© Board of the Foundation of the Scandinavian Journal of Statistics 1995. Published by Blackwell Publishers Ltd., 108 Cowley Road, Oxford OX4 1JF and 238 Main Street, Cambridge, MA 02142, USA. Vol. 22: 305–322, 1995

Permutation and Bootstrap Kolmogorov–Smirnov Tests for the Equality of two Distributions

JENS THOMAS PRÆSTGAARD

University of Iowa

ABSTRACT. We consider a generalized version of the two-sample Kolmogorov–Smirnov statistic which is calculated as the supremum distance, taken over a general class of indexing functions, $\mathscr{F}_N$, possibly depending on the joint sample size N, between the empirical measures of the two samples. The resulting class of tests is very flexible, encompassing both previous suggestions by Bickel (1969) and Romano (1989), and tests which have so far not been viewed as Kolmogorov–Smirnov tests, e.g. the supremum-based generalized log-rank tests of Fleming *et al.* (1987). We investigate the permutation and bootstrap versions of these tests; i.e. the procedure by which the critical values are found from the permutation or bootstrap distribution of the statistics. If the indexing class of functions $\mathscr{F}_N$ "converges" to a limiting collection $\mathscr{F}_0$, then we find conditions under which these procedures yield consistent tests of H_0: $\|\mathbf{P}-\mathbf{Q}\|_{\mathscr{F}_0}=0$ against any alternative $\|\mathbf{P}-\mathbf{Q}\|_{\mathscr{F}_0}\neq 0$. We furthermore calculate the local asymptotic power of both tests and find it to be the same for bootstrap and permutation resampling.

Key words: Bootstrap, permutation test, Kolmogorov–Smirnov statistic, empirical processes, contiguity

1. Introduction

The purpose of the present article is to investigate a general class of tests for equality of two distributions while arguing that modern empirical process methods are useful in this framework. Given an i.i.d. sample $X_1, \ldots, X_m$ from an unknown distribution P on a probability space $(\mathbf{Z}, \mathscr{Z})$ and, independent thereof, $Y_1, \ldots, Y_n$ independent from Q on the same probability space, the problem is to test the hypothesis that P equals Q. If the data are real-valued, one such test is based on the Kolmogorov–Smirnov statistic which is the supremum distance between the cumulative distribution functions of the two samples. A generalization of this test to a setting where the data are not necessarily real valued is to use the statistic D_N, $N=m+n$, defined by

$$D_N \equiv \sup\left\{\left|\frac{1}{m}\sum_{j=1}^{m} f(X_j)-\frac{1}{n}\sum_{j=1}^{n} f(Y_j)\right| : f\in\mathscr{F}_N\right\}\equiv\|\mathbf{P}_m-\mathbf{Q}_n\|_{\mathscr{F}_N} \tag{1.1}$$

where the supremum is taken over a class of functions $\mathscr{F}_N$ which map $\mathbf{Z}$ to $\mathbf{R}$. To get more general and interesting results, we allow the collection $\mathscr{F}_N$ to vary with N, in which case we shall assume that all $\mathscr{F}_N$ are contained in some collection $\mathscr{F}$, and that $\mathscr{F}_N$ "converges" in a sense to be specified to some collection $\mathscr{F}_0\subset\mathscr{F}$. The most common case is, of course, that $\mathscr{F}_0=\mathscr{F}_N=\mathscr{F}$ for all N. For the case $\mathbf{Z}=\mathbf{R}$, and $\mathscr{F}$ being the collection of indicator functions $1_{]-\infty,x]}$, $x\in\mathbf{R}$, (1.1) is exactly the Kolmogorov–Smirnov statistic.

The null distribution of (1.1), in the real as well as more general case, is in practice unknown. The asymptotic null distribution is usually known, but intractable because it depends on the unknown parameter P = Q; and even if P were known, it would still be computationally difficult to evaluate it. Thus it is wise to look for other, data driven, ways of obtaining critical values for a test based on D_N. One possibility is to use its permutation or randomization distribution. This would entail pooling the data into

$(Z_1^\omega, \ldots, Z_N^\omega) = (X_1^\omega, \ldots, X_m^\omega, Y_1^\omega, \ldots, Y_n^\omega)$ (the superscript ω signifies that we now consider the observed data fixed) and considering the statistic

$$\tilde{D}_N^\omega \equiv \sup\left\{\left|\frac{1}{m}\sum_{j=1}^{m} f(Z^\omega_{R_N(j)}) - \frac{1}{n}\sum_{j=m+1}^{N} f(Z^\omega_{R_N(j)})\right| : f \in \mathscr{F}_N\right\} \tag{1.2}$$

where R_N is a random permutation of $\{1, \ldots, N\}$ which takes on every permutation with equal probability $1/N!$. If $\tilde{c}_N^\omega(\alpha)$ denotes the $1-\alpha$ quantile of the distribution of (1.2),

$$\tilde{c}_N^\omega(\alpha) \equiv \inf\{t > 0: \Pr(\tilde{D}_N^\omega > t \mid Z_1, \ldots, Z_N) < \alpha\} \tag{1.3}$$

(which depends on the data, hence the superscript ω), then a permutation test based on (1.1) has test function

$$\tilde{\phi}_N(D_N) \equiv \begin{cases} 1 & \text{if } D_N > \tilde{c}_N^\omega(\alpha) \\ 0 & \text{if } D_N \leqslant \tilde{c}_N^\omega(\alpha). \end{cases} \tag{1.4}$$

Another possibility is to use the bootstrap. Here the idea is to sample $\hat{Z}_{N_1}, \ldots, \hat{Z}_{NN}$ with replacement from $(Z_1^\omega, \ldots, Z_N^\omega)$ and compute the distribution of the bootstrapped Kolmogorov–Smirnov statistic

$$\hat{D}_N^\omega \equiv \sup\left\{\left|\frac{1}{m}\sum_{j=1}^{m} f(\hat{Z}_{N_j}) - \frac{1}{n}\sum_{j=m+1}^{N} f(\hat{Z}_{N_j})\right| : f \in \mathscr{F}_N\right\}. \tag{1.5}$$

By defining $\hat{c}_N^\omega(\alpha)$ as the data dependent $1-\alpha$ quantile of $\hat{D}_N^\omega$ in the same way as (1.3), then the test function for the bootstrap Kolmogorov–Smirnov test is

$$\hat{\phi}_N(D_N) \equiv \begin{cases} 1 & \text{if } D_N > \hat{c}_N^\omega(\alpha) \\ 0 & \text{if } D_N \leqslant \hat{c}_N^\omega(\alpha). \end{cases} \tag{1.6}$$

The strength of the bootstrap and permutation methods outlined above is that although an exact evaluation of $\tilde{c}_N^\omega(\alpha)$ and $\hat{c}_N^\omega(\alpha)$ would involve, respectively, $N!$ and N^N evaluations of f, they can be approximated arbitrarily close by Monte-Carlo: by taking M ordered samples without and with replacement, respectively, from $(Z_1^\omega, \ldots, Z_N^\omega)$ and recomputing M times $\tilde{D}_N^\omega$ and $\hat{D}_N^\omega$. Based on these M permutation and bootstrap samples empirical permutation and bootstrap quantiles can easily be computed; the law of large numbers then guarantees that these converge to the "true" permutation and bootstrap quantiles, respectively $\tilde{c}_N^\omega(\alpha)$ and $\hat{c}_N^\omega(\alpha)$, as $M \to \infty$. For a more detailed treatment of the connections between bootstrap and permutation tests see Efron and Tibshirani (1993), chapter 15.

In this paper we first investigate the tests (1.4) and (1.6) under the fixed alternative $\|\mathrm{P}-\mathrm{Q}\|_{\mathscr{F}_0} \neq 0$ and null hypothesis H_0: $\|\mathrm{P}-\mathrm{Q}\|_{\mathscr{F}_0} = 0$. (Recall that $\mathscr{F}_0$ is the "limit" of $\mathscr{F}_N$.) H_0 is usually a simple hypothesis ($\mathrm{P} = \mathrm{Q}$); however, in some interesting examples such as example 1 below, it may be composite. It turns out that if the collection $\mathscr{F}$ which contains $\mathscr{F}_N$ and $\mathscr{F}_0$ satisfies the central limit theorem for P and Q, and the ratio of the P-sample size to the Q-sample size stabilizes, then the power of both tests approaches 1 as $N \to \infty$. The tests are furthermore found to be asymptotically equivalent under a sequence of local alternatives $\{\mathrm{P}_N\}$ and $\{\mathrm{Q}_N\}$ that tend to P and Q with $\|\mathrm{P}-\mathrm{Q}\|_{\mathscr{F}_0} = 0$, and the common asymptotic rejection probability is found to be larger than the level α, which is yet another desirable property of the tests.

Permutation tests based on the Kolmogorov–Smirnov statistic have previously been investigated by Bickel (1969) for $\mathbf{Z} = \mathbf{R}^p$ and $\mathscr{F} = \{1_{]-\infty, t]}: t \in \mathbf{R}^p\}$. Romano (1989), ex. 4, p. 152, shows consistency of permutation and bootstrap tests for the equality of k distributions based on a test statistic similar to (1.1). In his case the supremum is taken over a class $\mathscr{F} = \{1_V: V \in \mathscr{V}\}$ where $\mathscr{V}$ is a Vapnik–Chervonenkis class of sets. The contribution of the present paper is to generalize this to arbitrary classes of indexing functions $\mathscr{F}_N \subset \mathscr{F}$, possibly varying with the sample size, and to point out that the key feature in order for the permutation and bootstrap test to work is that $\mathscr{F}$ be P-as well as Q-Donsker (to be defined later).

The test statistics in (1.2) and (1.5) are both made up of empirical measures randomly weighted by exchangeable multipliers (this observation is also what drives the proofs we present). In view of the discovery of Ledoux & Talagrand (1986) (that a central limit theorem for the multipliers and $\mathscr{F}$'s being a Donsker class does not imply that the weighted sum converges in law), it is not automatic that the results of Bickel (1969) and Romano (1989) should carry through.

The following is an example of a test which may naturally be viewed as a Kolmogorov–Smirnov test with a more general class of indexing functions dependent on the sample size.

