Submitted to Bernoulli

Empirical process of concomitants for partly categorial data and applications in statistics

DANIEL GAIGALL[1], JULIAN GERSTENBERG[1] and THI THU HA TRINH[1]

[1]*Institute of Probability and Statistics, Leibniz University Hannover, Welfengarten 1, 30167 Hannover, Germany. E-mail: jgerst@stochastik.uni-hannover.de (Julian Gerstenberg)*

On the basis of independent and identically distributed bivariate random vectors, where the components are categorial and continuous variables, respectively, the related concomitants, also called induced order statistic, are considered. The main theoretical result is a functional central limit theorem for the empirical process of the concomitants in a triangular array setting. A natural application is hypothesis testing. An independence test and a two-sample test are investigated in detail. The fairly general setting enables limit results under local alternatives and bootstrap samples. For the comparison with existing tests from the literature simulation studies are conducted. The empirical results obtained confirm the theoretical findings.

Keywords: categorial variable, concomitant, induced order statistic, empirical process, independence test, two-sample test, bootstrap, triangular array, local alternatives.

Revision

1. Introduction

Given that $n \in \mathbb{N}$ is the sample size, let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent and identically distributed bivariate random vectors. We suppose that the random variable X_1 is categorial with values in a set Σ of cardinality $m \in \mathbb{N}$. For convenience, let $\Sigma = \{1, \ldots, m\}$ with the note that the categories are not necessarily ordinal but can be chosen arbitrary. It is supposed that the probability that X_1 takes the value $1, \ldots, m$ is different from zero, respectively. In addition, we assume that the random variable Y_1 takes values in $\mathbb{R}$ with continuous distribution function F. Let the joint distribution of (X_1, Y_1) as well as the marginal distributions of X_1 and Y_1 be unknown. We assume that $Y_1, \ldots, Y_n$ are pair-wise distinct without loss of generality. Let $R_{1:n}, \ldots, R_{n:n}$ be the ranks of $Y_1, \ldots, Y_n$,

$$R_{j:n} = \sum_{l=1}^{n} 1(Y_l \leqslant Y_j),\ j = 1, \ldots, n,$$

and denote by $R_{1:n}^{-1}, \ldots, R_{n:n}^{-1}$ the inverse ranks of $Y_1, \ldots, Y_n$ such that

$$Y_{R_{1:n}^{-1}} < \cdots < Y_{R_{n:n}^{-1}}.$$

Then, the random variables

$$X_{[j:n]} = X_{R_{j:n}^{-1}},\ j = 1, \dots, n,$$

are called the concomitants, also called induced order statistic, of $X_1, \dots, X_n$ with respect to $Y_1, \dots, Y_n$. In what follows, we speak shortly about the concomitants $X_{[1:n]}, \dots, X_{[n:n]}$.

Concomitants appear when we sort the values of X-attributes according to real-valued Y-attributes. There is a large number of works dealing with concomitants. We refer to David and Nagaraja [6] or Kamps [14] for an overview. Applications are given in change-point analysis of time series, see Robins, Lund, Gallagher and Lu [18]. The authors consider a dataset (HURDAT) which contains historic records on $n = 1410$ Atlantic basin cyclones. The cyclone is described by one of $m = 5$ categories, which yields the X-attributes. The Y-attributes are given by the related points in time of the cyclones. The question whether or not there is a change-point in the cyclone category with respect to the time is treated on the basis of the time-ordered observed cyclone categories, i.e., the concomitants. A cumulative sum type test statistic (CUSUM) is used and the asymptotic of the test statistic is obtained. In fact, the CUSUM test detects change-points by application to the HURDAT dataset. Further examples for the application of concomitants are outlined in David and Nagaraja [5], where in the first example the top $k \in \{1, \dots, n-1\}$ out of $n \geqslant 2$ rams, as judged by their genetic make-up according to the Y-attributes, are selected for breeding, where the X-attributes represent the quality of the wool of one of their female offspring. These examples show the relevance of categorial attributes in the study of concomitants. They are also topic in more theoretical works, such as in the study of size-biased permutations, see Pitman and Tran [16], we refer to section 5.2. in that work for short historical notes on induced order statistics. Another example is in exchangeability theory, see Gerstenberg [12].

The unknown joint distribution of (X_1, Y_1) is determined by the probabilities $P(X_1 = i, Y_1 \leqslant y)$, $(i, y) \in \Sigma \times \mathbb{R}$. Similar, the unknown joint distribution of X_1 and $F(Y_1)$ is determined by

$$\rho(i, t) = P(X_1 = i, F(Y_1) \leqslant t),\ (i, t) \in \Sigma \times [0, 1]. \tag{1}$$

The joint distribution of (X_1, Y_1) decomposes uniquely into the pair (F, ρ) according to $P(X_1 = i, Y_1 \leqslant y) = \rho(i, F(y))$, $(i, y) \in \Sigma \times \mathbb{R}$, where F corresponds to the marginal distribution of Y_1 and ρ determines the marginal distribution of X_1 and the dependency structure between X_1 and Y_1. In what follows we focus on inference of ρ. Later it will be seen that interesting testing problems can be reduced to inference of ρ. For a moment, suppose that $(X_1, Y_1), \dots, (X_n, Y_n)$ are our observations. For inference of ρ it would be natural to deal with the independent and identically distributed bivariate random vectors $(X_j, F(Y_j))$, $j = 1, \dots, n$. Unfortunately, these random vectors are not available as observations since F is unknown. Estimating F with the empirical distribution function of $Y_1, \dots, Y_n$,

$$\widehat{F}_n(y) = \frac{1}{n} \sum_{j=1}^{n} I(Y_j \leqslant y),\ y \in \mathbb{R},$$

yields the bivariate random vectors $(X_j, \widehat{F}_n(Y_j)), j = 1, \ldots, n$ as observations. From

$$\widehat{F}_n(Y_j) = R_{j:n}/n, \; j = 1, \ldots, n,$$

rearranging according to the inverse ranks of $Y_1, \ldots, Y_n$ yields $(X_{[j:n]}, j/n)$, $j = 1, \ldots, n$, as observations. These deliberations show that the concomitants contain the same information as the bivariate random vectors $(X_j, \widehat{F}_n(Y_j)), j = 1, \ldots, n$. For that reason, it is reasonable to deal with the concomitants as data for inference of ρ in what follows.

Setting

$$\widehat{N}_n(i,l) = \sum_{j=1}^{l} 1(X_{[j:n]} = i), \; i \in \Sigma, \; l = 1, \ldots, n,$$

the unknown values $\rho(i,t)$ can be consistently estimated on the basis of the concomitants by

$$\begin{aligned}\frac{1}{n}\widehat{N}_n(i, \lfloor nt \rfloor) &= \frac{1}{n}\sum_{j=1}^{\lfloor nt \rfloor} 1\Big(X_{[j:n]} = i\Big) \\ &= \frac{1}{n}\sum_{j=1}^{n} 1\Big(X_{[j:n]} = i, \frac{j}{n} \leqslant t\Big) \\ &= \frac{1}{n}\sum_{j=1}^{n} 1(X_j = i, \widehat{F}_n(Y_j) \leqslant t), \; (i,t) \in \Sigma \times [0,1].\end{aligned}$$

This estimator is strong uniformly consistent in $i \in \Sigma, \; t \in [0,1]$ meaning that

$$\sup_{i \in \Sigma, t \in [0,1]} \left| \frac{\widehat{N}_n(i, \lfloor nt \rfloor)}{n} - \rho(i,t) \right| \longrightarrow 0 \text{ almost surely as } n \to \infty. \tag{2}$$

We define

$$\mathcal{G}_n(i,t) = \sqrt{n}\Big(\frac{\widehat{N}_n(i, \lfloor nt \rfloor)}{n} - \rho(i,t)\Big), \; (i,t) \in \Sigma \times [0,1],$$

and introduce the *empirical process of the concomitants* $\mathcal{G}_n = (\mathcal{G}_n(i,t))_{(i,t) \in \Sigma \times [0,1]}$ as the subject of our investigation. A functional Central Limit Theorem (CLT) for the concomitants is under consideration in Theorem 24.3.1 in Davydov and Egorov [7], but in a very general setting. In particular, the concrete structure of the related limit processes are not transparent there. In contrast, we present in Section 2 a functional central limit theorem for the empirical process of the concomitants $\mathcal{G}_n$, where the argumentation in our proof is rather straightforward and use classical empirical process theory as in Ziegler [22] and knowledge about the well-known Bahadur-Kiefer process, see Bahadur [1]. Furthermore, we obtain the concrete distribution of the limit process in a closed form. Thereby, we deal with a fairly general setting of triangular arrays of random variables and by that extend the result of Davydov and Egorov [7].

A natural application of the empirical process of the concomitants is hypothesis testing. Consistent tests can be constructed with the help of functionals of the empirical process of the concomitants. For that reason, limit results for those tests can be deduced from our functional Central Limit Theorem. In Section 3, we consider the testing problem of independence

$$\begin{aligned} \mathcal{H}: &\ \forall\ (i,t) \in \Sigma \times \mathbb{R} : P(X_1 = i, Y_1 \leqslant t) = P(X_1 = i)P(Y_1 \leqslant t), \\ \mathcal{K}: &\ \exists\ (i,t) \in \Sigma \times \mathbb{R} : P(X_1 = i, Y_1 \leqslant t) \neq P(X_1 = i)P(Y_1 \leqslant t), \end{aligned} \tag{3}$$

as the first example for the application of our main result. Tests of independence for partly continuous and partly categorial data are also under consideration in Genest and Remillard [10], Genest et al. [11], Heller et al. [13], Székely, Rizzo and Bakirov [21], Robins, Lund, Gallagher and Lu [18], and Chatterjee [4]. Noticing that

$$P(X_1 = i) = \rho(i,1),\ i \in \Sigma,$$

it is clear that the testing problem is equivalent to

$$\mathcal{H}:\ \forall\ (i,t) \in \Sigma \times [0,1] : \rho(i,t) = t\rho(i,1),\ \mathcal{K}:\ \exists\ (i,t) \in \Sigma \times [0,1] : \rho(i,t) \neq t\rho(i,1).$$

