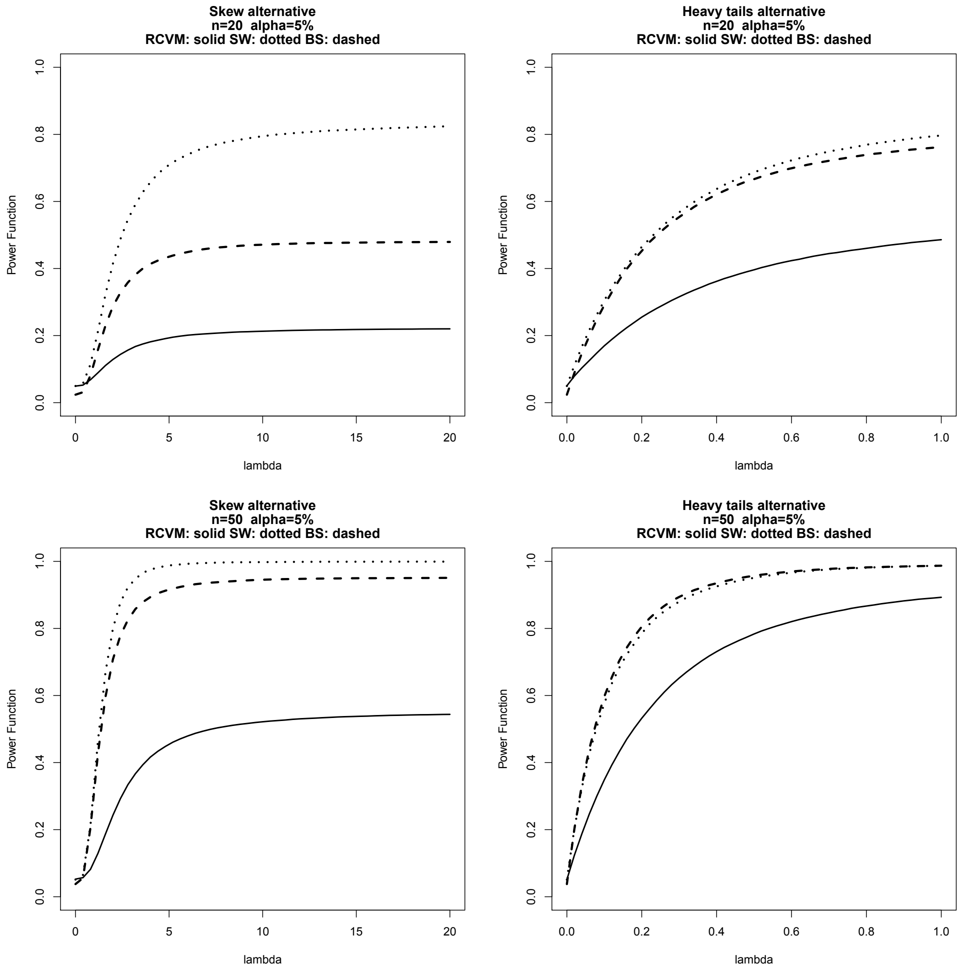
To compare the new randomized tests with the Shapiro–Wilk and the Bowman–Shenton test we consider two types of parameterized alternatives.

1. Deviation from normality in the direction of skewness

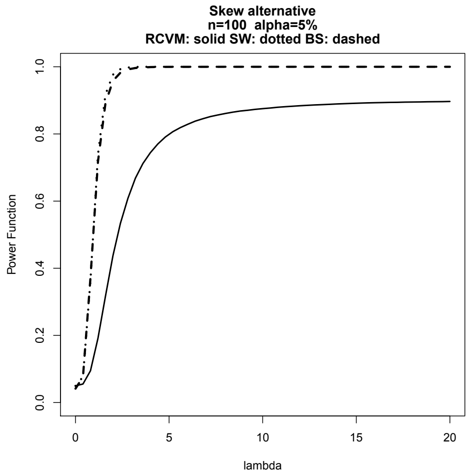
Let X has a standard normal distribution and suppose that Y has an exponential distribution with expectation 1. Put $P_\lambda = \mathcal{L}(X + \lambda Y), \quad \lambda > 0$. For $\lambda = 0$ the distribution P_0 is a standard normal one and λ controls the deviation from normality.