Example 1. Fleming *et al.* (1987) considered a class of test statistics for the equality of two survival time distributions based on randomly right censored data. Letting, respectively, $(N_1(t), Y_1(t))$ and $(N_2(t), Y_2(t))$ denote the counting process of observed deaths and risk-set at t for populations 1 and 2, which have sample sizes m and n, the authors investigated statistics of the form $\sup_{t\geqslant 0} T(\hat{S}^\rho, t)$ where (except for a scale factor $((m+n)/mn)^{1/2}$)

$$T(\hat{S}^\rho, t) \equiv \int_0^t \hat{S}(t)^\rho \frac{Y_1(s)Y_2(s)}{Y_1(s)+Y_2(s)}\left(\frac{dN_1(s)}{Y_1(s)} - \frac{dN_2(s)}{Y_2(s)}\right). \tag{1.7}$$

Here $\hat{S}(t)$ denotes the Kaplan–Meier estimator of the pooled sample, and $0 < \rho < 1$. To see how this fits nicely into an indexing-by-functions framework, let P denote the underlying probability which generates data $X_i = (L_i^x, C_i^x, J_i^x)$, $i = 1, \ldots, m$ where we think of L_i^x and C_i^x as the ith death and censoring time in population 1, respectively, and $J_i^x \equiv 1$ is deterministic. Similarly, let Q be the probability generating $Y_i = (L_i^y, C_i^y, J_i^y)$, $i = 1, \ldots, n$, the underlying death and censoring times for the second population, where we set $J_i^y \equiv 2$. A key observation is that

$$m\mathbf{P}_m(1_{\{j=1\}}1_{]0,t]}(l)1_{\{l\leqslant c\}}w(l)) = \int_0^t w(l)\,dN_1(l)$$

and similarly for $n\mathbf{Q}_n$ and $N_2(\cdot)$. By letting

$$\begin{aligned} f_N(l,c,j;t) \equiv{}& 1_{\{j=1\}}1_{]0,t]}(l)1_{\{l\leqslant c\}}\frac{m}{m+n}\frac{Y_2(l)}{Y_1(l)+Y_2(l)}\hat{S}(l)^\rho \\ &+ 1_{\{j=2\}}1_{]0,t]}(l)1_{\{1\leqslant c\}}\frac{n}{m+n}\frac{Y_1(l)}{Y_1(l)+Y_2(l)}\hat{S}(l)^\rho \end{aligned} \tag{1.8}$$

we have by (1.7) that $T(\hat{S}^\rho, t) = (m+n)(\mathbf{P}_m f_N(\cdot\,;t) - \mathbf{Q}_n f_N(\cdot\,;t))$. If we define $\mathscr{F}_N$ as the (random) collection of all functions of the form (1.8) with t ranging over $]0, 1]$, it follows that

$$\sup_{0<t\leqslant 1} |T(\hat{S}(t)^\rho, t)| = (m+n)\|\mathbf{P}_m - \mathbf{Q}_n\|_{\mathscr{F}_N}. \tag{1.9}$$

Our results show that a consistent test may be based on the statistic (1.9) by getting critical values from its permutation or bootstrap distribution. For clarity we describe the permutation test in detail: pool the triplets (L_i^x, C_i^x, J_i^x), $i = 1, \ldots, m$ and (L_j^y, C_j^y, J_j^y), $j = 1, \ldots, n$. Divide these pooled data randomly into two groups with, respectively, m and n elements. From these groups, form two counting processes, $\tilde{N}_1$ and $\tilde{N}_2$, counting the observed deaths in the two newly formed groups. (Notice that in this way the censoring pattern is kept fixed;

if an observation was censored in the original sample, so is it in the permuted sample). Let $\hat{S}(l)$, $Y_1(l)$, and $Y_2(l)$ be as in (1.7), i.e. calculated from the original data. Then the permutation version of (1.9) uses critical values obtained from

$$\sup_{0<t<1}\left|\int_0^t \hat{S}(l)\left(\frac{Y_2(l)}{Y_1(l)+Y_2(l)}\,d\tilde{N}_1(l)-\frac{Y_1(l)}{Y_1(l)+Y_2(l)}\,d\tilde{N}_2(l)\right)\right|$$

fixing the original data so that the only randomness is from the permutation. By sampling instead with replacement, we would get the bootstrap test. □

In order to find the distribution of statistics like (1.1), (1.2) and (1.5), we use the modern theory of empirical processes indexed by functions. The remainder of this introduction contains some definitions from this field; for a full account see e.g. Giné & Zinn (1984, 1986). From an i.i.d. sample $X_1, \ldots, X_m$ from $(\mathbf{Z}, \mathscr{Z}, \mathrm{P})$ the empirical measure is defined as $\mathbf{P}_m \equiv m^{-1}\Sigma_{j=1}^m \delta_{X_j}$, (where δ_{X_j} denotes the one-point measure concentrated in X_j) and the empirical process is $\mathbf{X}_m \equiv m^{1/2}(\mathbf{P}_m - \mathrm{P})$. If the collection $\mathscr{F} \subset L_2(P)$ satisfies $\sup_{f\in\mathscr{F}} |f(x)| \equiv \|\delta_x\|_{\mathscr{F}} < \infty$ for all $x \in \mathbf{Z}$ and $\sup_{f\in\mathscr{F}} |Pf| = \|P\|_{\mathscr{F}} < \infty$, then the empirical measure and process take values in $l^\infty(\mathscr{F})$: the space of real-valued, bounded functions defined on $\mathscr{F}$.

A P-Brownian bridge process G_{P} is a 0-mean Gaussian process indexed by $\mathscr{F}$ with covariance function

$$\mathrm{cov}\,(G_{\mathrm{P}}(f), G_{\mathrm{P}}(g)) = \mathrm{P}f_g - \mathrm{P}f\,\mathrm{P}g, \quad f, g \in \mathscr{F}.$$

Let ρ_{P} be the pseudometric on $L_2(\mathrm{P})$ given by

$$\rho_{\mathrm{P}}^2(f, g) \equiv \mathrm{var}\,(f(X) - g(X)) = \mathrm{P}((f-g)^2) - \mathrm{P}(f-g)^2. \tag{1.10}$$

If there exists a version G_{P} of a P-Brownian bridge, indexed by $\mathscr{F}$, which has bounded and ρ_{P}-uniformly continuous sample paths, we say that $\mathscr{F}$ is P-preGaussian. We say that $\mathscr{F}$ is P-Donsker or, shorter, that $\mathscr{F} \in CLT(\mathrm{P})$, if $\mathscr{F}$ is P-preGaussian and

$$\sqrt{n}(\mathbf{P}_m - \mathrm{P}) \Rightarrow G_{\mathrm{P}}.$$

This convergence is convergence in distribution in $l^\infty(\mathscr{F})$ in the sense of Hoffmann-Jørgensen (1984): for a sequence Φ_n of not necessarily Borel measurable mappings from a background probability space $(\Sigma, \mathscr{S}, \mathrm{Pr})$ into $l^\infty(\mathscr{F})$ and a random element Φ_0 of $l^\infty(\mathscr{F})$ with a Radon, Borel law, convergence in distribution, denoted $\Phi_n \Rightarrow \Phi_0$, takes place iff $E(H(\Phi_n))^* \to EH(\Phi_0)$ for all bounded and continuous $H: l^\infty(\mathscr{F}) \to \mathbf{R}$. Here the notation is for a mapping $h: (\Sigma, \mathscr{S}, \mathrm{Pr}) \to \mathbf{R}$

$$h^* \equiv \inf\{h' : h' \text{ is } \mathscr{S}\text{-measurable, } h' \geqslant h, \mathrm{Pr} \text{ a.s.}\}.$$

As is standard in empirical process theory, we shall use the notation $h_n \xrightarrow{\mathrm{Pr}^*} 0$ to indicate that the measurable envelopes converge to 0 in probability, i.e. that $h_n^* \xrightarrow{\mathrm{Pr}} 0$. See Anderson (1985), Dudley (1985), or Van der Vaart & Wellner (1994) for more on Hoffmann-Jørgensen weak convergence.

It is important in the Hoffmann-Jørgensen weak convergence setting to specify the probability space on which the possibly non-measurable random elements are defined. Hence we take $(X_1, X_2, \ldots, Y_1, Y_2, \ldots)$ to be defined on the canonical probability space

$$(\mathbf{Z}, \mathscr{Z}, \mathrm{P})^\infty \times (\mathbf{Z}, \mathscr{Z}, \mathrm{Q})^\infty \equiv (\Omega, \mathscr{O}, \mathrm{Pr}).$$

The bootstrap and permutation randomness, and what we need in terms of auxiliary random variables (Rademachers, symmetrized Poissons etc.) are taken to be defined on ([0, 1], $\mathscr{B}$, Lebesgue). The notation X_j^{ω} will indicate that the data are considered fixed; a typical example would be $E\|\Sigma_{j=1}^{n} \varepsilon_j \delta_{X_j^{\omega}}\|$ where the expectation is over $\varepsilon_1, \ldots, \varepsilon_n$, while in $E^*\|\Sigma_{j=1}^{n} \varepsilon_j \delta_{X_j}\|$ the expectation is over $\varepsilon_1, \ldots, \varepsilon_n$ and $X_1, \ldots, X_n$ jointly.

We define the $L_2(\mathrm{P})$ pseudometric by $e_{\mathrm{P}}^2(f, g) \equiv \mathrm{P}(f-g)^2$; for any pseudometric ρ on $\mathscr{F}$ and $\delta > 0$ we further need to define the collections

$$\mathscr{F}(\rho, \delta) \equiv \{f, g \in \mathscr{F}: \rho(f, g) < \delta\}.$$

Finally, we assume that $\mathscr{F}$ possesses enough measurability for randomization with i.i.d. multipliers to be possible; such a set of conditions is $\mathscr{F} \in \mathrm{NLDM(P)}$, and $\mathscr{F}^2, \mathscr{F}'^2 \in \mathrm{NLSM(P)}$ in the terminology of Giné & Zinn (1984, 1986, 1990). Here $\mathscr{F}^2$ and $\mathscr{F}'^2$ denote the classes of squared functions and squared differences of functions from $\mathscr{F}$, respectively. When all of these conditions hold, we write $\mathscr{F} \in M(\mathrm{P})$. It is known that $\mathscr{F} \in M(\mathrm{P})$ if $\mathscr{F}$ is countable, or if the empirical processes $\mathbf{X}_n$ are stochastically separable, or if $\mathscr{F}$ is image admissible Suslin (see Giné & Zinn, 1990, pp. 853, 854).

2. Power under fixed alternatives

In order to examine the permutation and bootstrap Kolmogorov–Smirnov tests under fixed alternatives, we shall first study what may be termed the two-sample bootstrap and permutation empirical processes. Let $\mathbf{H}_N \equiv N^{-1} \Sigma_{j=1}^{N} \delta_{Z_j}$ denote the pooled empirical measure for the combined sample; then the permutation empirical measures and process are, respectively,

$$\tilde{\mathbf{P}}_{m,N}^{\omega} \equiv \frac{1}{m} \sum_{j=1}^{m} \delta_{Z_{R_N(j)}^{\omega}},$$

and

$$\tilde{\mathbf{X}}_{m,N}^{\omega} \equiv \sqrt{m}(\tilde{\mathbf{P}}_{m,N}^{\omega} - \mathbf{H}_N^{\omega})$$

where R_N is a random, uniformly distributed permutation of $\{1, \ldots, N\}$.