In particular, the testing problem can be reduced to ρ. As it is explained above, this shows that it is reasonable to consider an independence test based on the concomitants. Similar as above, we can estimate $\rho(i,1)$, $i \in \Sigma$, consistently on the basis of the concomitants $X_{[1:n]}, \ldots, X_{[n:n]}$ by

$$\frac{1}{n}\hat{N}_n(i,n) = \frac{1}{n}\sum_{j=1}^{n} 1(X_{[j:n]} = i),\ i \in \Sigma.$$

It is easily seen that

$$\sup_{i\in\Sigma, t\in[0,1]} \left| \frac{\hat{N}_n(i,\lfloor nt \rfloor) - t\hat{N}_n(i,n)}{\hat{N}_n(i,n)} \right| \longrightarrow 0 \text{ almost surely as } n \to \infty$$

holds under the null hypothesis $\mathcal{H}$ as well. These deliberations motivate to use the test statistic

$$T_n = \int_0^1 \sum_{i=1}^{m} \hat{N}_n(i,n) \left(\frac{\hat{N}_n(i,\lfloor nt \rfloor) - t\hat{N}_n(i,n)}{\hat{N}_n(i,n)} \right)^2 dt. \tag{4}$$

Large values of T_n should be significant. We establish a simple expression for the test statistic T_n. Under the null hypothesis, it is seen that the test statistic T_n converges in distribution while the limit is distribution free. Using the quantiles of the limit distribution as critical values, we obtain a test which reaches the significance level and is consistent in the limit. The focus is the application of our main result for the limit distribution of the test statistic T_n under local alternatives. Behavior of a test statistic under local alternatives is of interest, e.g., for efficiency deliberations, in particular for the Pitman-efficiency, see Serfling [20] and Puri and Sen [17], or the Volume-efficiency, see Baringhaus and Gaigall [2]. In the finite sample case, simulations compare this test with tests in Chatterjee [4] and Genest et al. [11] in terms of size and power. The simulation results indicate that the test based on T_n is a serious competitor to concurring tests from the literature.

As another example for the application of our main result, we consider a two-sample problem in Section 4. Consider two independent samples of sizes $n_1, n_2 \in \mathbb{N}$ consisting of bivariate random vectors

$$(X_{1,1}, Y_{1,1}), \ldots, (X_{1,n_1}, Y_{1,n_1}) \quad \text{and} \quad (X_{2,1}, Y_{2,1}), \ldots, (X_{2,n_2}, Y_{2,n_2}),$$

such that for each $k = 1, 2$ the pairs $(X_{k,j}, Y_{k,j})$, $j = 1, \ldots, n_k$ are independent and identically distributed and $(X_{k,1}, Y_{k,1})$ satisfies the properties stated above for (X_1, Y_1) with continuous distribution functions $F_k(t) = P(Y_{k,1} \leqslant t)$. Within each group we consider the concomitants of the X-with respect to the Y-attributes,

$$X_{1,[1:n_1]}, \ldots, X_{1,[n_1:n_1]} \quad \text{and} \quad X_{2,[1:n_2]}, \ldots, X_{2,[n_2:n_2]}.$$

Introducing the set $\mathcal{T}$ of all monotonic transformations from $\mathbb{R}$ to $\mathbb{R}$, we are interested in the two-sample testing problem

$$\begin{aligned} &\mathcal{H} : \exists T \in \mathcal{T} \ \forall (i,t) \in \Sigma \times \mathbb{R} : P(X_{1,1} = i, Y_{1,1} \leqslant t) = P(X_{2,1} = i, T(Y_{2,1}) \leqslant t), \\ &\mathcal{K} : \forall T \in \mathcal{T} \ \exists (i,t) \in \Sigma \times \mathbb{R} : P(X_{1,1} = i, Y_{1,1} \leqslant t) \neq P(X_{2,1} = i, T(Y_{2,1}) \leqslant t). \end{aligned} \tag{5}$$

Setting for $k = 1, 2$

$$\rho_k(i,t) = P(X_{k,1} = i, F_k(Y_{k,1}) \leqslant t), \quad (i,t) \in \Sigma \times [0,1],$$

it will be shown that the two-sample testing problem is equivalent to

$$\mathcal{H} : \forall \ (i,t) \in \Sigma \times [0,1] : \rho_1(i,t) = \rho_2(i,t), \ \mathcal{K} : \ \exists \ (i,t) \in \Sigma \times [0,1] : \rho_1(i,t) \neq \rho_2(i,t), \tag{6}$$

i.e., the testing problem can be reduced to ρ. As it is explained above, this motivates to consider a two-sample test based on the two concomitant sequences $X_{k,[1:n_k]}, \ldots, X_{k,[n_k:n_k]}$, $k = 1, 2$. With

$$\widehat{N}_{k,n_k}(i,l) = \sum_{j=1}^{l} 1(X_{k,[j:n_k]} = i), \quad i \in \Sigma, \ l = 1, \ldots, n_k, \ k = 1, 2,$$

it is easily seen that

$$\sup_{i \in \Sigma, t \in [0,1]} \left| \frac{\widehat{N}_{1,n_1}(i, \lfloor n_1 t \rfloor)}{n_1} - \frac{\widehat{N}_{2,n_2}(i, \lfloor n_2 t \rfloor)}{n_2} \right| \longrightarrow 0 \text{ as } n \to \infty \text{ almost surely as}$$

under the null hypothesis $\mathcal{H}$. For that reason, it is reasonable to use the test statistic

$$T_{n_1,n_2} = \sqrt{\frac{n_1 n_2}{n_1+n_2}} \sup_{(i,t)\in\Sigma\times[0,1]} \Big| \frac{\widehat{N}_{1,n_1}(i,\lfloor n_1 t\rfloor)}{n_1} - \frac{\widehat{N}_{2,n_2}(i,\lfloor n_2 t\rfloor)}{n_2} \Big|. \tag{7}$$

Clearly, large values of T_{n_1,n_2} should indicate that the alternative $\mathcal{K}$ is true. It is seen that the test statistic T_{n_1,n_2} converges in distribution. Because the test statistic is not distribution free, neither under the null hypothesis, nor asymptotically, an approximation of the unknown quantiles of the distribution of T_{n_1,n_2} is demanded. For this purpose, we suggest a bootstrap procedure. We apply our main result and a general result of Bücher and Kojadinovic [3] to show that the resulting bootstrap test reaches the significance level and is consistent in the limit. Simulations investigate size and power of the test in the finite sample case. The empirical results confirm the theoretical findings. We note that it is possible to apply our functional Central Limit Theorem for the concomitants in the context of other testing problems as well, for example a one-sample goodness-of-fit problem analogously to the two-sample case.

2. Convergence of the empirical process of the concomitants

In this section we provide a functional CLT for the empirical process of the concomitants in the situation of triangular arrays. To be more precise, we assume that $(X_j^{(n)}, Y_j^{(n)}),\ j = 1,\dots,n$, are independent and identically distributed and $(X_1^{(n)}, Y_1^{(n)})$ has properties as stated above. We denote by $X_{[j:n]}^{(n)},\ j = 1,\dots,n$, the concomitants. Let $F^{(n)}$ be the continuous distribution function of the real-valued random variable $Y_1^{(n)}$. Now we have

$$\rho^{(n)}(i,t) = P(X_1^{(n)} = i, F^{(n)}(Y_1^{(n)}) \leqslant t),\ (i,t) \in \Sigma\times[0,1],$$

and

$$\mathcal{G}_n(i,t) = \sqrt{n}\Big(\frac{\widehat{N}_n(i,\lfloor nt\rfloor)}{n} - \rho^{(n)}(i,t)\Big),\ (i,t)\in\Sigma\times[0,1],$$

where $\widehat{N}_n(i,l) = \sum_{j=1}^{l} 1(X_{[j:n]}^{(n)} = i),\ l = 0,\dots,n$. For each $i \in \Sigma$ we introduce the i-th component process $\mathcal{G}_n(i,\cdot) = (\mathcal{G}_n(i,t))_{t\in[0,1]}$ which takes values in $D([0,1])$, that is the Skorohod space of right-continuous real-valued functions with existing left-hand limits defined on the unit interval $[0,1]$. We identify $\mathcal{G}_n$ with $(\mathcal{G}_n(1,\cdot),\dots,\mathcal{G}_n(m,\cdot))$ such that $\mathcal{G}_n$ takes values in $D([0,1])^m$. We equip $D([0,1])^m$ with the m-fold product topology such that we deal with a polish space. Let $C([0,1]) \subset D([0,1])$ be the subset of continuous functions. We say that a $D([0,1])^m$-valued random variable has continuous paths if it takes values in $C([0,1])^m$, i.e., if all component processes have continuous paths. We call a stochastic process $\mathcal{G} = (\mathcal{G}(i,t))_{(i,t)\in\Sigma\times[0,1]}$ *Gaussian process* if for all $h\in\mathbb{N}$ and all $(i_1,t_1),\dots,(i_h,t_h) \in \Sigma\times[0,1]$ the h-dimensional random vector $(\mathcal{G}(i_1,t_1),\dots,\mathcal{G}(i_h,t_h))$

has a h-dimensional multivariate normal distribution. Suitable structural assumptions on the sequence $\rho^{(n)}$, $n \in \mathbb{N}$, ensure that $\mathcal{G}_n$ converges in distribution to a centered Gaussian process with continuous paths. If not otherwise specified, all convergences are meant as $n \to \infty$. These are usual assumptions to ensure distributional convergence of test statistics under local alternatives, see, e.g., Section 6 and Assumption 3 in Gaigall [8], for instance. Local alternatives are a theoretical tool, e.g., for efficiency deliberations, in particular for the Pitman-efficiency, see Puri and Sen [17], or the Volume-efficiency, see Baringhaus and Gaigall [2]. In Section 3, we will verify these assumptions to obtain asymptotic properties of the test statistic of independence under local alternatives. Moreover, in Section 4 it is also seen that the assumptions are satisfied in the context of the bootstrap procedure considered there. Note that the assumptions include the common case of a fixed distribution independent of n. The assumptions are given as follows.