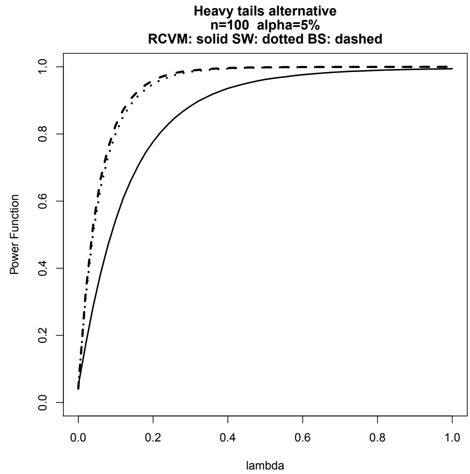
2. Deviation from normality in the direction of heavy tails

Let X has a standard normal distribution and suppose that C has a Cauchy distribution. Put $Q_\lambda = \mathcal{L}(X + \lambda C), \lambda > 0$. Again, for $\lambda = 0$ the distribution Q_0 is a standard normal one and λ controls the deviation from normality.



The results of the computer simulations are in a complete agreement with the power functions π_l in (47) and the course of the density functions in the pictures Figure 1 and Figure 2. Indeed the power function π_1 has a small slope. The density $p_{1,\eta}$ in Figure 2 is skew. Therefore it is not surprising that the randomized Cramér–von Mises test has poor power for skew alternatives. Similarly, as the power function π_2 has a big slope and the densities $p_{2,\eta}$ are broader than the normal density we may expect a good power for the randomized Cramér–von Mises test. This conjecture has been confirmed by the results of the computer simulations for alternatives with





heavy tails.

6. APPENDIX

In this appendix we collect known and prove some new technical results.

Lemma 6.1. If the conditions (25) and (26) are satisfied then $a \in L_2^0(P)$. Conversely, for every $a \in L_2^0(P)$ there is a sequence $\{P_n\}$ of distributions with $P_n \ll P$ and tangent a.

P r o o f. As g_n is a probability density we get

$$\begin{aligned} 1 &= \int g_n \,\mathrm{d}P = \int \left(1 + \frac{1}{\sqrt{n}} a_n + \frac{1}{4n} a_n^2\right) \mathrm{d}P \\ \left|\int a \,\mathrm{d}P\right| &\leq \frac{1}{4\sqrt{n}} \int a_n^2 \,\mathrm{d}P + \left(\int (a_n - a)^2 \,\mathrm{d}P\right)^{1/2}. \end{aligned}$$

The assumption (26) yields that $\int a_n^2 \,\mathrm{d}P$ is bounded. Hence $\int a \,\mathrm{d}P = 0$ by taking $n \to \infty$. To prove the second statement we assume $a \in L_2^0(P)$ and set for normalizing constants C_n

$$P_n(A) = C_n \int_A \left(1 + \frac{1}{2\sqrt{n}} a\right)^2 \mathrm{d}P.$$

It is not hard to see that $a_n = 2\sqrt{n}\big(\sqrt{\frac{\mathrm{d}P_n}{\mathrm{d}P}} - 1\big)$ satisfies (26). For details we refer to [9]. □

Lemma 6.2. Suppose $X_1, \ldots, X_n$ are the projection of $\mathcal{X}^n$ on $\mathcal{X}$ and assume $P_n \ll P$. If $g_n = \mathrm{d}P_n/\mathrm{d}P$ and $a_n = 2\sqrt{n}(\sqrt{g_n} - 1)$ satisfies (26) for some $a \in L_2^0(P)$ then

$$\ln L_n = \frac{1}{\sqrt{n}} \sum\nolimits_{i=1}^{n} a(X_i) - \frac{1}{2}\int_0^1 a^2(x)P(\mathrm{d}x) + o_{P^{\otimes n}}(1). \tag{51}$$

$$L_n = \frac{\mathrm{d}P_n^{\otimes n}}{\mathrm{d}P^{\otimes n}}(X_1, \ldots, X_n)$$

$$P_n^{\otimes n} \lhd P^{\otimes n}. \tag{52}$$

Proof. The statement (51) is a special case of Theorem 3 in [15], p. 154. The fact that (51) implies (52) is the content of the first lemma of LeCam, see Exercise 2, p. 157 in [15]. □

Lemma 6.3. The condition (52) implies

$$P_n^{\otimes n} \otimes \mathsf{N}(0, I) \lhd P^{\otimes n} \otimes \mathsf{N}(0, I).$$