The following theorem investigates the almost sure limiting behaviour of the empirical permutation process.

Theorem 1

Assume that $\mathscr{F} \in M(\mathrm{P}) \cap M(\mathrm{Q})$ *satisfies* $\mathscr{F} \in CLT(\mathrm{P}) \cap CLT(\mathrm{Q})$, *and has* P *and* Q *square integrable envelope F. Assume further that* $m/N \to \lambda$, $0 < \lambda < 1$, $m, N \to \infty$. *Then*

$$\tilde{\mathbf{X}}_{m,N}^{\omega} \Rightarrow \sqrt{1-\lambda} \cdot G_{\mathrm{H}}, \text{ Pr a.s.}$$

where $\mathbf{H} \equiv \lambda \mathrm{P} + (1-\lambda)\mathrm{Q}$.

The proof is contained in section 4.

The square integrability condition on the envelope function F is crucial for getting almost sure convergence in distribution of the empirical permutation process. This is an interesting result in its own right; for the purpose of investigating the permutation test (1.2) it is, however, slightly stronger than necessary. It suffices to know that $\tilde{\mathbf{X}}_{m,N}^{\omega}$ converges weakly "in probability" which may be explained as follows: the dual bounded Lipschitz metric d_{BL^*} metrizes weak convergence in $l^{\infty}(\mathscr{F})$, see e.g. Van der Vaart and Wellner (1994). Hence the conclusion of theorem 1 may be rephrased as

$$d_{\mathrm{BL}^*}(\tilde{\mathbf{X}}_{m,N}^{\omega}, \sqrt{1-\lambda} \cdot G_{\mathrm{H}}) \to 0, \text{ Pr a.s.,} \tag{2.11}$$

and it is natural to define convergence in distribution "in probability" as (2.11) taking place in Pr* probability, i.e. $d_{\mathrm{BL}*}(\tilde{\mathbf{X}}^{\omega}_{m,N}, \sqrt{1-\lambda}\cdot G_{\mathrm{H}}) \xrightarrow{\mathrm{Pr}^*} 0$. This definition is adapted from Giné & Zinn (1990) who define in this way a bootstrap central limit theorem "in probability". With this definition we get the following result; as before the proof is postponed to section 4.

Theorem 2
Assume that $\mathscr{F} \in M(\mathrm{P}) \cap M(\mathrm{Q})$ satisfies $\mathscr{F} \in CLT(\mathrm{P}) \cap CLT(\mathrm{Q})$, and $\|\mathrm{P}\|_{\mathscr{F}} < \infty$, $\|\mathrm{Q}\|_{\mathscr{F}} < \infty$. Assume further that $m/N \to \lambda$, $0 < \lambda < 1$, as $m, N \to \infty$. Then

$$\tilde{\mathbf{X}}^{\omega}_{m,N} \Rightarrow \sqrt{1-\lambda}\cdot G_H, \quad \textit{in } \mathrm{Pr}^* \textit{ probability.}$$

To see the usefulness of the empirical permutation process, define now $\tilde{\mathrm{Q}}^{\omega}_{n,N} \equiv n^{-1}\sum_{j=m+1}^{N} \delta_{Z^{\omega}_{R_N(j)}}$ where the random permutation R_n is the same as that used in defining $\tilde{\mathrm{P}}^{\omega}_{m,N}$. Let $\mathscr{F}_N \subset \mathscr{F}$ be as in the introduction. By noting that $\mathbf{H}^{\omega}_N = (m/N)\tilde{\mathrm{P}}^{\omega}_{m,N} + (n/N)\tilde{\mathrm{Q}}^{\omega}_{n,N}$, we may write the permutation Kolmogorov–Smirnov statistic (1.2) as

$$\tilde{D}^{\omega}_N = \|\tilde{\mathrm{P}}^{\omega}_{m,N} - \tilde{\mathrm{Q}}^{\omega}_{n,N}\|_{\mathscr{F}_N} = \frac{N}{n}\|\tilde{\mathrm{P}}^{\omega}_{m,N} - \mathbf{H}^{\omega}_N\|_{\mathscr{F}_N} = \frac{N}{n\sqrt{m}}\|\tilde{\mathbf{X}}^{\omega}_{m,N}\|_{\mathscr{F}_N},$$

that is

$$\sqrt{\frac{N}{n}}\|\tilde{\mathbf{X}}^{\omega}_{m,N}\|_{\mathscr{F}_N} = \sqrt{\frac{mn}{N}}\tilde{D}^{\omega}_N. \tag{2.12}$$

The statistic itself, defined in (1.1), may be written as

$$D_N = \|\mathbf{P}_m - \mathbf{Q}_n\|_{\mathscr{F}_N} = \|(\mathbf{P}_m - \mathrm{P}) - (\mathbf{Q}_n - \mathrm{Q}) + (\mathrm{P} - \mathrm{Q})\|_{\mathscr{F}_N}. \tag{2.13}$$

Assume now that $\mathscr{F}_0$ is the set of all limit points of sequences from $\mathscr{F}_N$, and that this convergence takes place uniformly, in the sense that

$$\lim_{N\to\infty}\sup_{f_0\in\mathscr{F}_0}\inf_{f\in\mathscr{F}_N}\rho_{\mathrm{H}}(f, f_0) = 0. \tag{2.14}$$

and

$$\lim_{N\to\infty}\sup_{f\in\mathscr{F}_N}\inf_{f_0\in\mathscr{F}_0}\rho_{\mathrm{H}}(f, f_0) = 0. \tag{2.15}$$

(The pseudometric ρ_{H} is defined in (1.10); note that if $\mathscr{F}_N \equiv \mathscr{F} \equiv \mathscr{F}_0$, then (2.14) and (2.15) are trivially satisfied.) Then any random elements $\mathbf{Z}_N$ which converge in distribution to G_{H} in $l^{\infty}(\mathscr{F})$ satisfy, by virtue of the $\mathbf{Z}_N$s being asymptotically equicontinuous, see e.g. Pollard (1990, th. 10.2 ii),

$$\|\mathbf{Z}_N\|_{\mathscr{F}_N} \Rightarrow \|G_{\mathrm{H}}\|_{\mathscr{F}_0}. \tag{2.16}$$

We hence get the following corollary to theorems 1 and 2. Its contents are that the permutation test based on the Kolmogorov–Smirnov statistic (1.1) is consistent against all alternatives $\mathrm{P} \neq \mathrm{Q}$ that can be distinguished over the class $\mathscr{F}_0$.

Corollary 1
Let the assumptions be as in theorem 2, and assume that $\mathscr{F}_N$ and $\mathscr{F}_0$ are related by (2.14) and (2.15). Then if further $\mathrm{P} \neq \mathrm{Q}$ and $\mathscr{F}_0$ is sufficiently rich that $\|\mathrm{P} - \mathrm{Q}\|_{\mathscr{F}_0} > 0$, the permutation Kolmogorov–Smirnov test with test function (1.4) is consistent, i.e. $\tilde{\phi}_N(D_N) \xrightarrow{\mathrm{Pr}^*} 1$.

Proof of corollary 1. By the definition (1.4) we see that the test function $\tilde{\phi}_N(D_N)$ may also be written as

$$\tilde{\phi}_N(D_N) = 1_{\{\sqrt{(mn)/N}D_N > \sqrt{(mn)/N}\tilde{c}_N^{\omega}(\alpha)\}}. \tag{2.17}$$

By (2.12) the $1-\alpha$ quantile of $(N/n)^{1/2}\|\tilde{\mathbf{X}}^{\omega}_{m,N}\|_{\mathscr{F}_N}$ is $((mn)/N)^{1/2}\tilde{c}_N^{\omega}(\alpha)$. By theorem 2 (and $N/n \to (1-\lambda)^{-1}$) any sequence $\{N'\}$ has a further sub-sequence $\{N''\}$ along which $(N''/n'')^{1/2}\tilde{\mathbf{X}}^{\omega}_{m'',N''} \Rightarrow G_{\mathrm{H}}$, Pr a.s. Hence, it follows from theorem 2 and (2.16) that $(N''/n'')^{1/2}$ $\|\tilde{\mathbf{X}}^{\omega}_{m'',N''}\|_{\mathscr{F}_N} \Rightarrow \|G_{\mathrm{H}}\|_{\mathscr{F}_0}$. Since G_{H} has a.s. uniformly continuous sample paths in the ρ_{H} pseudometric, it follows that $\|G_{\mathrm{H}}\|_{\mathscr{F}_0}$ has a continuous distribution except for a possible atom at 0; see e.g. Ledoux & Talagrand (1991, p. 60, 61), and references to Tsirelson (1975) given therein. Hence, by the continuous mapping theorem the $1-\alpha$ quantile converges along the same sub-sequence, i.e. $((m''n'')/N'')^{1/2}\tilde{c}^{\omega}_{N''}(\alpha) \xrightarrow{\text{a.s.}} c_{\mathrm{H}}(\alpha)$ where $c_{\mathrm{H}}(\alpha)$ denotes the $1-\alpha$ quantile of $\|G_{\mathrm{H}}\|_{\mathscr{F}_0}$. In other words, $((mn)/N)^{1/2}\tilde{c}_N^{\omega}(\alpha) \xrightarrow{\text{Pr*}} c_{\mathrm{H}}(\alpha)$. By (2.13) it follows that $(mn/N)^{1/2}D_N \xrightarrow{\text{Pr*}} \infty$; hence also jointly

$$\sqrt{\frac{mn}{N}}(D_N, \tilde{c}_N^{\omega}(\alpha)) \xrightarrow{\text{Pr*}} (\infty, c_{\mathrm{H}}(\alpha)). \tag{2.18}$$

Finally $\tilde{\phi}_N(D_N) \xrightarrow{\text{Pr*}} 1$ follows by (2.17), (2.18) and the continuous mapping theorem. □