Assumption 1. *For each $n \in \mathbb{N}$ we have a map $\psi^{(n)} : \Sigma \times [0,1] \to \mathbb{R}$ such that*

$$\rho^{(n)}(i,t) = \rho(i,t) + \frac{\psi^{(n)}(i,t)}{\sqrt{n}}, \; (i,t) \in \Sigma \times [0,1], \tag{8}$$

and that

a) $\rho(i,t) = P(X = i, U \leqslant t)$ *for some bivariate random variable* (X,U) *with* $P(X \in \Sigma) = 1$ *and* U *real-valued with* $P(U \leqslant t) = t$ *for each* $t \in [0,1]$*,*
b) $t \mapsto \rho(i,t)$ $t \in [0,1]$*, is continuously differentiable for each* $i \in \Sigma$ *with derivative* $\rho'(i,t)$*,*
c) there is a continuous limit function $\psi : \Sigma \times [0,1] \to \mathbb{R}$ *such that* $\sup_{i,t} |\psi^{(n)}(i,t) - \psi(i,t)| \to 0$*.*

Under Assumption 1 the sequence of bivariate random vectors $(X_1^{(n)}, F^{(n)}(Y_1^{(n)}))$ converges in distribution to the random vector (X,U) and $\rho(i,t) = P(X = i, U \leqslant t)$. Note that under Assumption 1 the point-wise convergence of $\rho^{(n)}$ to ρ holds also uniformly over $\Sigma \times [0,1]$ since this space is compact and we assume all functions being continuous. The following is the main result of this section. We denote almost sure convergence by $\overset{a.s.}{\to}$, convergence in probability by $\overset{P}{\to}$ and distributional convergence by $\overset{d}{\to}$.

Theorem 2.1. *Under Assumption 1 it holds that $\mathcal{G}_n \xrightarrow{d} \mathcal{G}$, where $\mathcal{G}$ is a centered Gaussian process with continuous paths and covariance function*

$$\begin{aligned}\operatorname{Cov}\big(\mathcal{G}(i,t), \mathcal{G}(h,s)\big) = {} & 1(i = h)\rho(i, s \wedge t) - \rho(i,t)\rho(h,s) + \rho'(i,t)\rho'(h,s)\big[s \wedge t - ts\big] \\ & - \rho'(i,t)\big[\rho(h, s \wedge t) - t\rho(h,s)\big] - \rho'(h,s)\big[\rho(i, s \wedge t) - s\rho(i,t)\big],\end{aligned}$$

where $(i,t), (h,s) \in \Sigma \times [0,1]$*.*

A proof of Theorem 2.1 can be found in Appendix A.1.

3. Application: Independence test

We consider the testing problem of independence (3), treated on the basis of the test statistic (4). In fact, by splitting the integration in parts of length $1/n$ we obtain the following simple expression of the test statistic

$$T_n = \frac{1}{6n} - \frac{1}{2} - \frac{n}{3} + \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{\hat{N}_n(i,j)^2}{n\hat{N}_n(i,n)}, \tag{9}$$

useful for calculation purposes in practice. A simple consequence of our results in this section is that under the null hypothesis of independence $\mathcal{H}$ it holds the following convergence in distribution

$$T_n \xrightarrow{d} T = \sum_{k=1}^{\infty} \frac{W_k}{k^2\pi^2}, \tag{10}$$

where $W_k, k \in \mathbb{N}$, is a sequence of independent χ^2-distributed random variables with $m-1$ degrees of freedom. In particular, T is distribution free. For that reason, the related asymptotic test at significance level $\alpha \in (0,1)$, that is the test which rejects the null hypothesis $\mathcal{H}$ if and only if $T_n > c$, where c is the $(1-\alpha)$-quantile of T, is suitable for the treatment of the testing problem of independence. Critical values for selected m and significance levels α, obtained by Monte-Carlo simulation with $2 \cdot 10^6$ replications, where the infinite sum in (10) is truncated after 800 terms, are displayed in Table 1.

m \ α	0.1	0.05	0.01	0.005	0.001
2	0.3477	0.4621	0.7456	0.8718	1.1654
3	0.6065	0.7473	1.073	1.2148	1.5355
4	0.8407	0.9997	1.3591	1.513	1.8565
5	1.0632	1.2378	1.6202	1.7766	2.1444
10	2.0947	2.3237	2.811	3.0081	3.4342
15	3.0731	3.3431	3.9025	4.1243	4.6125
20	4.0281	4.3308	4.9506	5.1965	5.7387

Table 1. Critical values for the implementation of the independence test.

Remark 1. *Based on the full observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ it is possible to translate our testing problem of independence to the multi-sample testing problem of homogeneity in Kiefer [15] by grouping the Y-observations with respect to the X-observations. Among others, Kiefer [15] treat a Cramér-von Mises type test statistic. This approach leads to an asymptotic test equivalent to the independence test. For that reason, it is possible to obtain (10) from the results in Kiefer [15]. Nevertheless, it is not obvious at first sight that the Cramér-von Mises type test statistic introduced by Kiefer [15] is measurable with respect to the concomitants. We emphasize that our aim in this section is to use our* *main result Theorem 2.1 not only to prove (10) but also to obtain the distributional limit of T_n under local alternatives, hence our results can be seen as an extension of the result of Kiefer [15] to the case of local alternatives around the hypotheses of homogeneity.*

In what follows, we aim to prove distributional convergence under local alternatives for the test statistic T_n.

3.1. Local alternatives

Now we consider a sequence of local alternatives $\rho^{(n)}, n \in \mathbb{N}$, which satisfies the following structural assumption.

Assumption 2. *For each $n \in \mathbb{N}$ we have a map $\psi^{(n)} : \Sigma \times [0,1] \to \mathbb{R}$ as well as $p_1, \dots, p_m \in (0,1)$ with $\sum_{i=1}^m p_i = 1$ such that*

$$\rho^{(n)}(i,t) = t \cdot p_i + \frac{\psi^{(n)}(i,t)}{\sqrt{n}}, \ (i,t) \in \Sigma \times [0,1], \tag{11}$$

and that $\sup_{i \in \Sigma, t \in [0,1]} |\psi^{(n)}(i,t) - \psi(i,t)| \to 0$ with a continuous limit function $\psi : \Sigma \times [0,1] \to \mathbb{R}$.

Note that, under Assumption 2, the independence of $X_1^{(n)}$ and $Y_1^{(n)}$ is equivalent to $\psi^{(n)}(i,t) = t \cdot \psi^{(n)}(i,1)$ for all $(i,t) \in \Sigma \times [0,1]$. for that reason, we say that $\rho^{(n)}$ *approaches independence locally* iff the limiting function ψ fulfills $\psi(i,t) = t\psi(i,1)$ for all $(i,t) \in \Sigma \times [0,1]$.

Our aim is to prove the distributional convergence of T_n under Assumption 2 using Theorem 2.1. Note that Assumption 2 is stronger than Assumption 1, where the function $\rho(i,t) = p_i t$, $(i,t) \in \Sigma \times [0,1]$, is differentiable with constant derivative $\rho'(i,t) \equiv p_i$, $(i,t) \in \Sigma \times [0,1]$. Hence we deduce directly from Theorem 2.1 the following result.

Corollary 3.1. *Under Assumption 2 it is $\mathcal{G}_n \xrightarrow{d} \mathcal{G}$, where $\mathcal{G}$ is a centered Gaussian process with covariance function*

$$\operatorname{Cov}\big(\mathcal{G}(i,t), \mathcal{G}(h,s)\big) = \big[1(i=h)p_i - p_i p_h\big] \cdot (t \wedge s), \quad (i,t),(h,s) \in \Sigma \times [0,1].$$

Defining the stochastic process $\mathcal{U}_n = (\mathcal{U}_n(t))_{t \in [0,1]}$ by

$$\mathcal{U}_n(t) = \sum_{i=1}^m \frac{\big(\widehat{N}_n(i, \lfloor nt \rfloor) - t\widehat{N}_n(i,n)\big)^2}{\widehat{N}_n(i,n)}, \ t \in [0,1],$$

the test statistic can be rewritten as

$$T_n = \int_0^1 \mathcal{U}_n(t) dt.$$

We first prove a result about the distributional convergence for $\mathcal{U}_n$. Then, we use the Karhuen-Loève expansion to prove the distributional convergence of T_n. Thereby, Assumption 2 is supposed.

Note that the limiting random vector $(\mathcal{G}(1,1),\dots,\mathcal{G}(m,1))$, that is the stochastic process $\mathcal{G}$ evaluated at time $t=1$, which appears in Corollary 3.1 is similar to the limit in the classical CLT for multinomial distributions. Let

$$\Xi = (\xi(i,h))_{i,h=1,\dots,m}$$

be the telated covariance matrix, i.e. $\Xi(i,h) = 1(i=h)p_i - p_i p_h,\ i,h=1,\dots,m$, and define $\Lambda = \mathrm{diag}(p_1^{-1/2},\dots,p_m^{-1/2}) \cdot \Xi \cdot \mathrm{diag}(p_1^{-1/2},\dots,p_m^{-1/2})$, that is

$$\Lambda = (\lambda(i,h))_{i,h=1,\dots,m},\ \ \lambda(i,h) = 1(i=h) - \sqrt{p_i p_h},\ \ i,h=1,\dots,m.$$

Note that Λ is the covariance matrix of $(\mathcal{G}(1,1)/\sqrt{p_1},\dots,\mathcal{G}(m,1)/\sqrt{p_m})$, which is symmetric and positive semi-definite with rank $m-1$. The eigenvalues of Λ are $0,1,1,\dots,1$, hence there is a orthogonal matrix $A = (a(i,h))_{i,h=1,\dots,m}$ such that

$$A\Lambda A^T \ = \ \mathrm{diag}(0,1,1,\dots,1).$$

For each $i=1,\dots,m$ let

$$a_i = (a(i,1),\dots,a(i,m))^T$$

be the i-th column of A^T. The m vectors $a_1,\dots,a_m$ form an orthonormal basis of $\mathbb{R}^m$ and it is $\Lambda a_1 = 0$ and $\Lambda a_i = a_i$ for each $i=2,\dots,m$. Note that for the first eigenvector

$$a_1 = (\sqrt{p_1},\dots,\sqrt{p_m})^T$$

holds. For vectors $a,b\in\mathbb{R}^m$ we write $\langle a,b\rangle = \sum_{i=1}^m a_i b_i$ for the euclidean scalar product. Further we write

$$\psi(t) = \big(\psi(1,t),\dots,\psi(m,t)\big)^T,\ \ t\in[0,1],$$

where ψ is the limiting function from Assumption 2, and

$$\overline{\psi}(t) = \Big(\frac{\psi(1,t)}{\sqrt{p_1}},\dots,\frac{\psi(m,t)}{\sqrt{p_m}}\Big)^T,\ \ t\in[0,1].$$

As a application of Theorem 2.1 we obtain the following result. Proofs of the following statements can be found in Appendix A.2.