Proof. As $P_n^{\otimes n} \otimes \mathsf{N}(0, I) \ll P^{\otimes n} \otimes \mathsf{N}(0, I)$ it follows from part B) in Theorem 6.26 in [11] that the statement is equivalent with the uniform integrability of

$$\frac{d(P_n^{\otimes n} \otimes \mathsf{N}(0, I))}{d(P^{\otimes n} \otimes \mathsf{N}(0, I))}(x, y) = \frac{\mathrm{d}P_n^{\otimes n}}{\mathrm{d}P^{\otimes n}}(x)$$

with respect to $P^{\otimes n} \otimes \mathsf{N}(0, I)$. But this is equivalent with uniform integrability of the right hand term with respect to $P^{\otimes n}$ which follows from (52) by applying Theorem 6.26 in [11] again. □

The next Lemma follows from Theorem 3 in [13], p. 92.

Lemma 6.4. Suppose $Z, Z_1, Z_2, \ldots$ are random elements of $\mathbb{D}[-\infty, \infty]$ (under its uniform metric and projection σ-algebra) defined on $(\Omega, \mathfrak{F}, Q)$ and assume that Z is continuous. Then $\mathcal{L}(Z_n) \Rightarrow \mathcal{L}(Z)$ if and only if the fidis of Z_n converge to the fidis of Z and for each $\varepsilon > 0$ and each $\delta > 0$ there exists a grid $-\infty = t_0 < t_1 < \cdots < t_N = \infty$ such that

$$\limsup_{n\to\infty} Q\left(\max_{0 \le i < N-1} \sup_{t_i \le t < t_{i+1}} |Z_n(t) - Z_n(t_i)| > \delta \right) < \varepsilon. \tag{53}$$

Lemma 6.5. Suppose $Y, Y_1, Y_2, \ldots$ are random elements of $\mathbb{D}[-\infty, \infty]$ and the fidis of Y_n converge to the fidis of Y where Y is continuous. If $\mathcal{L}(Y_n) \lhd \mathcal{L}(Z_n)$ and Z_n satisfies (53) then $\mathcal{L}(Y_n) \Rightarrow \mathcal{L}(Y)$.

Proof. By the Jurečková characterization of contiguity, see e.g. [15], Lemma 1, p. 157, for every $\eta > 0$ there is some $\varepsilon > 0$ such that (53) implies

$$\limsup_{n\to\infty} Q\left(\max_{0 \le i < N-1} \sup_{t_i \le t < t_{i+1}} |Y_n(t) - Y_n(t_i)| > \delta \right) < \eta$$

and the statement follows from the preceding Lemma. □

Lemma 6.6. Suppose $Z, Z_1, Z_2, \ldots$ are random elements of $\mathbb{D}[-\infty, \infty]$ defined on $(\Omega, \mathfrak{F}, Q)$ and assume that Z is continuous. Assume $X_1, X_2, \ldots$ are i.i.d. with c.d.f. F and empirical c.d.f. $\widehat{F}_n$. Then $\mathcal{L}(Z_n) \Rightarrow \mathcal{L}(Z)$ implies

$$\mathcal{L}\left(\int Z_n^2(t)\,\mathrm{d}\widehat{F}_n(t)\right) \Rightarrow \mathcal{L}\left(\int Z^2(t)\,\mathrm{d}F(t)\right).$$

Proof. The Glivenko–Cantelli Theorem yields

$$\lim_{n\to\infty} \sup_t |\widehat{F}_n(t) - F(t)| = 0 \quad \text{a.s.} \tag{54}$$

Set $W_n = Z_n^2, W = Z^2$. For every partition $\mathcal{P} = \{t_0, \ldots, t_N\}$, $-\infty = t_0 < t_1 < \cdots < t_N = \infty$ we denote by $W_{n,\mathcal{P}}$ and $W_{\mathcal{P}}$ the piecewise constant processes that have the values $W_n(t_i)$ and $W(t_i)$, respectively, in $[t_i, t_{i+1})$. The convergence of the fidis and (54) yield

$$\mathcal{L}\left(\int W_{n,\mathcal{P}}\,\mathrm{d}\widehat{F}_n\right) \Rightarrow \mathcal{L}\left(\int W_{\mathcal{P}}\,\mathrm{d}F\right). \tag{55}$$