Remark 1. If in fact the hypothesis H_0: $\|\mathrm{P}-\mathrm{Q}\|_{\mathscr{F}_0}=0$ is true, it follows from (2.13) and (2.16) that

$$\begin{aligned}\sqrt{\frac{mn}{N}}D_N &= \left\|\sqrt{\frac{n}{N}}\mathbf{X}_{m,N} - \sqrt{\frac{m}{N}}\mathbf{Y}_{n,N} - \sqrt{\frac{mn}{N}}(\mathrm{P}-\mathrm{Q})\right\|_{\mathscr{F}_N} \\ &\Rightarrow \|\sqrt{1-\lambda}G_{\mathrm{P}} - \sqrt{\lambda}G_{\mathrm{Q}}\|_{\mathscr{F}_0} \stackrel{\mathscr{L}}{\equiv} \|G_{\mathrm{H}}\|_{\mathscr{F}_0},\end{aligned} \tag{2.19}$$

where the last equality follows because G_{P}, G_{Q}, and G_{H} all have the same distribution when regarded as a random element indexed by $\mathscr{F}_0$. Now (2.18) becomes

$$\sqrt{\frac{mn}{N}}(D_N, \tilde{c}_N^{\omega}(\alpha)) \xrightarrow{\text{Pr*}} (\|G_{\mathrm{H}}\|_{\mathscr{F}_0}, c_{\mathrm{H}}(\alpha)),$$

and the permutation Kolmogorov–Smirnov test attains asymptotically the correct level α. (The level for finite samples may, due to the discreteness of $\tilde{D}_N^{\omega}$, be slightly different from α.) □

Turning to the bootstrap Kolmogorov–Smirnov test given by (1.6), we now introduce the following bootstrap empirical measure

$$\hat{\mathrm{P}}^{\omega}_{m,N} \equiv m^{-1}\sum_{j=1}^{m}\delta_{\hat{Z}_{Nj}},$$

and process

$$\hat{\mathbf{X}}^{\omega}_{m,N} \equiv m^{1/2}(\hat{\mathrm{P}}^{\omega}_{m,N} - \mathbf{H}^{\omega}_N).$$

Letting $\underline{M}_N$ be distributed as $\underline{M}_N \sim \mathrm{Mult}_N(m, (N^{-1}, \ldots, N^{-1}))$, we may also express the bootstrap empirical measure and process as, respectively,

$$\hat{\mathrm{P}}^{\omega}_{m,N} = m^{-1}\sum_{j=1}^{N} M_{N_j}\delta_{Z_{j^{\omega}}}$$

and

$$\hat{\mathbf{X}}^{\omega}_{m,N} = \frac{1}{\sqrt{m}}\sum_{j=1}^{N}\left(M_{N_j} - \frac{m}{n}\right)\delta_{Z_{j^{\omega}}} = \sqrt{\frac{N}{m}}\frac{1}{\sqrt{N}}\sum_{j=1}^{N}\left(M_{N_j} - \frac{m}{N}\right)\delta_{Z_{j^{\omega}}}. \tag{2.20}$$

The following almost sure limit theorem for the bootstrap empirical process is analogous to theorem 1 for sampling with replacement instead of without.

Theorem 3
Assume $\mathscr{F} \in M(\mathrm{P}) \cap M(\mathrm{Q})$ *satisfies* $\mathscr{F} \in \mathrm{CLT}(\mathrm{P}) \cap \mathrm{CLT}(\mathrm{Q})$, *and has* P *and* Q *square integrable envelope. Assume further that* $m/N \to \lambda$, $0 < \lambda < 1$, *as* $m, N \to \infty$. *Then*

$$\hat{X}^{\omega}_{m,N} \Rightarrow G_{\mathrm{H}}, \text{ Pr a.s.}$$

where $\mathrm{H} \equiv \lambda \mathrm{P} + (1-\lambda)\mathrm{Q}$.

As in theorem 2, we obtain an "in probability" result by relaxing the integrability assumptions on F.

Theorem 4
Assume that $\mathscr{F} \in M(\mathrm{P}) \cap M(\mathrm{Q})$ *satisfies* $\mathscr{F} \in CLT(\mathrm{P}) \cap CLT(\mathrm{Q})$, $\|\mathrm{P}\|_{\mathscr{F}} < \infty$, $\|\mathrm{Q}\|_{\mathscr{F}} < \infty$, *and that* $m/N \to \lambda$, $0 < \lambda < 1$, *as* $m, N \to \infty$. *Then*

$$\hat{\mathbf{X}}^{\omega}_{m,N} \Rightarrow G_{\mathrm{H}}, \textit{ in } \mathrm{Pr}^* \textit{ probability}.$$

To use the theorem, define $\hat{\mathrm{Q}}^{\omega}_{n,N} \equiv n^{-1} \Sigma_{j=1}^{n} \delta_{\hat{Z}_{N(m+j)}}$, and $\hat{\mathbf{Y}}^{\omega}_{n,N} = \sqrt{n}(\hat{\mathrm{Q}}^{\omega}_{n,N} - \mathrm{Q})$ where $\hat{Z}_{N(m+1)}, \ldots, \hat{Z}_{NN}$ are i.i.d. $\mathbf{H}^{\omega}_{N}$ and independent of $(\hat{\mathbf{Z}}_{N_1}, \ldots, \hat{\mathbf{Z}}_{Nm})$. Then we can write the bootstrapped Kolmogorov–Smirnov statistic (1.5) as

$$\hat{D}^{\omega}_{N} = \|\hat{\mathbf{P}}^{\omega}_{m,N} - \hat{\mathrm{Q}}^{\omega}_{n,N}\|_{\mathscr{F}_N} = \|(\hat{\mathbf{P}}^{\omega}_{m,N} - \mathbf{H}_N) - (\hat{\mathrm{Q}}^{\omega}_{n,N} - \mathbf{H}_N)\|_{\mathscr{F}_N}, \tag{2.21}$$

and we get the following corollary to theorem 3 and 4. It shows that if $\mathscr{F}$ is both P and Q-Donsker, then the bootstrap Kolmogorov–Smirnov test is consistent for testing $\|\mathrm{P}-\mathrm{Q}\|_{\mathscr{F}_0} > 0$ against $\|\mathrm{P}-\mathrm{Q}\|_{\mathscr{F}_0} = 0$ (if we have a little bit of additional assumptions on $\mathscr{F}$ for P and Q).

Corollary 2
Let the assumptions be as in theorem 4 *and assume further that* $\|\mathrm{P}-\mathrm{Q}\|_{\mathscr{F}_0} > 0$. *Assume that* $\mathscr{F}_N$ *and* $\mathscr{F}_0$ *satisfy* (2.14) *and* (2.15). *Then the test given by* (1.6) *is consistent, i.e.* $\hat{\phi}_N \to 1$, *in* Pr^* *probability*.

Proof of corollary 2. By theorem 4, (2.16), and (2.21) it follows that

$$\begin{aligned}\sqrt{\frac{mn}{N}}\,\hat{D}^{\omega}_{N} &= \left\|\sqrt{\frac{n}{N}}\,\hat{\mathbf{X}}^{\omega}_{m,N} - \sqrt{\frac{m}{N}}\,\hat{\mathbf{Y}}^{\omega}_{n,N}\right\|_{\mathscr{F}_N} \\ &\Rightarrow \|\sqrt{1-\lambda}G_{\mathrm{H}} - \sqrt{\lambda}G'_{\mathrm{H}}\|_{\mathscr{F}_0} \text{ in } \mathrm{Pr}^* \text{ probability} \\ &\overset{\mathscr{L}}{=} \|G_{\mathrm{H}}\|_{\mathscr{F}_0}\end{aligned} \tag{2.22}$$

where G'_{H} is an independent copy of G_{H}. Hence it holds as in corollary 1 that $(mn/N)^{1/2}\hat{c}_N(\alpha) \xrightarrow{\mathrm{Pr}^*} c_{\mathrm{H}}(\alpha)$, and since, as also seen in corollary 1, $(nm/N)^{1/2}D_N \xrightarrow{\mathrm{Pr}^*} \infty$, the conclusion follows. □

Remark 2. If $\|\mathrm{P}-\mathrm{Q}\|_{\mathscr{F}_0} = 0$, we have that (2.19) holds. It follows from (2.22) that in this case the bootstrap Kolmogorov–Smirnov test attains asymptotically the correct level α (which it might not for finite values due to discreteness). □

Example 1. (continued). Define $S_x(t) \equiv \Pr(L_1^x > t)$, $U_x(t) \equiv \Pr(C_1^x > t)$, and similarly $S_y(t)$ and $U_y(t)$, so that $\mathrm{P} \equiv (1 - S_x) \times (1 - U_x) \times \delta_1$ where δ_1 denotes the Dirac measure with mass at 1. Define similar quantities for the second population, so that $\mathrm{Q} \equiv (1 - S_y) \times (1 - U_y) \times \delta_2$. Define $\pi_x(t) \equiv S_x(t)U_x(t)$, $\pi_y(t) \equiv S_y(t)U_y(t)$, and the hazard measured $d\Lambda_x(t) \equiv -dS_x(t)/S_x(t-)$, $d\Lambda_y(t) \equiv -dS_y(t)/S_y(t-)$. Note that $U_x(l)\,d[-S_x(l)] = \pi_x(l)\,d\Lambda_x(l)$. Under the assumption H_0: $\Lambda_x(t) = \Lambda_y(t)$ it then holds for all functions $w(l)$ that

$$\begin{aligned}\int w(l)\pi_y(l)1_{\{l \leqslant c\}}\,d\mathrm{P}(l, c, 1) &= \int w(l)\pi_y(l)U_x(l)\,d[-S_x(l)]\\ &= \int w(l)\pi_y(l)\pi_x(l)\,d\Lambda_x(l)\\ &= \int w(l)\pi_x(l)\pi_y(l)\,d\Lambda_y(l)\\ &= \int w(l)\pi_x(l)U_y(l)\,d[-S_y(l)]\\ &= \int w(l)\pi_x(l)1_{\{l \leqslant c\}}\,d\mathrm{Q}(l, c, 2).\end{aligned}$$

Under H_0 we have $S_x(l) = S_y(l)$, all l, and if we choose

$$w(l) = \frac{\lambda(1-\lambda)S_x(l)^\rho}{\lambda\pi_x(l) + (1-\lambda)\pi_y(l)}\,1_{]0,t]}(l), \quad 0 \leqslant t \leqslant 1,$$

we see that in particular the class of functions $\mathscr{F}_0$ indexed by $0 < t < 1$ with members

$$\begin{aligned}f_0(l, c, j; t) \equiv{}& \lambda(1-\lambda)S_x(t)^\rho \frac{\pi_y(l)}{\lambda\pi_x(l) + (1-\lambda)\pi_y(l)}\,1_{]0,t]}(l)1_{\{l \leqslant c\}}1_{\{j=1\}}\\ &+ \lambda(1-\lambda)S_x(t)^\rho \frac{\pi_x(l)}{\lambda\pi_x(l) + (1-\lambda)\pi_y(l)}\,1_{]0,t]}(l)1_{\{1 \leqslant c\}}1_{\{j=2\}}\end{aligned}$$

satisfies $\|\mathrm{P} - \mathrm{Q}\|_{\mathscr{F}_0} = 0$. Note that this does not imply $\mathrm{P} = \mathrm{Q}$; the test is only for equality of the lifetime distributions while the censoring distributions may be different.