Theorem 3.1. *Under Assumption 2 it holds that*

$$\mathcal{U}_n \xrightarrow{d} \Big(\sum_{i=2}^m \big[\mathcal{B}_i(t) + \langle a_i, \overline{\psi}(t) - t\overline{\psi}(1)\rangle\big]^2\Big)_{t\in[0,1]}, \tag{12}$$

where $\mathcal{B}_2,\dots,\mathcal{B}_m$ are $m-1$ independent Brownian bridges. In the case where $\rho^{(n)}$ approaches independence locally the limiting distribution of $\mathcal{U}_n$ is that of the sum of $m-1$ independent squared Brownian bridges. In the latter case, the limiting distribution of $\mathcal{U}_n$ does not depend on the concrete values of $p_1,\dots,p_m$.

Using Theorem 3.1 yields the following statement.

Theorem 3.2. *Under Assumption 2 we have*

$$T_n \xrightarrow{d} T = \sum_{k=1}^{\infty} \sum_{i=2}^{m} \Big[\frac{Z_{k,i}}{k\pi} + c_{k,i} \Big]^2,$$

where $Z_{k,i}$, $k \in \mathbb{N}$, $i = 2, \ldots, m$, are independent standard normal distributed random variables and

$$c_{k,i} = \int_0^1 \langle a_i, \overline{\psi}(t) \rangle \sqrt{2} \sin(k\pi t) dt + (-1)^k \frac{\sqrt{2}}{k\pi} \langle a_i, \overline{\psi}(1) \rangle, \ k \in \mathbb{N}, \ i = 2, \ldots, m.$$

In the case that the map $t \mapsto \overline{\psi}(t)$, $t \in [0,1]$, is differentiable for each $i = 2, \ldots, m$ with (the vector of) derivatives $\overline{\psi}'(t)$, $t \in [0,1]$, it is

$$T_n \xrightarrow{d} T = \sum_{k=1}^{\infty} \sum_{i=2}^{m} \frac{[Z_{k,i} + d_{k,i}]^2}{k^2 \pi^2}$$

with

$$d_{k,i} = \int_0^1 \langle a_i, \overline{\psi}'(t) \rangle \sqrt{2} \cos(k\pi t) dt, \ k \in \mathbb{N}, \ i = 2, \ldots, m.$$

In the case that $\rho^{(n)}$ approaches independence locally we have

$$T_n \xrightarrow{d} T = \sum_{k=1}^{\infty} \frac{W_k}{k^2 \pi^2},$$

where W_k, $k \in \mathbb{N}$, are independent random variables, each with the same χ^2-distribution with $m-1$ degrees of freedom.

Let us summarize the consequences of Theorem 3.2 for the independence test.

Corollary 3.2. *The following properties of the independence test are valid.*

(i) *Under local alternatives of the form of Assumption 2 it is $P(T_n > c) \to \beta$ with $\beta \in [\alpha, 1]$, i.e., the test is asymptotically unbiased.*
(ii) *In addition to (i) it is $\beta = \alpha$ if and only if $\psi^{(n)}$ approaches independence locally. This holds in particular in the case $\psi^{(n)} \equiv 0$, i.e., under the null hypothesis $\mathcal{H}$. In other words, the test reaches the significance level exactly in the limit.*
(iii) *Under any fixed alternative $\mathcal{K}$ it is $P(T_n > c) \to 1$, i.e., the test is consistent.*

3.2. Simulations

We investigate size and power of our asymptotic independence test T_n with simulations under the null hypothesis of independence and under alternatives. The performance of T_n is compared with the recently proposed tests in Chatterjee [4] and Genest et al. [11]. The test in Chatterjee [4], denoted by C_n in what follows, shares several features to our test. For that reason this test is chosen as competitor. In detail, the test statistic is also distribution free in the limit under the null hypothesis of independence and use the asymptotic quantiles as critical values, C_n only depends on the concomitants as well, and the test is also consistent. Beyond that, the test is asymptotically standard normal distributed under the null hypothesis of independence. For the test in Genest et al. [11], denoted by S_n in what follows, it is seen in Genest et al. [11] that the test has much power, hence we included it in our simulation studies. The test is also consistent, but the test statistic is not distribution-free, even asymptotically or under the null hypothesis of independence. For that reason, a wild Bootstrap procedure is suggested to obtain critical values. In addition, the complexity of S_n is high which limits the applicability of the test for large sample sizes. We point out that the tests in Chatterjee [4] and Genest et al. [11] are applicable for random vectors (X_1, Y_1), where the components take values in the real line. To enable the application of the tests in our situation, we consider the special case that $\Sigma = \{1, \dots, m\}$ and note that our test T_n is applicable also for general categories Σ.

The simulations are based on a Monte-Carlo simulation with 10.000 replications. For the implementation of our test T_n, critical values are obtained from from Table 1. For the implementation of the test C_n, quantiles of the standard normal distribution are used as critical values. For the determination of critical values for the test S_n, we use a wild Bootstrap with 1.000 replications. Due to the complexity of this test we restrict the simulations for S_n to sample sizes less than 200. Table 2 shows our simulation results. The presented values are average rejection rates. We consider the case that the joint distribution of (X_1, Y_1) follows a mixed-normal model, i.e., the conditional distribution $Y_1|X_1 = i$ is a normal distribution with mean μ_i and standard deviation $\sigma_i > 0$, and $P(X_1 = i) = p_i > 0$, $i = 1, \dots, m$. We set $p = (p_1, \dots, p_m)$, $\mu = (\mu_1, \dots, \mu_m)$, and $\sigma = (\sigma_1, \dots, \sigma_m)$. Note that in this model X_1 and Y_1 are independent if and only if $\mu_1 = \cdots = \mu_m$ and $\sigma_1 = \cdots = \sigma_m$. For that reason, the alternative $\mathcal{K}$ is valid if μ or σ has entries which are not all the same.

For all tests T_n, C_n, and S_n, it is seen that the empirical size values tend to the significance level and that the empirical power values tend to one as the sample size increases. In this regard, the empirical results confirm our theoretical findings for the test T_n. For $m = 2$ and $m = 5$, all tests T_n, C_n, and S_n keep the significance level quite well. For $m = 20$, the test T_n does not keep the significance level quite well for small sample sizes in contrast to the tests C_n and S_n. For $m = 2$, all tests T_n, C_n, and S_n have much power, where the power of T_n is comparable with the power of S_n and much higher than the power of C_n. For $m = 5$, the tests T_n and S_n have power, where the power of T_n is clearly higher than the power of S_n, and C_n has little power. For $m = 20$, all tests

T_n, C_n, and S_n have power, where for small sample sizes the power of T_n is little and the power of C_n is higher than the power of S_n, for moderate sample sizes the power of S_n is higher than the power of C_n, and for large sample sizes the test T_n gains more power. The simulation results indicate that the sample size should be sufficiently large compared with the number of categories to ensure a suitable quality of the test T_n in terms of size and power. Taking the latter into account, the simulation results indicate that the test T_n is a serious competitor to concurring tests from the literature.

4. Application: Two-Sample test

In this section we consider the two-sample testing problem (5). Proofs can be found in Appendix A.3. Throughout this section we assume that both maps $t \mapsto \rho_1(i,t)$, $t \in [0,1]$, and $t \mapsto \rho_2(i,t)$, $t \in [0,1]$, are continuously differentiable for all $i \in \Sigma$. At first, we show that the two-sample testing problem is equivalent to the testing problem (6), which motivates the usage of the test statistic (7). For this purpose, we need the following general preparatory result.

Lemma 4.1. *Let $h, g : \mathbb{R} \to \mathbb{R}$ be two non-decreasing functions and let Y a real-valued random variable with $h(Y) \stackrel{d}{=} g(Y)$. Then it is $h(Y) = g(Y)$ almost surely.*

Finally the equivalence of the testing problems (5) and (6) follows from the following general result.

Lemma 4.2. *Let $(X_1, Y_1), (X_2, Y_2)$ be two bivariate random vectors where X_k takes values in an arbitrary measurable space and Y_k is real valued with a continuous distribution function F_k for each $k = 1, 2$. In this case the following two statements are equivalent.*

- *(i) There exists a non-decreasing function $g : \mathbb{R} \to \mathbb{R}$ such that $(X_1, Y_1) \stackrel{d}{=} (X_2, g(Y_2))$.*
- *(ii) $(X_1, F_1(Y_1)) \stackrel{d}{=} (X_2, F_2(Y_2))$.*

For each $k = 1, 2$ we define the process $\mathcal{G}_{k,n_k} = (\mathcal{G}_{k,n_k}(i,t))_{(i,t)\in\Sigma\times[0,1]}$ by

$$\mathcal{G}_{k,n_k}(i,t) = \sqrt{n_k}\Big(\frac{\widehat{N}_{k,n_k}(i, \lfloor n_k t\rfloor)}{n_k} - \rho_k(i,t)\Big),\ (i,t) \in \Sigma \times [0,1]. \tag{13}$$

Defining the stochastic process $\mathcal{F}_{n_1,n_2} = (\mathcal{F}_{n_1,n_2}(i,t))_{(i,t)\in\Sigma\times[0,1]}$, by

$$\mathcal{F}_{n_1,n_2}(i,t) = \sqrt{\frac{n_1 n_2}{n_1+n_2}}\Big(\frac{\widehat{N}_{1,n_1}(i, \lfloor n_1 t\rfloor)}{n_1} - \frac{\widehat{N}_{2,n_2}(i, \lfloor n_2 t\rfloor)}{n_2}\Big),\ (i,t) \in \Sigma \times [0,1], \tag{14}$$

and the map $\Psi : D[0,1]^m \to \mathbb{R}$ by

$$\Psi(g(i,t)_{(i,t)\in\Sigma\times[0,1]}) = \sup_{i\in\Sigma, t\in[0,1]} |g(i,t)|,$$