The continuous mapping theorem and $\mathcal{L}(Z_n) \Rightarrow \mathcal{L}(Z)$ imply $\mathcal{L}(W_n) \Rightarrow \mathcal{L}(W)$. Consequently by Lemma 6.4, for every m there is a partitions $\mathcal{P}_m$ such that for every n

$$\left|\int W_n\,\mathrm{d}\widehat{F}_n - \int W_{n,\mathcal{P}_m}\,\mathrm{d}\widehat{F}_n\right| < \frac{1}{m} \quad \text{and} \quad \left|\int W\,\mathrm{d}F - \int W_{\mathcal{P}_m}\,\mathrm{d}F\right| < \frac{1}{m}.$$

The statement (55) implies for every Lipschitz continuous function φ with Lipschitz constant L

$$\limsup_{n\to\infty} \left|E\varphi\left(\int W_n\,\mathrm{d}\widehat{F}_n\right) - E\varphi\left(\int W\,\mathrm{d}F\right)\right| \le \frac{2L}{m}.$$

Taking $m \to \infty$ we get the statement. □

7. ACKNOWLEDGEMENT

The authors sincerely thank the Associated Editor and the referees for their careful reading of the paper and the suggestions that led to a substantial improvement of the manuscript. The work of the second author was supported by Deutsche Forschungsgemeinschaft DFG GRK 1505/1 Welisa.

(Received November 1, 2010)

REFERENCES

[1] P. Billingsley: Covergence of Probability Measures. Wiley, New York 1986.

[2] K. O. Bowman and L. R. Shenton: Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2. Biometrica *62* (1975), 243-250.

[3] S. Csörgő and J. J. Faraway: The exact and asymptotic distributions of Cramér–von Mises statistics. J. Roy. Statist. Soc. Ser. B (Methodological) *58* (1996), 1, 221–234.

[4] R. B. D'Agostino and M. A. Stephens: Goodness-of-fit Techniques. Marcel Decker, New York and Basel 1986.

[5] J. Durbin: Distribution theory for tests based on the sample distribution function. Regional Conference Series in Applied Mathematics *9*, SIAM, Philadelphia 1973.

[6] J. Durbin: Weak convergence of the sample distribution function when parameters are estimated. Ann. Statist. *1* (1973), 279–290.

[7] M. Genz and E. Haeusler: Empirical processes with estimated parameters under auxiliary information. J. Comput. Appl. Math. *186* (2006), 191–216.

[8] J. Haywood and E. Khmaladze: On distribution-free goodness-of-fit tetsting of exponentiality. Econometrics *143* (2008), 5–18.

[9] A. Janssen: Asymptotic relative efficiency of tests at the boundary of regular statistical modells. J. Statist. Plann. Inference *126* (2004), 461–477.

[10] E. Khmaladze: Martingale approach in the theory of goodness of fit tests. Theory Probab. Appl. *24* (1981), 2, 283–302.

[11] F. Liese and K. J. Miescke: Statistical Decision Theory, Estimation, Testing and Selection. Springer-Verlag, New York 2008.

[12] M. Matsumoto and T. Nishimura: Mersenne-Twister: A 623-dimensionally equidistributed uniform pseudo-random generator. ACM Trans. Model. Comp. Simul. *8* (1998), 19, 3–30.

[13] D. Pollard: Convergence of Stochastic Processes. Springer-Verlag, New York 1984.

[14] S. S. Shapiro and R. S. Francia: Approximate analysis of variance test for normality. J. Amer. Statist. Assoc. *67* (1972), 215–216.

[15] G. Shorack and J. A. Wellner: Empirical Processes with Applications to Statistics. Wiley, New York 1986.

[16] M. A. Stephens: EDF statistics for goodness of fit and some comparisons. J. Amer. Statist. Assoc. *69* (1974), 730–737.

[17] M. A. Stephens: Goodness of fit for the extreme value distribution. Biometrika *64* (1976), 583–588.

[18] W. Stute, W. G. Manteiga, and M. P. Quindimil: Bootstrap based goodness-of-fit-tests. Metrika *40* (1993), 243–256.

[19] A. W. van der Vaart and J. A. Wellner: Weak convergence and Empirical Processes. With Applications to Statstics. Second edition 2000. Springer, New York 1996.

[20] A. W. van der Vaart: Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge Univ. Press. 1998.

Friedrich Liese, University of Rostock, Department of Mathematics, D-18057 Rostock, Ulmenstrasse 69, HS 3. Germany.
e-mail: friedrich.liese@uni-rostock.de

Bing Liu, University of Rostock, Department of Mathematics, D-18057 Rostock, Ulmenstrasse 69, HS 3. Germany.
e-mail: bing.liu@uni-rostock.de