Consider now the statistic $\sup_{0 < t \leqslant 1} |(\hat{S}(t)^\rho, t)|$ where we take the absolute value of the supremum and restrict this to a finite interval, which we take without loss of generality take to be $]0, 1]$. Letting $\mathscr{F}_N$ denote the class of functions from (1.8), we showed in the introduction that

$$\sup_{0 < t \leqslant 1} |T(\hat{S}(t)^\rho, t)| = (m + n)\|\mathbf{P}_n - \mathbf{Q}_n\|_{\mathscr{F}_N}.$$

In order to invoke corollary 1 and 2, we now need to find a Donsker class $\mathscr{F}$ for P and Q which contains both $\mathscr{F}_N$ and $\mathscr{F}_0$. Note for this purpose that we may write the functions in $\mathscr{F}_N$ as

$$\begin{aligned}f_N(l, c, j; t) \equiv{}& 1_{\{j=1\}}1_{]0,t]}(l)1_{\{1 \leqslant c\}} \frac{nm}{(m+n)^2} \frac{Y_2(l)/n}{(Y_1(l) + Y_2(l))/(m+n)}\,\hat{S}(l)^\rho\\ &+ 1_{\{j=2\}}1_{]0,t]}(l)1_{\{l \leqslant c\}} \frac{mn}{(m+n)^2} \frac{Y_1(l)/m}{(Y_1(l) + Y_2(l))/(m+n)}\,\hat{S}(l)^\rho.\end{aligned}$$

Assume, as is standard in survival analysis that $\pi_x(1) > 0$, $\pi_y(1) > 0$. Note that $\hat{S}(l)^\rho$ is a non-increasing function bounded by 0 and 1. By the Glivenco–Cantelli theorem

$((Y_1(l) + Y_2(l))/(m+n))^{-1} \to (\lambda\pi_x(l) + (1-\lambda)\pi_y(l))^{-1}$, uniformly, a.s., and hence $((Y_1(l) + Y_2(l))/(m+n))^{-1}$ is eventually a non-decreasing, bounded function on [0, 1] by some constant M. Since $Y_1(l)/m$ and $Y_2(l)/n$ are both non-increasing bounded functions, we have that the functions

$$l \to \hat{S}(l)^\rho \frac{Y_1/m}{(Y_1(l) + Y_2(l))/(m+n)}, \quad l \to \hat{S}(l)^\rho \frac{Y_2/n}{(Y_1(l) + Y_2(l))/(m+n)}$$

will eventually be members of the collection $\mathscr{H}$ of real functions on [0, 1] that can be written as a product $h = fg$ for any non-decreasing $|f| < M$ and non-increasing $|g| < M$. It is known, e.g. from Pollard (1990, lem. 5.3), or Van der Vaart & Wellner (1993, lem. 2.6.14), that $\mathscr{H}$ is a universal Donsker class (i.e. a Donsker class for every probability on [0, 1]); hence also all functions

$$\mathscr{F} \equiv \{(l, c, j) \to 1_{\{j=1\}} 1_{\{1 \leq c\}} h_1(l) + 1_{\{j=2\}} 1_{\{l \leq c\}} h_2(l) \colon h_1, h_2 \in \mathscr{H}\}$$

form a universal Donsker class on $\mathbf{Z} = [0, 1]^2 \times \{1, 2\}$ which will eventually contain both $\mathscr{F}_N$ and $\mathscr{F}_0$.

In order for corollary 1 and 2 to hold for this example, we must finally argue that (2.14) and (2.15) are satisfied. By a strong law of large numbers for the Kaplan–Meier estimator we have that $\hat{S}(l)^\rho \to S_x(l)^\rho$ uniformly, a.s., which together with the Glivenco–Cantelli theorem for $Y_1(l)/m$, $Y_2(l)/n$, and $Y_1(l) + Y_2(l)/(m+n)$ shows that

$$\sup_{0<t<1} \sup_{(l,c,j) \in [0,1]^2 \times \{1,2\}} |f_N(l, c, j; t) - f_0(l, c, j)| \to 0, \text{ a.s.}$$

By dominated convergence this in particular implies convergence in the ρ_{H} seminorm for any probability H on $[0, 1]^2 \times \{0, 1\}$.

Hence, corollaries 1 and 2 may be invoked to show consistency of the permuation and bootstrap tests based on $\sup_{0<t\leq 1} |T(\hat{S}(t)^\rho, t)|$. □

Remark 3. The assumptions $\|\mathrm{P}\|_{\mathscr{F}} < \infty$ and $\|\mathrm{Q}\|_{\mathscr{F}} < \infty$ in theorem 2 and 4 can not in general be dispensed with. To see this, let P = uniform [0, 1[, Q = uniform [1, 2]; let g be any function with $\int_0^2 g^2\, dx < \infty$, and define a function class by

$$\mathscr{F} \equiv \{f_{a,b} \equiv g + a1_{[0,1[} + b1_{[1,2]} \colon a, b \in \mathbf{R}\}.$$

Then $\mathscr{F}$ is trivially P and Q Donsker and $\|\mathrm{P}\|_{\mathscr{F}} = \|\mathrm{Q}\|_{\mathscr{F}} = \infty$. Let H be given as in theorem 1 and define the pseudometric ρ_{H} by $\rho_{\mathrm{H}}^2(f, g) \equiv \mathrm{Var}_{\mathrm{H}}(f(Z) - g(Z)) = \mathrm{H}(f-g)^2 - (H(f-g))^2$; then for $a, b, c, d \in \mathbf{R}$

$$\begin{aligned}\rho_{\mathrm{H}}^2(f_{a,b}, f_{c,d}) &= \mathrm{var}_{\mathrm{H}}((a-c)1_{[0,1[} + (b-d)1_{[1,2]}) \\ &= (a-c)^2 \mathrm{var}_{\mathrm{H}} 1_{[0,1[} + (b-d)^2 \mathrm{var}_{\mathrm{H}} 1_{[1,2]} \\ &\quad + 2(a-c)(b-d)\,\mathrm{cov}_{\mathrm{H}}(1_{[0,1]}, 1_{[1,2]}) \\ &= (a-b-c+d)^2(\lambda(1-\lambda))\end{aligned}$$

by noting that if $Z \sim \mathrm{H}$, then

$$(1_{[0,1[}(Z), 1_{[1,2[}(Z)) \sim \mathrm{mult}_2(1, (\lambda, 1-\lambda)).$$

Hence if $0 < \lambda < 1$, it follows that $\sup_{f,g \in \mathscr{F}} \rho_{\mathrm{H}}(f, g) = \infty$; i.e. $(\mathscr{F}, \rho_{\mathrm{H}})$ is not totally bounded and does not support an H-bridge.

3. Power under a sequence of local alternatives

The results of the preceding section show that both the bootstrap and permutation methods applied to the two-sample Smirnov statistic yield a consistent test against any fixed alternative for which $\|\mathrm{P}-\mathrm{Q}\|_{\mathscr{F}_0} \neq 0$. In order to compare the power of the two tests it may be more interesting to consider a situation where the Nth stage data $X_{N_1}, \ldots, X_{N_m}$ and $Y_{N_1}, \ldots, Y_{N_n}$, $N = m + n \equiv m(N) + n(N)$, are sampled i.i.d. from measures P_N and Q_N, respectively, and these probabilities tend "smoothly" to limiting measures P and Q with $\|\mathrm{P}-\mathrm{Q}\|_{\mathscr{F}_0} = 0$ as $N \to \infty$. Finding the asymptotic power of the tests under this model may be regarded as expanding their power around P and Q to the first order. For these power calculation we shall first study the permutation and bootstrap empirical processes when the Nth stage probability measures P_N and Q_N are Hellinger differentiable at two probabilities P and Q:

$$\int (\sqrt{N}(\sqrt{d\mathrm{P}_N} - \sqrt{d\mathrm{P}}) - \tfrac{1}{2}h_{\mathrm{P}}\sqrt{d\mathrm{P}})^2 \to 0, \tag{3.23}$$

and

$$\int (\sqrt{N}(\sqrt{d\mathrm{Q}_N} - \sqrt{d\mathrm{Q}}) - \tfrac{1}{2}h_{\mathrm{Q}}\sqrt{d\mathrm{Q}})^2 \to 0. \tag{3.24}$$

Here the "tangents" $h_{\mathrm{P}} \in L^2(\mathrm{P})$, $h_{\mathrm{Q}} \in L^2(\mathrm{Q})$ satisfy $\mathrm{P}h_{\mathrm{P}} = \mathrm{Q}h_{\mathrm{Q}} = 0$. The densities $d\mathrm{P}_N$ and $d\mathrm{P}$ in (3.23) are with respect to any measure dominating them both, e.g. $\mathrm{P}_N + \mathrm{P}$; the value of h_{P} remains the same. The same remark applies to (3.24). For more on Hellinger differentiability see e.g. Bickel *et al.* (1992).

We retain all definitions of empirical measures and processes and permutation and bootstrap empirical measures and processes, keeping in mind the new distributional assumptions. One way to model these is to assume that the data $\{((X_{N_1}, \ldots, X_{N_{m(N)}}), (Y_{N_1}, \ldots, Y_{N_{n(N)}}), N = 1, 2, \ldots\}$ are defined jointly on the canonical probability space

$$\begin{aligned}(\Omega, \mathscr{O}, \mathrm{Pr}) \equiv ((\mathbf{Z}, \mathscr{Z}, \mathrm{P}_1)^{m(1)} \times (\mathbf{Z}, \mathscr{Z}, \mathrm{Q}_1)^{n(1)}) \times \cdots \times \\ ((\mathbf{Z}, \mathscr{Z}, \mathrm{P}_N)^{m(N)} \times (\mathbf{Z}, \mathscr{Z}, \mathrm{Q}_N)^{n(N)}) \times \cdots.\end{aligned} \tag{3.25}$$

With this setup we have the following result concerning the limiting distribution of the permutation and bootstrap empirical processes. The proof is contained in section 4.

Theorem 5

Let $\{\mathrm{P}_N\}$ *and* $\{\mathrm{Q}_N\}$ *be sequences of probability measures on* $(\mathbf{Z}, \mathscr{Z})$ *satisfying* (3.23) *and* (3.24). *Let* $\mathscr{F}$ *be a* P- *and* Q-*Donsker class of functions with* $\mathscr{F} \subset L_2(\mathrm{P}) \cap L_2(\mathrm{Q})$, $\|\mathrm{P}\|_{\mathscr{F}} < \infty$, $\|\mathrm{Q}\|_{\mathscr{F}} < \infty$ *and* $\mathscr{F} \in M(\mathrm{P}) \cap M(\mathrm{Q})$. *Assume further that* $m/N \equiv m(N)/N \to \lambda \in (0, 1)$ *as* $N \to \infty$. *Then it holds for the permutation and bootstrap empirical process that*

$$\tilde{\mathbf{X}}^{\omega}_{m,N} \Rightarrow (1-\lambda)G_{\mathrm{H}}, \text{ in } \mathrm{Pr}^* \text{ probability},$$

and

$$\hat{\mathbf{X}}^{\omega}_{m,N} \Rightarrow G_{\mathrm{H}}, \text{ in } \mathrm{Pr}^* \text{ probability}$$

where G_{H} *is defined in theorem* 1.