	$\mathcal{H}$						$\mathcal{K}$					
m	$m = 2$											
p	$(p_1, p_2) = (0.5, 0.5)$											
μ	$(\mu_1, \mu_2) = (0, 0)$						$(\mu_1, \mu_2) = (0, 1)$					
σ	$(\sigma_1, \sigma_2) = (1, 1)$											
Level	$\alpha = 0.1$			$\alpha = 0.05$			$\alpha = 0.1$			$\alpha = 0.05$		
Test	T_n	C_n	S_n	T_n	C_n	S_n	T_n	C_n	S_n	T_n	C_n	S_n
$n = 25$	.120	.109	.130	.058	.055	.073	.748	.397	.761	.627	.245	.660
$n = 50$	.109	.097	.115	.055	.052	.061	.951	.527	.953	.906	.430	.913
$n = 75$	.103	.090	.107	.051	.053	.055	.993	.65	.993	.983	.515	.984
$n = 100$	.099	.097	.101	.051	.045	.053	.998	.754	.998	.997	.635	.997
$n = 200$	.111	.094	.11	.055	.052	.057	1	.913	1	1	.856	1
$n = 500$	.104	.101	–	.053	.049	–	1	.998	–	1	.995	–
$n = 1000$	.098	.096	–	.050	.048	–	1	1	–	1	1	–
m	$m = 5$											
p	$(p_1, \ldots, p_5) = (0.05, 0.1, 0.1, 0.25, 0.5)$											
μ	$\mu_i = 0, i = 1, \ldots, 5$						$(\mu_1, \ldots, \mu_5) = (1, 0.5, -0.5, 0, 0)$					
σ	$\sigma_i = 1, i = 1, \ldots, 5$											
Level	$\alpha = 0.1$			$\alpha = 0.05$			$\alpha = 0.1$			$\alpha = 0.05$		
Test	T_n	C_n	S_n	T_n	C_n	S_n	T_n	C_n	S_n	T_n	C_n	S_n
$n = 25$	.060	.091	.123	.023	.044	.065	.165	.118	.159	.074	.061	.088
$n = 50$	.084	.095	.111	.033	.046	.059	.385	.127	.191	.242	.065	.107
$n = 75$	.101	.094	.106	.045	.045	.054	.561	.132	.227	.416	.071	.135
$n = 100$	.095	.100	.105	.044	.050	.052	.702	.141	.279	.577	.076	.169
$n = 200$	.094	.096	.106	.045	.046	.053	.948	.157	.474	.905	.086	.319
$n = 500$	.102	.091	–	.050	.046	–	.999	.199	–	.999	.117	–
$n = 1000$	.101	.097	–	.050	.048	–	1	.235	–	1	.135	–
m	$m = 20$											
p	$p_i = 0.05, i = 1, \ldots, 20$											
μ	$\mu_i = 0, i = 1, \ldots, 20$											
σ	$\sigma_i = 1, i = 1, \ldots, 20$						$\sigma_i = 0.5, i = 1, \ldots, 5$ $\sigma_i = 1, i = 6, \ldots, 15$ $\sigma_i = 1.5, i = 16, \ldots, 20$					
Level	$\alpha = 0.1$			$\alpha = 0.05$			$\alpha = 0.1$			$\alpha = 0.05$		
Test	T_n	C_n	S_n	T_n	C_n	S_n	T_n	C_n	S_n	T_n	C_n	S_n
$n = 25$	.0	.090	.119	.0	.045	.064	.0	.227	.204	.0	.133	.104
$n = 50$	.023	.099	.111	.004	.049	.059	.046	.337	.330	.013	.225	.172
$n = 75$	.061	.095	.102	.022	.044	.052	.127	.427	.518	.055	.297	.296
$n = 100$	.078	.101	.106	.029	.051	.055	.191	.504	.681	.093	.375	.443
$n = 200$	.100	.099	.105	.046	.052	.052	.390	.712	.969	.248	.584	.899
$n = 500$	.090	.096	–	.042	.048	–	.931	.948	–	.841	.900	–
$n = 1000$	.101	.099	–	.050	.048	–	1	.997	–	.999	.993	–

Table 2. Average rejection rates for our independence test T_n, the independence test C_n in Chatterjee [4], and the independence test S_n in Genest and Remillard [10].

the test statistic can be rewritten as

$$T_{n_1,n_2} = \Psi(\mathcal{F}_{n_1,n_2}) = \sup_{i\in\Sigma, t\in[0,1]} |\mathcal{F}_{n_1,n_2}(i,t)|. \tag{15}$$

If $\mathcal{G} = (\mathcal{G}(i,t))_{(i,t)\in\Sigma\times[0,1]}$ is a centered Gaussian process with continuous paths and covariance structure as given in Theorem 2.1 we write

$$\mathcal{G} \sim \mathrm{GP}(\rho) \tag{16}$$

and $\mathcal{G}_n \xrightarrow{d} \mathrm{GP}(\rho)$ if convergence in distribution towards a $\mathrm{GP}(\rho)$-distributed Gaussian process holds. Application of Theorem 2.1 yields the following result.

Corollary 4.1. *Under the null hypothesis $\mathcal{H}$ (i.e., $\rho_1 = \rho_2 = \rho$) and if $n_1, n_2 \to \infty$ such that $n_1/(n_1+n_2) \to \gamma \in (0,1)$ it is $\mathcal{F}_{n_1,n_2} \xrightarrow{d} \mathrm{GP}(\rho)$.*

In order to treat the two-sample testing problem with the test statistic T_{n_1,n_2}, suitable critical values are needed. As a consequence of Corollary 4.1 combined with the continuous mapping theorem, the test statistics T_{n_1,n_2} converges in distribution towards a random variable T with distribution $\mathrm{GP}(\rho)^{\Psi}$ under the null hypothesis $\mathcal{H}$ (i.e., if $\rho_1 = \rho_2 = \rho$). In general, $\mathrm{GP}(\rho)^{\Psi}$ depends on the unknown distribution $\rho = \rho_1 = \rho_2$ and hence is not distribution free. To resolve this problem and to obtain suitable critical values, a Bootstrap approach is suggested.

4.1. Bootstrap procedure

Let us define the pooled set of concomitants by

$$X_{n_1,n_2} = (X_{1,[1:n_1]}, \ldots, X_{1,[n_1:n_1]}, X_{2,[1:n_2]}, \ldots, X_{2,[n_2:n_2]}),$$

that is X_{n_1,n_2} takes values in $\Sigma^{n_1+n_2}$. We now explain the procedure for generating a suitable Bootstrap sample of X_{n_1,n_2}.

1. Independently from X_{n_1,n_2} generate independent random variables $B_{k,j}$, $U_{k,j}$, $j = 1,\ldots,n_k$, $k = 1,2$, such that each $U_{k,j}$ is uniformly distributed on $[0,1]$ and each $B_{k,j}$ satisfies

$$P(B_{k,j} = 1) = \frac{n_1}{n_1+n_2} \quad \text{and} \quad P(B_{k,j} = 2) = \frac{n_2}{n_1+n_2}.$$

2. For $k = 1,2$ and $j = 1,\ldots,n_k$ set $I_{k,j} = \lceil n_1 \cdot U_{k,j} \rceil$ and $J_{k,j} = \lceil n_2 \cdot U_{k,j} \rceil$ and define

$$\widehat{X}_{k,j} = \begin{cases} X_{1,[I_{k,j}:n_1]}, & \text{if } B_{k,j} = 1 \\ X_{2,[J_{k,j}:n_2]}, & \text{if } B_{k,j} = 2. \end{cases}$$

3. For $k = 1, 2$, define the concomitants $\hat{X}_{k,[j:n_k]}$, $j = 1, \ldots, n_k$, of the $\hat{X}_{k,j}$, $j = 1, \ldots, n_k$, with respect to $U_{k,j}$, $j = 1, \ldots, n_k$, and set
$$\hat{X}_{n_1,n_2} = (\hat{X}_{1,[1:n_1]}, \ldots, \hat{X}_{1,[n_1:n_1]}, \hat{X}_{2,[1:n_2]}, \ldots, \hat{X}_{2,[n_2:n_2]}).$$

Let $\hat{\mathcal{F}}_{n_1,n_2}$ be the stochastic process obtained by replacement of X_{n_1,n_2} with $\hat{X}_{n_1,n_2}$ in (14) and let $\hat{T}_{n_1,n_2} = \Psi(\hat{\mathcal{F}}_{n_1,n_2})$ be the associated statistic. Denote by K_{n_1,n_2} the (random) distribution function of $\hat{T}_{n_1,n_2}$ given X_{n_1,n_2}, that is
$$K_{n_1,n_2}(x) = P\big(\hat{T}_{n_1,n_2} \leqslant x | X_{n_1,n_2}\big), \; x \in \mathbb{R}, \tag{17}$$
and let $K^{-1}_{n_1,n_2}$ be the inverse quantile function of K_{n_1,n_2}. Given that $\alpha \in (0,1)$ is the significance level, we suggest to use the two-sample test which rejects the null hypothesis $\mathcal{H}$ if and only if $T_{n_1,n_2} > K^{-1}_{n_1,n_2}(1-\alpha)$. In general, $K^{-1}_{n_1,n_2}(1-\alpha)$ is not available in a closed form. For that reason, $K^{-1}_{n_1,n_2}(1-\alpha)$ has to be approximated by Monte-Carlo simulation. For this purpose, we generate $B \in \mathbb{N}$, where B is sufficiently large, independent Bootstrap samples $\hat{X}^{(b)}_{n_1,n_2}$, $b = 1, \ldots, B$, the associated statistics $\hat{T}^{(b)}_{n_1,n_2}$, $b = 1, \ldots, B$, and estimate K_{n_1,n_2} by
$$\hat{K}_{B,n_1,n_2}(x) = \frac{1}{B} \sum_{b=1}^{B} I\big(\hat{T}^{(b)}_{n_1,n_2} \leqslant x\big), \; x \in \mathbb{R}.$$

We state the main result of this section in advance.

Theorem 4.1. *If $n_1, n_2 \to \infty$ such that*
$$\frac{n_1}{n_1+n_2} \to \gamma \in (0,1) \quad \textit{and} \quad \frac{(1-\gamma)n_1 - \gamma n_2}{\sqrt{n_1+n_2}} \to \eta \in \mathbb{R}, \tag{18}$$
it follows that

(i) $P\Big(T_{n_1,n_2} > K^{-1}_{n_1,n_2}(1-\alpha)\Big) \longrightarrow \alpha$ *under the null hypothesis* $\mathcal{H}$,

(ii) $P\Big(T_{n_1,n_2} > K^{-1}_{n_1,n_2}(1-\alpha)\Big) \longrightarrow 1$ *under alternatives* $\mathcal{K}$.

For the proof of this result, some effort is needed. We present all the necessary interim statements and the final proof step here, the missing proof can be found in Appendix A.3. For the rest of the section we regard the sample sizes n_1 and n_2 as two functions $n_1, n_2 : \mathbb{N} \to \mathbb{N}$ with $\lim_{n\to\infty} n_1(n) = \infty$, $\lim_{n\to\infty} n_2(n) = \infty$ and
$$\lim_{n\to\infty} \frac{n_1(n)}{n_1(n)+n_2(n)} = \gamma \in (0,1) \quad \text{and} \quad \lim_{n\to\infty} \frac{(1-\gamma)n_1(n) - \gamma n_2(n)}{\sqrt{n_1(n)+n_2(n)}} = \eta \in \mathbb{R}. \tag{19}$$

All limits considered in the following are as $n \to \infty$. If we write n_k we always mean $n_k(n)$ for some n. Sometimes we refer to (n, n_1, n_2) in the sense that we understand n_k as $n_k(n)$.

Lemma 4.3. *Suppose that the null hypothesis $\mathcal{H}$ is true, where the underlying joint distribution is defined by*

$$\rho^{(n)}(i,t) = \rho(i,t) + \frac{\psi^{(n)}(i,t)}{\sqrt{n_1}}, \ (i,t) \in \Sigma \times [0,1], \tag{20}$$

with a map $\psi^{(n)} : \Sigma \times [0,1] \to \mathbb{R}$ and that

a) $\rho(i,t) = P(X = i, U \leqslant t)$ *where* $P(X \in \Sigma) = 1$ *and* $P(U \leqslant t) = t$ *for each* $t \in [0,1]$,
b) $t \mapsto \rho(i,t) \quad t \in [0,1]$, *is continuously differentiable for each* $i \in \Sigma$ *with derivative* $\rho'(i,t)$,
c) *there is a continuous limit function* $\psi : \Sigma \times [0,1] \to \mathbb{R}$ *such that* $\sup_{i\in\Sigma, t\in[0,1]} |\psi^{(n)}(i,t) - \psi(i,t)| \to 0$.

Then, it holds that $\mathcal{F}_{n_1,n_2} \xrightarrow{d} \mathrm{GP}(\rho)$.

In what follows we work in a particular probability space for $X_{n_1,n_2} = X_{n_1(n),n_2(n)}$, $n \in \mathbb{N}$. For this purpose, we specify this special construction and prove its existence next.

Lemma 4.4. *There exists a probability space $(\Omega, \mathcal{A}, P)$ with random variables X_{n_1,n_2}, $n \in \mathbb{N}$, such that the following properties hold.*

(i) $\lim_n \sup_{i\in\Sigma, t\in[0,1]} |\widehat{N}_{k,n_k}(i, \lfloor n_k t \rfloor)/n_k - \rho_k(i,t)| = 0$ *almost surely for* $k = 1,2$.
(ii) *There exist independent centered Gaussian processes with continuous paths* $\mathcal{G}' \sim \mathrm{GP}(\rho_1)$ *and* $\mathcal{G}'' \sim \mathrm{GP}(\rho_2)$ *such that* $\mathcal{G}_{1,n_1} \to \mathcal{G}'$ *and* $\mathcal{G}_{2,n_2} \to \mathcal{G}''$ *almost surely.*
(iii) *There exists a random variable U, uniformly distributed on $[0,1]$, such that U, X_{n_1,n_2}, $n \in \mathbb{N}$, $\mathcal{G}', \mathcal{G}''$, are independent.*

Using (30) we obtain the following result.

Corollary 4.2. *On the probability space constructed in Lemma 4.4 and under $\mathcal{H}$, i.e., in the case of $\rho_1 = \rho_2 = \rho$, the process $\mathcal{F}_{n_1,n_2}$ converges almost surely towards the centered Gaussian process with continuous paths $\mathcal{G} = \sqrt{1-\gamma}\mathcal{G}' - \sqrt{\gamma}\mathcal{G}'' \sim \mathrm{GP}(\rho)$.*

On the probability in Lemma 4.4 we can use the independent randomization variable U to construct countable many Bootstrap samples of X_{n_1,n_2} which are conditional independent given X_{n_1,n_2}. We write $\widehat{X}_{n_1,n_2}$ for a Bootstrap sample of X_{n_1,n_2} and $\widehat{\mathcal{F}}_{n_1,n_2}$ for the stochastic process (14) on the basis of $\widehat{X}_{n_1,n_2}$ instead of X_{n_1,n_2}. We denote by $\widehat{X}^{(b)}_{n_1,n_2}$, $b \in \mathbb{N}$, a (conditional independent given X_{n_1,n_2}) sequence of Bootstrap samples, and we define by $\widehat{\mathcal{F}}^{(b)}_{n_1,n_2}$, $b \in \mathbb{N}$, the related stochastic process (14).
The key proposition used in the proof of Theorem 4.1 is the following.

Proposition 4.1. *On the probability space in Lemma 4.4 it holds that*

$$\widehat{\mathcal{F}}_{n_1,n_2} \xrightarrow{d} \mathrm{GP}(\gamma\rho_1 + (1\text{-}\gamma)\rho_2) \ \textit{ given } X_{n_1,n_2} \textit{ almost surely.}$$

If in addition $\mathcal{H}$ holds, i.e., if $\rho_1 = \rho_2 = \rho$, then $\widehat{\mathcal{F}}_{n_1,n_2} \xrightarrow{d} \mathrm{GP}(\rho)$ given X_{n_1,n_2} almost surely.

Because the map Ψ is continuous and almost sure weak convergence of random probability measures always implies weak convergence of expectations, we obtain the following corollary from the previous proposition.

Corollary 4.3. *On the probability space in Lemma 4.4 it holds that*

$$\widehat{T}_{n_1,n_2} \xrightarrow{d} \mathrm{GP}(\gamma\rho_1 + (1\text{-}\gamma)\rho_2)^{\Psi} \text{ given } X_{n_1,n_2} \text{ almost surely.}$$

On any probability space it is

$$\widehat{\mathcal{F}}_{n_1,n_2} \xrightarrow{d} \mathrm{GP}(\gamma\rho_1 + (1\text{-}\gamma)\rho_2) \quad \text{and} \quad \widehat{T}_{n_1,n_2} \xrightarrow{d} \mathrm{GP}(\gamma\rho_1 + (1\text{-}\gamma)\rho_2)^{\Psi}. \tag{21}$$

In order to prove Theorem 4.1 we proceed with the following lemma.

Lemma 4.5. *Under $\mathcal{H}$, i.e., if $\rho_1 = \rho_2 = \rho$, it holds that*

$$\big(\mathcal{F}_{n_1,n_2}, \widehat{\mathcal{F}}^{(1)}_{n_1,n_2}, \widehat{\mathcal{F}}^{(2)}_{n_1,n_2}\big) \xrightarrow{d} \mathrm{GP}(\rho) \otimes \mathrm{GP}(\rho) \otimes \mathrm{GP}(\rho).$$

Applying the map Ψ in the sense that $(g, g_1, g_2) \mapsto (\Psi(g), \Psi(g_1), \Psi(g_2))$ yields a continuous map. Hence as a direct consequence of the previous lemma we obtain the following result.

Corollary 4.4. *Under $\mathcal{H}$, i.e., if $\rho_1 = \rho_2 = \rho$, it holds that*

$$\big(T_{n_1,n_2}, \widehat{T}^{(1)}_{n_1,n_2}, \widehat{T}^{(2)}_{n_1,n_2}\big) \xrightarrow{d} \mathrm{GP}(\rho)^{\Psi} \otimes \mathrm{GP}(\rho)^{\Psi} \otimes \mathrm{GP}(\rho)^{\Psi}.$$

The following lemma is needed.

Lemma 4.6. *Let $T = \Psi(\mathcal{G}) = \sup_{i\in\Sigma, t\in[0,1]} |\mathcal{G}(i,t)|$ with $\mathcal{G} \sim \mathrm{GP}(\rho)$. Then the map $x \mapsto P(T \leqslant x)$, $x \in \mathbb{R}$, is continuous and strictly increasing on $(0,\infty)$.*

Finally we prove the main result in this section .

Proof of Theorem 4.1. For (i) we apply Lemma 4.2 in Bücher and Kojadinovic [3]. For this purpose, it is sufficient to show that

$$(T_{n_1,n_2}, \widehat{T}^{(1)}_{n_1,n_2}, \widehat{T}^{(2)}_{n_1,n_2}) \xrightarrow{d} (T, T^{(1)}, T^{(2)})$$

such that

(a) $T, T^{(1)}, T^{(2)}$ are independent and identically distributed under $\mathcal{H}$,
(b) T has a continuous distribution function under $\mathcal{H}$.

For ii) we show that

(c) $\widehat{T}_{n_1,n_2}$ converges in distribution under $\mathcal{K}$.

Well, (a) is Corollary 4.4, (b) is Lemma 4.6, and (c) follows from Corollary 4.3. □

4.2. Simulations

In what folows, size and power of the two-sample test is investigated via simulations. The simulations are based on a Monte-Carlo simulation with 3.000 replications, where the Monte-Carlo approximation of the Bootstrap critical values based on 500 replications in each case. In the simulations, the joint distribution of $(X_{k,1}, Y_{k,1})$ follow a mixed-normal model in the simulations, i.e., the conditional distribution $Y_{k,1}|X_{k,1} = i$ is a normal distribution with mean $\mu_{k,i}$ and standard deviation $\sigma_{k,i} > 0$, and $P(X_{k,1} = i) = p_{k,i} > 0$, $i = 1, \ldots, m$, $k = 1, 2$. We set $p_k = (p_{k,1}, \ldots, p_{k,m})$, $\mu_k = (\mu_{k,1}, \ldots, \mu_{k,m})$, and $\sigma_k = (\sigma_{k,1}, \ldots, \sigma_{k,m})$. Empirical size and power values of the he two-sample test are displayed in Table 3 and Table 4, respectively. It is seen that the empirical size values of the test tent to the significance level and that and that the empirical power values of the test tend to one as the sample size increases. The empirical results confirm the theoretical findings.

m		$m = 2$		$m = 3$		$m = 5$	
$p_1 = p_2$		$p_1 = p_2 = (0.5, 0.5)$		$p_1 = p_2 = (0.5, 0.25, 0.25)$		$p_1 = p_2 = (0.4, 0.2, 0.2, 0.1, 0.1)$	
$\mu_1 = \mu_2$		$\mu_1 = \mu_2 = (0, 0)$		$\mu_1 = \mu_2 = (0, -1, 1)$		$\mu_1 = \mu_2 = (0, 0, 0, 0, 0)$	
$\sigma_1 = \sigma_2$		$\sigma_1 = \sigma_2 = (1, 1)$		$\sigma_1 = \sigma_2 = (1, 1, 1)$		$\sigma_1 = \sigma_2 = (1, 2, 3, 4, 5)$	
n_1	n_2	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$
25	25	0.086	0.050	0.074	0.044	0.069	0.041
50	25	0.091	0.040	0.101	0.049	0.097	0.044
50	50	0.090	0.041	0.088	0.038	0.071	0.033
100	50	0.094	0.044	0.086	0.043	0.083	0.037
100	100	0.081	0.036	0.093	0.046	0.072	0.037
500	100	0.096	0.052	0.102	0.054	0.093	0.049
1000	200	0.105	0.057	0.089	0.049	0.101	0.052
1000	1000	0.090	0.051	0.093	0.045	0.103	0.056

Table 3. Empirical size values for the two-sample test.