As corollaries to theorem 5 we find the asymptotic power of the bootstrap and permutation tests (1.2) and (1.5).

Corollary 3. (The local asymptotic power of the permutation Kolmogrov–Smirnov test) *Let the assumptions be as in theorem* 5, *and assume further that* $\mathscr{F}_N, \mathscr{F}_0 \subset \mathscr{F}$ *are subclasses for which* (2.14) *and* (2.15) *both hold. Assume that* $\|\mathrm{P}-\mathrm{Q}\|_{\mathscr{F}_0} > 0$. *Then the power of the test given by* (1.2) *satisfies*

$$\Pr{}^*(\tilde{\phi}_N(D_N)=1) \to \Pr(\|G_\mathrm{H} + \sqrt{1-\lambda}\,\delta_{h_\mathrm{P}} - \sqrt{\lambda}\,\delta_{h_\mathrm{Q}}\|_{\mathscr{F}_0} > c_\mathrm{H}(\alpha))$$

where $\delta_{h_\mathrm{P}}(f) \equiv \mathrm{P}(h_\mathrm{P} f) = \int h_\mathrm{P} f \, d\mathrm{P}$ *for* $f \in \mathscr{F}$.

Proof of corollary 3. By theorem 5 it follows, as in the proof of corollary 1, that $((mn)/N)^{1/2}\tilde{c}_N^\omega(\alpha) \xrightarrow{\Pr^*} c_\mathrm{H}(\alpha)$, the $1-\alpha$ quantile of $\|G_\mathrm{H}\|_{\mathscr{F}_0}$. By Wellner (1992, th. III.2.1), it holds that $\mathbf{X}_m \Rightarrow G_\mathrm{P} + \delta_{h_\mathrm{P}}$ and $\mathbf{Y}_n \Rightarrow G'_\mathrm{Q} + \delta_{h_\mathrm{Q}}$ as $N \to \infty$. By noting (2.19) we find that

$$\begin{aligned}\sqrt{\frac{mn}{N}}\,D_N &= \left\|\sqrt{\frac{n}{N}}\,\mathbf{X}_m - \sqrt{\frac{m}{N}}\,\mathbf{Y}_n\right\|_{\mathscr{F}_N} \\ &\Rightarrow \|\sqrt{1-\lambda}(G_\mathrm{P} + \delta_{h\mathrm{P}}) - \sqrt{\lambda}(G'_\mathrm{Q} + \delta_{h_\mathrm{Q}})\|_{\mathscr{F}_0} \\ &\overset{\mathscr{L}}{=} \|G_\mathrm{H} + \sqrt{1-\lambda}\,\delta_{h_\mathrm{P}} - \sqrt{\lambda}\,\delta_{h_\mathrm{Q}}\|_{\mathscr{F}_0}\end{aligned}$$

where the tangents $\delta_{h_\mathrm{P}} = \mathrm{P}(h_\mathrm{P}\,\cdot\,)$ and $\delta_{h_\mathrm{Q}} = \mathrm{P}(h_\mathrm{Q}\,\cdot\,)$ are given in (3.23) and (3.24). Hence it also holds, as in (2.18), that

$$\sqrt{\frac{mn}{N}}\,(D_N, \tilde{c}_N^\omega(\alpha)) \xrightarrow{\Pr^*} (\|G_\mathrm{P} + \sqrt{1-\lambda}\,\delta_{h_\mathrm{P}} - \sqrt{\lambda}\,\delta_{h_\mathrm{Q}}\|_{\mathscr{F}_0}, c_\mathrm{H}(\alpha)).$$

The remaining part of the proof follows by the expression (2.17) and the Portmanteau theorem. □

Corollary 4. (The local asymptotic power of the bootstrap Kolmogorov–Smirnov test) *Let the assumptions be as in corollary* 3. *Then it further holds for the test* (1.5) *that*

$$\Pr{}^*(\hat{\phi}_N(D_N)=1) \to \Pr(\|G_\mathrm{P} + \sqrt{1-\lambda}\,\delta_{h_\mathrm{P}} - \sqrt{\lambda}\,\delta_{h_\mathrm{Q}}\|_{\mathscr{F}_0} > c_\mathrm{H}(\alpha))$$

Proof of corollary 4. The proof is similar to that of corollary 3, noting that by the same reasoning as in corollary 2 it holds that $((mn)/N)^{1/2}\hat{c}_N^\omega(\alpha) \xrightarrow{\Pr^*} c_\mathrm{H}(\alpha)$. □

4. Proof for sections 2 and 3

Proof of theorem 1. Notice first of all that the $L_2(\mathrm{P})$, $L_2(\mathrm{Q})$, and $L_2(\mathrm{H})$ pseudometrics are related by

$$\begin{aligned}e_\mathrm{H}^2(f,g) &\equiv \mathrm{H}(f-g)^2 = \lambda\mathrm{P}(f-g)^2 + (1-\lambda)\mathrm{Q}(f-g)^2 \\ &\equiv \lambda e_\mathrm{P}^2(f,g) + (1-\lambda)e_\mathrm{Q}^2(f,g).\end{aligned} \tag{4.26}$$

Since $\mathscr{F}$ is a P- as well as Q-Donsker class and $\|\mathrm{P}\|_{\mathscr{F}}, \|\mathrm{Q}\|_{\mathscr{F}} < \infty$, it follows by e.g. Giné & Zinn (1986, th. 1.2.8), that $\mathscr{F}$ is totally bounded under e_P as well as under e_Q. Hence by (4.26), $(\mathscr{F}, e_\mathrm{H})$ is totally bounded. By Pollard (1990, th. 10.2), and the Cramér-Wold device, it now suffices for proving theorem 1 to show that

$$\tilde{\mathbf{X}}_{m,N}^\omega(\phi) \Rightarrow \sqrt{1-\lambda}\,G_\mathrm{H}(\phi), \quad \forall\phi \equiv \sum_{l=1}^p c_l f_l,\ c_1,\ldots,c_p \in \mathbf{R}, f_1,\ldots,f_p \in \mathscr{F}, \text{ Pr a.s.} \tag{4.27}$$

and

$$\lim_{\delta\to 0}\limsup_{N\to\infty} E\|\tilde{\mathbf{X}}_{m,N}^\omega\|_{\mathscr{F}'(e_\mathrm{H},\delta)} = 0, \quad \text{Pr a.s.} \tag{4.28}$$

We prove finite dimensional convergence (4.27) first. For this, we use the following variation on a rank statistics central limit theorem due to Hájek (1961). It is due to Mason & Newton (1992); see also Præstgaard & Wellner (1993).

Lemma 1

Let $\{k\}$ be a sequence of natural numbers $\{a_{kj}\}$ a triangular array of constants, and let B_{kj}, $j=1, \ldots, k$, $k \in \{k\}$ be a triangular array of row-exchangeable random variables such that

$$\frac{1}{k}\sum_{j=1}^{k}(a_{kj}-\bar{a}_k)^2 \to \sigma^2 > 0; \quad \frac{1}{k}\max_{j\leqslant k}(a_{kj}-\bar{a}_k)^2 \to 0. \tag{4.29}$$

$$\frac{1}{k}\sum_{j=1}^{k}(B_{kj}-\bar{B}_k)^2 \to \alpha^2 > 0, \quad \textit{in probability}. \tag{4.30}$$

$$\lim_{K\to\infty}\limsup_{k\to\infty} E(B_{k1}-\bar{B}_k)^2 1_{\{|B_{k1}-\bar{B}_k|>K\}} = 0. \tag{4.31}$$

Then

$$\frac{1}{\sqrt{k}}\sum_{j=1}^{k}(a_{kj}B_{kj}-\bar{a}_k\bar{B}_k) \Rightarrow N(0, \alpha^2\sigma^2).$$

Proof of lemma 1. See Præstgaard & Wellner (1993, lem. 4.6). □

Returning to the proof of (4.27) we define constants

$$b_{N_j} \equiv \begin{cases} 1 & \text{if } 1 \leqslant j \leqslant m \\ 0 & \text{if } N \geqslant j > m \end{cases}; \quad \bar{b}_N = \frac{m}{N},$$

and for fixed ϕ and ω

$$a^{\omega}_{N_k}(\phi) \equiv \phi(Z^{\omega}_j); \quad \bar{a}^{\omega}_N(\phi) = \mathbf{H}^{\omega}_N\phi. \tag{4.32}$$

The some algebra and the definition of $\tilde{\mathbf{X}}^{\omega}_{m,N}$ show that

$$\begin{aligned}\tilde{\mathbf{X}}^{\omega}_{m,N}(\phi) &= \frac{1}{\sqrt{m}}\sum_{j=1}^{m}(\phi(Z^{\omega}_{R_N(j)}) - \mathbf{H}^{\omega}_N\phi) \\ &= \frac{1}{\sqrt{m}}\sum_{j=1}^{N}\left(a^{\omega}_{N_j}(\phi)b_{NR_N(j)} - \frac{m}{N}\mathbf{H}^{\omega}_N\phi\right) \\ &= \sqrt{\frac{N}{m}}\frac{1}{\sqrt{N}}\sum_{j=1}^{N}(a^{\omega}_{N_j}(\phi)b_{NR_N(j)} - \bar{a}_N\bar{b}_N).\end{aligned}$$