Acknowledgment

The authors wish to thank the two Referees for very helpful comments and suggestions which improved the paper significantly. Special thanks goes to the Associate Editor who drew the attention of the authors to the work of Chatterjee [4].

m		$m=2$		$m=3$		$m=5$	
p_1		$p_1=(0.5,0.5)$		$p_1=(0.5,0.25,0.25)$		$p_1=(0.4,0.2,0.2,0.1,0.1)$	
μ_1		$\mu_1=(0,0)$		$\mu_1=(0,-1,1)$		$\mu_1=(0,0,0,0,0)$	
σ_1		$\sigma_1=(1,1)$		$\sigma_1=(1,1,1)$		$\sigma_1=(1,2,3,4,5)$	
p_2		$p_2=(0.3,0.7)$		$p_2=(0.5,0.25,0.25)$		$p_2=(0.4,0.2,0.2,0.1,0.1)$	
μ_2		$\mu_2=(0,0)$		$\mu_2=(0,0,0)$		$\mu_2=(0,0,0,0,0)$	
σ_2		$\sigma_2=(1,1)$		$\sigma_2=(1,1,1)$		$\sigma_2=(1,1,1,1,1)$	
n_1	n_2	$\alpha=0.1$	$\alpha=0.05$	$\alpha=0.1$	$\alpha=0.05$	$\alpha=0.1$	$\alpha=0.05$
25	25	0.359	0.263	0.114	0.062	0.066	0.040
50	25	0.483	0.341	0.140	0.065	0.115	0.052
50	50	0.613	0.490	0.167	0.080	0.108	0.056
100	50	0.728	0.621	0.215	0.115	0.132	0.074
100	100	0.863	0.777	0.316	0.175	0.167	0.092
500	100	0.970	0.952	0.572	0.368	0.249	0.140
1000	200	1.000	0.999	0.940	0.826	0.493	0.300
1000	1000	1.000	1.000	1.000	1.000	0.998	0.987

Table 4. Empirical power values for the two-sample test.

Appendix: Proofs

A.1. Proofs for Section 2

Proof of Theorem 2.1. Suppose without loss of generality that $Y_1^{(n)}$ has the uniform distribution on $[0,1]$, i.e., $F^{(n)}(t)=t$ for all $t\in[0,1]$. In particular, $\rho^{(n)}(i,t)=P(X_1^{(n)}=i,Y_1^{(n)}\leqslant t)$ for all $(i,t)\in\Sigma\times[0,1]$. A natural and consistent estimator for $\rho^{(n)}$ based on the full observations $(X_j^{(n)},Y_j^{(n)})$, $j=1,\ldots,n$, is given by

$$\hat{\rho}_n^{(n)}(i,t)=\frac{1}{n}\sum_{j=1}^{n}1(X_j^{(n)}=i,Y_j^{(n)}\leqslant t),\ \ (i,t)\in\Sigma\times[0,1].$$

Consider the *empirical process based on the full observations*

$$\mathcal{E}_n=(\mathcal{E}_n(i,t))_{(i,t)\in\Sigma\times[0,1]}\quad\text{with}\quad \mathcal{E}_n(i,t)=\sqrt{n}\big(\hat{\rho}_n^{(n)}(i,t)-\rho^{(n)}(i,t)\big),\ \ (i,t)\in\Sigma\times[0,1].$$

Classic empirical process theory for triangular arrays yields that

$$\mathcal{E}_n\xrightarrow{d}\mathcal{E},$$

where $\mathcal{E}=(\mathcal{E}(i,t))_{(i,t)\in\Sigma\times[0,1]}$ is a centered Gaussian process with continuous paths and covariance function

$$\mathrm{Cov}(\mathcal{E}(i,t),\mathcal{E}(h,s))=1(i=h)\rho(i,s\wedge t)-\rho(i,t)\rho(h,s),\ \ (i,t),(h,s)\in\Sigma\times[0,1],$$

see Ziegler [22]. Like before we write $\mathcal{E}_n(i,\cdot), \mathcal{E}(i,\cdot)$ for coordinate processes. Let $\widehat{F}_n$ be the empirical distribution function of $Y_j^{(n)}$, $j = 1, \ldots, n$. It is

$$\Big(\sum_{i=1}^{m} \mathcal{E}_n(i,t)\Big)_{t\in[0,1]} = \Big(\sqrt{n}\big(\widehat{F}_n(t) - t\big)\Big)_{t\in[0,1]} \xrightarrow{d} \mathcal{B},$$

where $\mathcal{B} = (\mathcal{B}(t))_{t\in[0,1]}$ is a Brownian bridge. It follows from the continuous mapping theorem that the limiting sum process $\sum_{i=1}^{m} \mathcal{E}(i,\cdot)$ is a Brownian bridge.
Let $Y_{[1:n]}^{(n)} < \cdots < Y_{[n:n]}^{(n)}$ be the order statistic of $Y_1^{(n)}, \ldots, Y_n^{(n)}$. For each $t \in [0,1]$ we define the empirical t-quantile of $Y_1, \ldots, Y_n$ as

$$\widehat{Q}_n(t) = Y_{[\lfloor nt\rfloor:n]}^{(n)},$$

where we set $Y_{[0:n]}^{(n)} = 0$. In the triangular situation, both the empirical distribution function $\widehat{F}_n$ and the process of empirical quantiles $\widehat{Q}_n$ converge uniformly to the identity in probability. As noted above, $\sqrt{n}(\widehat{F}_n(t) - t)_{t\in[0,1]} \xrightarrow{d} \mathcal{B}$, where $\mathcal{B}$ is a Brownian bridge. The so-called *Bahadur-Kiefer process*, that is the sum process $\big(\sqrt{n}(\widehat{F}_n(t) - t) + \sqrt{n}(\widehat{Q}_n(t) - t)\big)_{t\in[0,1]}$, converges uniformly to zero in probability, i.e.,

$$\sup_{t\in[0,1]} \left|\sqrt{n}\big(\widehat{F}_n(t) - t\big) + \sqrt{n}\big(\widehat{Q}_n(t) - t\big)\right| \xrightarrow{P} 0, \tag{22}$$

see Bahadur [1] and chapter 15 and chapter 18 Shorack and Wellner [19]. As a consequence of (22) we obtain the joint distributional convergence

$$\Big(\ \sqrt{n}(\widehat{F}_n(t) - t)_{t\in[0,1]},\ \sqrt{n}(\widehat{Q}_n(t) - t)_{t\in[0,1]}\ \Big) \ \xrightarrow{d}\ (\mathcal{B}, -\mathcal{B}),$$

where $\mathcal{B}$ is a Brownian Bridge.

The proof of Theorem 2.1 is now based on the following representation of $\mathcal{G}_n$ involving the empirical process based on the full observations $\mathcal{E}_n$ and the process of empirical quantiles $\widehat{Q}_n$. First note that for each $t \in [0,1]$, and $j = 1, \ldots, n$, it holds the equivalence

$$\widehat{F}_n(Y_j^{(n)}) \leqslant t \ \Leftrightarrow\ Y_j^{(n)} \leqslant \widehat{Q}_n(t).$$

This implies

$$\begin{aligned}\frac{\widehat{N}_n(i,\lfloor nt\rfloor)}{n} &= \frac{1}{n}\sum_{j=1}^{n} 1(X_j^{(n)} = i, \widehat{F}_n(Y_j^{(n)}) \leqslant t)\\ &= \frac{1}{n}\sum_{j=1}^{n} 1(X_j^{(n)} = i, Y_j^{(n)} \leqslant \widehat{Q}_n(t))\\ &= \rho_n^{(n)}(i, \widehat{Q}_n(t)),\ \ (i,t) \in \Sigma \times [0,1],\end{aligned}$$

and hence

$$\mathcal{G}_n(i,t) = \mathcal{E}_n(i,\widehat{Q}_n(t)) + \sqrt{n}\big[\rho^{(n)}(i,\widehat{Q}_n(t)) - \rho^{(n)}(i,t)\big],\ (i,t) \in \Sigma \times [0,1]. \tag{23}$$

Using Assumption 1 yields for each $i \in \Sigma, t \in [0,1]$

$$\mathcal{G}_n(i,t) = \mathcal{E}_n(i,\widehat{Q}_n(t)) + \sqrt{n}\big[\rho(i,\widehat{Q}_n(t)) - \rho(i,t)\big] + \big[\psi^{(n)}(i,\widehat{Q}_n(t)) - \psi^{(n)}(i,t)\big]. \tag{24}$$

We first consider the third summand on the right hand side of (24). By assumption is is $\sup_{(i,t)\in\Sigma\times[0,1]} |\psi^{(n)}(i,t) - \psi(i,t)| \to 0$ and ψ is continuous. Because $\sup_{t\in[0,1]} |\widehat{Q}_n(t) - t| \xrightarrow{P} 0$ we have

$$\sup\nolimits_{(i,t)\in\Sigma\times[0,1]} \big|\psi^{(n)}(i,\widehat{Q}_n(t)) - \psi^{(n)}(i,t)\big| \xrightarrow{P} 0. \tag{25}$$

We now investigate the first summand on the right hand side of (24) and show that

$$\sup\nolimits_{(i,t)\in\Sigma\times[0,1]} \big|\mathcal{E}_n(i,\widehat{Q}_n(t)) - \mathcal{E}_n(i,t)\big| \xrightarrow{P} 0. \tag{26}$$