Apply lemma 1 for fixed ϕ and ω with the sequence $\{k\}$ equal to the natural numbers, $a_{N_j} \equiv a^{\omega}_{N_j}(\phi)$, and $B_{N_j} \equiv b_{NR_N(j)}$, $j=1, \ldots, N$ which are exchangeable because of the permutation. We first argue that (4.29) is satisfied for all choices of ϕ, Pr a.s. We have under the hypotheses of theorem 1 that the class $\mathscr{F}\cdot\mathscr{F} \equiv \{fg : f, g \in \mathscr{F}\}$ is a Glivenko–Cantelli class for P and Q, (this is explained in Giné & Zinn (1990, eq. (2.17), p. 857)). Since ϕ^2 is a weighted sum of a finite number of functions from $\mathscr{F}\cdot\mathscr{F}$, it follows that

$$\begin{aligned}\frac{1}{N}\sum_{j=1}^{N}(a^{\omega}_{N_j}(\phi) - \bar{a}^{\omega}_N(\phi))^2 &= \frac{1}{N}\sum_{j=1}^{N}(\phi(Z^{\omega}_j) - \mathbf{H}^{\omega}_N\phi)^2 = \mathbf{H}^{\omega}_N\phi^2 - (\mathbf{H}^{\omega}_N\phi)^2 \\ &= \left(\frac{m}{N}\mathbf{P}^{\omega}_m + \frac{n}{N}\mathbf{Q}^{\omega}_n\right)\phi^2 - \left(\left(\frac{m}{N}\mathbf{P}^{\omega}_m + \frac{n}{N}\mathbf{Q}^{\omega}_n\right)\phi\right)^2 \\ &\to (\lambda\mathrm{P} + (1-\lambda)\mathrm{Q})\phi^2 - ((\lambda\mathrm{P}+(1-\lambda)\mathrm{Q})\phi)^2 \quad \forall\phi \text{ as in (4.27)}, \\ &\quad \text{Pr a.s.} \\ &= \mathrm{H}\phi^2 - (\mathrm{H}\phi)^2 = \operatorname{var} G_{\mathrm{H}}(\phi).\end{aligned} \tag{4.33}$$

As to the second condition in (4.29), we use again the Glivenko–Cantelli property of $\mathscr{F} \cdot \mathscr{F}$ to show that

$$\frac{1}{N} \max_{j \leqslant N} (a^{\omega}_{N_j}(\phi) - \bar{a}^{\omega}_N(\phi))^2 \leqslant \frac{4}{N} \max_{j \leqslant N} \phi^2(Z^{\omega}_j)$$

$$\leqslant \frac{4}{m} \max_{j \leqslant m} \phi^2(X^{\omega}_j) + \frac{4}{n} \max_{j \leqslant n} \phi^2(Y^{\omega}_j)$$

$$\to 0, \quad \forall \phi \text{ as in (4.27), Pr a.s.} \tag{4.34}$$

Hence (4.29) is satisfied with $\sigma^2 \equiv \text{Var}\, G_{\text{H}}(\phi)$, simultaneously for all ϕ of the form (4.27), Pr a.s. Next, a simple calculation shows that

$$\frac{1}{N} \sum_{j=1}^{N} (B_{N_j} - \bar{B}_N)^2 = \frac{mn}{N^2} \to \lambda(1-\lambda),$$

and

$$\frac{1}{N} \max_{j \leqslant N} (B_{N_j} - \bar{B}_N)^2 \leqslant \frac{1}{N} \max \left\{ \frac{n^2}{N^2}, \frac{m^2}{N^2} \right\} \to 0;$$

and hence (4.30) is satisfied with $\alpha^2 \equiv \lambda(1-\lambda)$. The condition (4.31) is obviously satisfied since by definition $|B_{N1} - \bar{B}_N| \leqslant 1$. Hence the conditions of lemma 1 are satisfied; since finally $(N/m)^{1/2} \to \lambda^{-1/2}$ was assumed, (4.27) follows.

For the asymptotic equicontinuity part (4.28) we use some results from the theory of probability in Banach spaces: randomization with Rademachers, Poissonization, sample path continuity of P and Q Brownian bridges, a key inequality due to Ledoux, Talagrand, and Zinn, and a finite-sampling inequality originally due to Hoeffding (1963). Only this latter is not well referenced in the probability in Banach spaces literature, but it is explained in Præstgaard & Wellner (1993). We state a version of it here.

Lemma 2

Let $\hat{Z}_{N1}, \ldots, \hat{Z}_{NN}$ denote an i.i.d. sample from $\mathbf{H}^{\omega}_N$, and let R be a random permutation of $\{1, \ldots, N\}$. Then for any $m \leqslant N$

$$E \left\| \sum_{i=1}^{m} (\delta_{Z^{\omega}_{NR(i)}} - \mathbf{H}^{\omega}_N) \right\| \leqslant E \left\| \sum_{i=1}^{m} (\delta_{\hat{Z}_{N_i}} - \mathbf{H}^{\omega}_N) \right\|.$$

Proof of lemma 2. See Præstgaard & Wellner (1993, corol. 4.1). □

As to the proof of (4.28), let $\|\cdot\| \equiv \|\cdot\|_{\mathscr{F}'(e_{\text{H}}, \delta)}$ for any $\delta > 0$. First, Hoeffding's inequality (lemma 2) shows that

$$E\|\tilde{\mathbf{X}}^{\omega}_{m,N}\| = E\left\| \frac{1}{\sqrt{m}} \sum_{j=1}^{m} (\delta_{Z^{\omega}_{RN(j)}} - \mathbf{H}^{\omega}_N) \right\| \leqslant E\left\| \frac{1}{\sqrt{m}} \sum_{j=1}^{m} (\delta_{\hat{Z}_{N_j}} - \mathbf{H}^{\omega}_N) \right\|, \tag{4.35}$$

and secondly, symmetrization with $\varepsilon_1, \ldots, \varepsilon_n$ i.i.d. Rademachers and Le Cam's Poissonization lemma (see Giné & Zinn, 1990, lem. 2.1) shows the bound

$$E\left\| \frac{1}{\sqrt{m}} \sum_{j=1}^{m} (\delta_{\hat{Z}_{N_j}} - \mathbf{H}^{\omega}_N) \right\| \leqslant 2E\left\| \frac{1}{\sqrt{m}} \sum_{j=1}^{m} \varepsilon_j \delta_{\hat{Z}_{N_j}} \right\| \leqslant 4E\left\| \frac{1}{\sqrt{m}} \sum_{j=1}^{N} \tilde{N}^{(m)}_j \delta_{Z^{\omega}_j} \right\| \tag{4.36}$$

where $\tilde{N}^{(m)}_1, \ldots, \tilde{N}^{(m)}_N$ are i.i.d. symmetrized Poisson $(m/2N)$. (A symmetrized Poisson (λ) random variable is the difference of two i.i.d Poisson (λ) random variables.) By the closedness of the Poisson family under convolution and Jensen's inequality, (4.36) if further bounded by

$$4E\left\|\frac{1}{\sqrt{m}}\sum_{j=1}^{N}\tilde{N}_j\delta_{Z_j^\omega}\right\| \leqslant 4E\left\|\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\tilde{N}_j\delta_{X_j^\omega}\right\| + 4E\left\|\frac{1}{\sqrt{m}}\sum_{j=1}^{n}\tilde{N}_j\delta_{Y_j^\omega}\right\|, \tag{4.37}$$

with $\tilde{N}_1, \ldots, \tilde{N}_N$ i.i.d. symmetrized Poisson (1/2). Notice finally by (4.26) that $e_{\mathrm{H}}(f, g) < \delta$ implies that $e_{\mathrm{P}}(f, g) < \delta/\lambda^{1/2}$ and $e_{\mathrm{Q}}(f, g) < \delta/(1-\lambda)^{1/2}$, so that $\mathscr{F}'(e_{\mathrm{H}}, \delta) \subset \mathscr{F}'(e_{\mathrm{P}}, \delta/\lambda^{1/2}) \cap \mathscr{F}'(e_{\mathrm{Q}}, \delta/(1-\lambda)^{1/2})$, and hence by (4.37) it suffices for (4.27) that

$$\lim_{\delta\to 0}\limsup_{n\to\infty} E\left\|\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\tilde{N}_j\delta_{X_j^\omega}\right\|_{\mathscr{F}'(e_{\mathrm{P}}, \delta)} = 0, \quad \text{Pr a.s.,}$$

and

$$\lim_{\delta\to 0}\limsup_{n\to\infty} E\left\|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\tilde{N}_j\delta_{Y_j^\omega}\right\|_{\mathscr{F}'(e_{\mathrm{Q}}, \delta)} = 0, \quad \text{Pr a.s.}$$

By the Ledoux–Talagrand–Zinn inequality (Ledoux & Talagrand, 1988 or Giné & Zinn, 1990, lem. 2.4), it holds for $\delta > 0$ that

$$\limsup_{n\to\infty} E\left\|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\tilde{N}_j\delta_{X_j^\omega}\right\|_{\mathscr{F}'(e_{\mathrm{P}}, \delta)} \leqslant 4\limsup_{n\to\infty} E^*\left\|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\tilde{N}_j\delta_{X_j}\right\|_{\mathscr{F}'(e_{\mathrm{P}}, \delta)} = 0$$

where the last equality follows by Giné & Zinn (1986, th. 1.2.8). The term involving $Y_1, \ldots, Y_n$ may be handled in the same manner. □

Proof of theorem 2. It suffices to show that any sequence $\{N'\}$ has a further sub-sequence $\{N''\}$ along which $d^*_{\mathrm{BL}}(\tilde{\mathbf{X}}^\omega_{m'', N''}, (1-\lambda)^{1/2}G_{\mathrm{H}})^* \to 0$, a.s. Since th. 10.2 of Pollard (1990) still holds along a sub-sequence of the natural numbers, it suffices for theorem 2 to prove that any sequence $\{N'\}$ has a further sub-sequence $\{N''\}$ along which

$$\tilde{\mathbf{X}}^\omega_{m'', N''}(\phi) \Rightarrow \sqrt{1-\lambda}G_{\mathrm{H}}(\phi), \quad \forall\phi \equiv \sum_{l=1}^{p} c_l f_l, c_1, \ldots, c_p \in \mathbf{R}, f_1, \ldots, f_p \in \mathscr{F}, \text{Pr a.s.,} \tag{4.38}$$

and

$$\lim_{\delta\to 0}\limsup_{N''\to\infty} E\|\tilde{\mathbf{X}}^\omega_{m'', N''}\|_{\mathscr{F}'(e_{\mathrm{H}}, \delta)} = 0, \quad \text{Pr a.s.} \tag{4.39}$$

Given any sequence $\{N'\}$ the only change from the proof of (4.27) needed to prove (4.38) is to find a sub-sequence $\{N''\}$ along which lemma 1 can be applied; that is along which (4.33) and (4.34) still hold. When $\mathscr{F} \in CLT(\mathrm{P}) \cap CLT(\mathrm{Q})$, it follows, again from Giné & Zinn (1990, p. 853), that $\mathscr{F}\cdot\mathscr{F}$ is a weak Glivenko–Cantelli class for P and Q; i.e. $\|\mathbf{P}_m - \mathrm{P}\|_{\mathscr{F}} \xrightarrow{\mathrm{Pr}^*} 0$ and $\|\mathbf{Q}_n - \mathrm{Q}\|_{\mathscr{F}} \xrightarrow{\mathrm{Pr}^*} 0$. Then any sub-sequence $\{N''\}$ so that $\|\mathbf{P}_{m''} - \mathrm{P}\|^*_{\mathscr{F}} \to 0$ and $\|\mathbf{Q}_{n''} - \mathrm{Q}\|^*_{\mathscr{F}} \to 0$ a.s. will satisfy (4.38).

For (4.39) it suffices to show that

$$\lim_{\delta\to 0}\limsup_{n\to\infty} E^*\|\tilde{\mathbf{X}}_{m, N}\|_{\mathscr{F}(e_{\mathrm{H}}, \delta)} = 0.$$

But this is contained in the proof of theorem 1; in fact the Ledoux–Talagrand–Zinn inequality is no longer needed. □

Proof of theorem 3. Again it suffices to show that (4.27) and (4.28) are satisfied with $\tilde{\mathbf{X}}^\omega_{m, N}$ replaced by $\hat{\mathbf{X}}^\omega_{m, N}$ and the limit $(1-\lambda)^{1/2}G_{\mathrm{H}}(\phi)$ replaced by $G_{\mathrm{H}}(\phi)$. But since by (4.35) $E\|\tilde{\mathbf{X}}^\omega_{m, N}\|_{\mathscr{F}'(e_{\mathrm{H}}, \delta)} \leqslant E\|\hat{\mathbf{X}}^\omega_{m, N}\|_{\mathscr{F}'(e_{\mathrm{H}}, \delta)}$, the present proof of (4.28) (that is a.s. asymptotic equicontinuity of $\hat{\mathbf{X}}^{\omega}_{m,N}$) is contained in the proof of theorem 1 from (4.36) and onwards. For (4.27) we use the expression (2.20) of the bootstrap empirical process, and we show that the conditions of lemma 1 are satisfied with $\{k\} \equiv 1, 2, \ldots, a^{\omega}_{N_j}(\phi)$ as in (4.32) and

$$(B_{N_1}, \ldots, B_{NN}) \sim \text{mult}_N (m, (N^{-1}, \ldots, N^{-1})).$$

For showing that (4.29) is satisfied with $\sigma^2 \equiv \text{var}\, G_{\mathrm{H}}(\phi)$ there is no change from the proof of theorem 1. All that needs to be shown is that the $\{B_{N_j}\}$ satisfy (4.30) and (4.31). The proof of these properties of multinomial vectors is straightforward if tedious, here is one possibility. The following moment calculations can, except for the covariance be found in Johnson & Kotz (1969, p. 51):

$$EB^2_{N_1} = \frac{m}{N} + \frac{m^{(2)}}{N^2},$$

$$EB^4_{N_1} = \frac{m}{N} + 7\frac{m^{(2)}}{N^2} + 6\frac{m^{(3)}}{N^3} + \frac{m^{(4)}}{N^4} \leqslant 15 \text{ (crude)}, \tag{4.40}$$

and

$$\text{cov}\,(B^2_{N_1}, B^2_{N_2}) = \frac{m^{(4)} - (m^{(2)})^2}{N^4} + 2\frac{m^{(3)} - nm^{(2)}}{N^3} + \frac{m^{(2)} - m^2}{N^2} < 0.$$

To see that (4.30) holds, notice first of all that

$$\begin{aligned}\Pr\left(\left|\frac{1}{N}\sum_{j=1}^{N} (B^2_{N_j} - EB^2_{N_j})\right| > \varepsilon\right) &\leqslant \varepsilon^{-2}N^{-2}\,\text{var}\left(\sum_{j=1}^{N} (B^2_{N_j} - EB^2_{N_j})\right)\\ &\leqslant \varepsilon^{-2}N^{-2}\sum_{j=1}^{N} E(B^2_{N_j} - EB^2_{N_j})^2\\ &= \varepsilon^{-2}N^{-1}E(B^2_{N_j} - EB^2_{N_j})^2\\ &\leqslant 15\varepsilon^{-2}N^{-1} \to 0\end{aligned}$$

where we have used Chebychev's inequality and the fact that by (4.40) $B^2_{N_i}$ and $B^2_{N_j}$ are non-positively correlated for $i \neq j$. Hence

$$\frac{1}{N}\sum_{j=1}^{N} B^2_{N_j} - EB^2_{N_1} \to 0 \quad \text{in probability;}$$

since $EB^2_{N_1} - \bar{B}^2_N = m/N + m^{(2)}/N^{(2)} - (m/N)^2 \to \lambda$, it follows that (4.30) is satisfied with $\alpha = \lambda$. The uniform square integrability (4.31) follows because of the uniform bound on the fourth moments (4.40).

It now follows from lemma 1 and $(N/m)^{1/2} \to \lambda^{-1/2}$ that (4.27) is satisfied for $\hat{\mathbf{X}}^{\omega}_{m,N}$, and hence the proof of theorem 3 is complete. □

Proof of theorem 4. The only difference from theorem 3 is to show that the conditions corresponding to (4.38) and (4.39) hold along a sub-sequence. For the finite-dimensional part (4.38, the array $\{B_{N_j}\}$ still satisfies (4.30) and (4.31), so as in the proof of theorem 2, all that is needed is the weak Glivenko–Cantelli property of $\mathscr{F} \cdot \mathscr{F}$. The asymptotic equicontinuity part (4.39) follows as in theorem 3; since we are no longer fixing ω, there is no need to invoke the Ledoux–Talagrand–Zinn inequality. □

Proof of theorem 5. When (3.23) and (3.24) are satisfied, it follows from e.g. Wellner (1992, th. III.2.1, p. 257) that the sequence $\{\mathrm{P}^{\infty}_N\}$ is contiguous to P^{∞} and $\{\mathrm{Q}^{\infty}_N\}$ is contiguous to Q^∞, when all of these measures are defined on the probability space $(\mathbf{Z}, \mathscr{Z})^\infty$. Hence the product sequence $\{(P_N \times Q_N)^\infty\}$ is contiguous to $(P \times Q)^\infty$. Denote $(P_N \times Q_N)^\infty \equiv \mathrm{Pr}_N$ and $(P \times Q)^\infty \equiv \mathrm{Pr}_0$, not to be confused for Pr of (3.25). The assumptions of theorem 5 include those of theorems 2 and 4, and hence it holds that

$$d_{\mathrm{BL}^*}(\tilde{\mathbf{X}}^{\omega}_{m,N}, \sqrt{1-\lambda}G_{\mathrm{H}}) \xrightarrow{\mathrm{Pr}_0^*} 0, \tag{4.41}$$

and hence by one of the properties of contiguity, see e.g. Van der Vaart & Wellner (1994), (4.41) also holds in Pr_N^* probability. (The same contiguity result for weak convergence in the ordinary, Billingsley, setting can be found in e.g. Shorack & Wellner (1986, p. 157). This observation proves theorem 5, by noting that for each N, the envelope function of $d_{\mathrm{BL}^*}(\tilde{\mathbf{X}}^{\omega}_{m,N}, \sqrt{1-\lambda}G_{\mathrm{H}})$ when defined on $(\mathbf{Z}^\infty, \mathscr{Z}^\infty, \mathrm{Pr}_N)$ is identical a.s. Pr to the envelope when defined on the probability space (3.25). □

Acknowledgement

I thank Jon Wellner for helpful advice. Two anonymous referees and an associate editor helped improve the presentation greatly.

References

Andersen, N. T. (1985). The central limit theorem for non-separable valued functions. *Zeit. Wahrsch. Verv. Geb.* **70,** 445–455.

Bickel, P. (1969). A distribution free version of the Smirnov two sample test in the p-variate case. *Ann. Math. Statist.* **50,** 1–23.

Bickel, P., Klaassen, C. J., Ritov, Y. & Wellner, J. A. (1992). *Efficient and adaptive estimation in semiparametric models.* Wiley, New York.

Dudley, R. M. (1985). An extended Wichura thoerem, definitions of Donsker classes, and weighted empirical distributions. *Lecture Notes Math.* **1143,** 141–178. Springer Verlag, New York.

Efron, B. & Tibshirani, R. J. (1993). *An introduction to the bootstrap.* Chapman & Hall, London.

Fleming, T. R., Harrington, D. P. & O'Sullivan, M. (1987). Supremum versions of the log-rank and generalized wilcoxon statistics. *J. Amer. Statist. Assoc.* **82,** 312–320.

Giné, E. & Zinn, J. (1984). Some limit theorems for empirical processes. *Ann. Probab.* **4,** 929–990.

Giné, E. & Zinn, J. (1986). Lectures on the central limit theorem for empirical processes. *Lecture Notes Math.* **1221,** 50–113. Springer Verlag, New York.

Giné, E. & Zinn, J. (1990). Bootstrapping general empirical functions. *Ann. Probab.* **18**; 851–869.

Hájek, J. (1961). Some extensions of the Wald–Wolfowitz–Noether theorem. *Ann. Math. Statist.* **32,** 506–523.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58,** 13–30.

Hoffmann-Jørgensen, J. (1984). Stochastic process on Polish spaces. Unpublished manuscript.

Johnson, N. & Kotz, S. (1969). *Distributions in statistics. Discrete distributions.* Houghton Mifflin, Boston.

Ledoux, M. & Talagrand, M. (1988). Un critère sur les petites boules dans le théorème limite central. *Probab. Theory Related Fields* **77,** 29–47.

Ledoux, M. & Talagrand, M. (1991). *Probability in Banach spaces.* Springer-Verlag, New York.

Mason, D. M. & Newton, M. A. (1992). A rank statistics approach to the consistency of a general bootstrap. *Ann. Statist.* **20,** 1611–1624.

Pollard, D. (1990). Empirical processes: theory and applications. *NSF-CMBS Regional Conference Series in Probability and Statistics.* **2.** Institute of Mathematical Statistics, Hayward, CA.

Præstgaard, J. T. & Wellner, J. A. (1993). Exchangeably weighted bootstraps of the general empirical process. *Ann. probability*, **21,** 2053–2086.

Romano, J. P. (1989). Bootstrap and randomization tests of some nonparametric hypotheses. *Ann. Statist.* **17,** 141–159.

Shorack, G. & Wellner, J. (1986). *Empirical processes.* Wiley, New York.

Tsirelson, V. S. (1975). The density of the maximum of a Gaussian process. *Theor. Probab. Appl.* **20,** 847–856.

Van der Vaart, A. W. and Wellner, J. A. (1994). Weak convergence and empirical processes. *Institute of Mathematical Statistics Lecture Notes—Monograph Series*, to appear. Institute of Mathematical Statistics, Hayward, CA.

Wellner, J. A. (1992). Empirical processes in action: a review. *Int. Statist. Review*, **60,** 247–269.

Received February 1993, in final form October 1994.

Jens Thomas Præstgaard, Department of Statistics and Actuarial Science, University of Iowa, Iowa City, Iowa 52242, USA.