We show (26) using the Skorokhod's representation theorem. First note that $\widehat{Q}_n$ is a $D([0,1])$-valued random variable witch converges in distribution to the constant identity function id. Since this limit is constant, $\mathcal{E}_n \xrightarrow{d} \mathcal{E}$ implies the joint convergence $(\mathcal{E}_n, \widehat{Q}_n) \xrightarrow{d} (\mathcal{E}, \mathrm{id})$. Now using Skorokhod's representation theorem we assume that the joint convergence is almost surely. Note that convergence takes place in the topological product space $D([0,1])^m \times D([0,1])$ and that the limit $(\mathcal{E}, \mathrm{id})$ is concentrated on the subspace of continuous functions $C([0,1])^m \times C([0,1])$. It is a well known fact that a sequence of functions $e_n \in D([0,1])$ converges to some continuous function $e \in C([0,1])$ if and only if $\sup_{t\in[0,1]} |e_n(t) - e(t)| \to 0$. This fact can be extended for the product spaces $C([0,1])^m$ and $D([0,1])^m$ and hence $\sup_{(i,t)\in\Sigma\times[0,1]} |\mathcal{E}_n(i,t) - \mathcal{E}(i,t)| \xrightarrow{a.s.} 0$ holds. Triangle inequality yields

$$\begin{aligned}
&\sup_{(i,t)\in\Sigma\times[0,1]} \big|\mathcal{E}_n(i,\widehat{Q}_n(t)) - \mathcal{E}_n(i,t)\big| \\
&\qquad\leqslant \sup_{(i,t)\in\Sigma\times[0,1]} \big|\mathcal{E}_n(i,\widehat{Q}_n(t)) - \mathcal{E}(i,\widehat{Q}_n(t))\big| + \sup_{(i,t)\in\Sigma\times[0,1]} \big|\mathcal{E}(i,\widehat{Q}_n(t)) - \mathcal{E}(i,t)\big|.
\end{aligned}$$

The first summand in the upper bound is bounded by $\sup_{(i,t)\in\Sigma\times[0,1]} \big|\mathcal{E}_n(i,t) - \mathcal{E}(i,t)\big|$ and hence converges to zero almost surely. The second summand in the upper bound converges to zero almost surely as well since $\mathcal{E}$ has continuous paths and $\sup_{t\in[0,1]} |\widehat{Q}_n(t) - t| \xrightarrow{a.s.} 0$. Hence (26) holds.

Now we investigate the second summand on the right hand side of (24). By Assumption 1 the map $t \mapsto \rho(i,t)$, $t \in [0,1]$, is continuously differentiable with derivative $\rho'(i,t)$ for each $i \in \Sigma$. We use a first order Taylor expansion and obtain

$$\rho(i,s) = \rho(i,t) + \rho'(i,t)(s-t) + r(i,t,s)(s-t),\ i \in \Sigma,\ t,s \in [0,1]$$

with

$$r(i,t,s) = \frac{\rho(i,s)-\rho(i,t)}{s-t} - \rho'(i,t),\ i \in \Sigma,\ t,s \in [0,1],$$

where we set $r(i,t,t) = 0$. With $s = \hat{Q}_n(t)$, $t \in [0,1]$, we obtain

$$\sqrt{n}\big(\rho(i,\hat{Q}_n(t)) - \rho(i,t)\big) = \big[\rho'(i,t) + r(i,t,\hat{Q}_n(t))\big] \cdot \big[\sqrt{n}(\hat{Q}_n(t) - t)\big],\ t \in [0,1]. \quad (27)$$

Since $t \mapsto \rho'(i,t)$, $t \in [0,1]$ is continuous for each $i \in \Sigma$ by assumption, $(t,s) \mapsto r(i,t,s)$, $t,s \in [0,1]$, is continuous for each $i \in \Sigma$ and hence $\sup_{t\in[0,1]} |\hat{Q}_n(t) - t| \xrightarrow{P} 0$ implies $\sup_{(i,t)\in\Sigma\times[0,1]} |r(i,t,\hat{Q}_n(t))| \xrightarrow{P} 0$ because $r(i,t,t) = 0$ for each $(i,t) \in \Sigma \times [0,1]$. We obtain

$$\sup_{(i,t)\in\Sigma\times[0,1]} \Big|\ \sqrt{n}\big(\rho(i,\hat{Q}_n(t)) - \rho(i,t)\big)\ -\ \rho'(i,t)\sqrt{n}\big(\hat{Q}_n(t) - t\big)\ \Big| \xrightarrow{P} 0.$$

Using (22) and $\sqrt{n}(\hat{F}_n(t) - t) = \sum_{i=1}^m \mathcal{E}_n(i,t)$, $t \in [0,1]$, we can replace $\sqrt{n}\big(\hat{Q}_n(t) - t\big)$ with $-\sum_{i=1}^m \mathcal{E}_n(i,t)$ in the previous equation for each $t \in [0,1]$ and obtain

$$\sup_{(i,t)\in\Sigma\times[0,1]} \Big|\ \sqrt{n}\big(\rho(i,\hat{Q}_n(t)) - \rho(i,t)\big)\ -\ \Big[-\rho'(i,t)\sum_{h=1}^m \mathcal{E}_n(h,t)\Big]\ \Big| \xrightarrow{P} 0. \quad (28)$$

Combining (24), (25), (26), and (28) yields

$$\sup_{(i,t)\in\Sigma\times[0,1]} \Big|\mathcal{G}_n(i,t) - \big[\mathcal{E}_n(i,t) - \rho'(i,t)\sum_{h=1}^m \mathcal{E}_n(h,t)\big]\Big| \xrightarrow{P} 0. \quad (29)$$

Because $t \mapsto \rho'(i,t)$, $t \in [0,1]$, is continuous for all $i \in \Sigma$ by assumption, the map $\Phi : D([0,1])^m \to D([0,1])^m$, defined by.

$$\Phi(e_1,\ldots,e_m) = (g_1,\ldots,g_m) \ \text{ where }\ g_i(t) = e_i(t) - \rho'(i,t)\sum_{h=1}^m e_h(t),\ t \in [0,1],$$

is continuous and hence $\mathcal{E}_n \xrightarrow{d} \mathcal{E}$ implies $\Phi(\mathcal{E}_n) \xrightarrow{d} \Phi(\mathcal{E})$. Combining this with (29) yields

$$\mathcal{G}_n \xrightarrow{d} \mathcal{G} = \Phi(\mathcal{E}) = \Big(\mathcal{E}(i,t) - \rho'(i,t)\sum_{h=1}^m \mathcal{E}(h,t)\Big)_{(i,t)\in\Sigma\times[0,1]}.$$

It is obvious that $\mathcal{G}$ has continuous paths. Moreover, the finite dimensional distributions of $\mathcal{G}$ are linear transformations of the finite dimensional distributions of $\mathcal{E}$ which are centered Gaussian, hence $\mathcal{G}$ is a centered Gaussian process. Simple calculation shows that the covariance function of $\mathcal{G}$ is as claimed in the theorem. □

A.2. Proofs for Section 3

Proof of Theorem 3.1. From Assumption 2 it follows that for each $i \in \Sigma, t \in [0,1]$

$$\sqrt{n}\Big(\frac{\widehat{N}_n(i,\lfloor nt \rfloor)}{n} - t \cdot \frac{\widehat{N}_n(i,n)}{n}\Big) = \Big(\mathcal{G}_n(i,t) + \psi^{(n)}(i,t)\Big) - t \cdot \Big(\mathcal{G}_n(i,1) + \psi^{(n)}(i,1)\Big).$$

Combining the distributional convergence $\mathcal{G}_n \xrightarrow{d} \mathcal{G}$ from Theorem 2.1, the uniformly convergence $\psi^{(n)} \to \psi$ and the almost sure convergence $\sqrt{n/\widehat{N}_n(i,n)} \to 1/\sqrt{p_i}$, $i = 1, \dots, m$, yields

$$\Big(\frac{\widehat{N}_n(i,\lfloor nt \rfloor) - t \cdot \widehat{N}_n(i,n)}{\sqrt{\widehat{N}_n(i,n)}}\Big)_{(i,t)\in\Sigma\times[0,1]} \xrightarrow{d} \Big(\frac{\mathcal{G}(i,t)}{\sqrt{p_i}} + \overline{\psi}(i,t) - t \cdot \Big(\frac{\mathcal{G}(i,1)}{\sqrt{p_i}} + \overline{\psi}(i,1)\Big)\Big)_{(i,t)\in\Sigma\times[0,1]}.$$

Let $\mathcal{W}_2, \dots, \mathcal{W}_m$ be independent Brownian Motions. Basic calculations show that

$$\Big(\frac{\mathcal{G}(1,\cdot)}{\sqrt{p_1}}, \dots, \frac{\mathcal{G}(m,\cdot)}{\sqrt{p_m}}\Big) \stackrel{d}{=} (0, \mathcal{W}_2, \dots, \mathcal{W}_m) \cdot A.$$

Because $\sum_{i=1}^m \psi^{(n)}(i,t) = 0$ for all $t \in [0,1]$ it is $\sum_i^m \psi(i,t) = 0$ for all $t \in [0,1]$ and hence with $\eta_i(t) = \langle a_i, \bar{\psi}(t) \rangle$ for $i = 2, \dots, m$ and $t \in [0,1]$ it follows that

$$\overline{\psi}(t) = \Big(\frac{\psi(1,t)}{\sqrt{p_1}}, \dots, \frac{\psi(m,t)}{\sqrt{p_m}}\Big) = (0, \eta_2(t), \dots, \eta_m(t)) \cdot A, \ t \in [0,1].$$

The orthogonality of A yields

$$\begin{aligned}\mathcal{U}_n &= \Big(\sum_{i=1}^m \Big[\frac{\widehat{N}_n(i,\lfloor nt \rfloor) - t \cdot \widehat{N}_n(i,n)}{\sqrt{\widehat{N}_n(i,n)}}\Big]^2\Big)_{t\in[0,1]} \\ &\xrightarrow{d} \Big(\sum_{i=2}^m \big[\mathcal{W}_i(t) - t\mathcal{W}_i(1) + \langle a_i, \overline{\psi}(t) - t\overline{\psi}(1)\rangle\big]^2\Big)_{t\in[0,1]}.\end{aligned}$$

The result follows since $\mathcal{B}_i(t) = \mathcal{W}_i(t) - t\mathcal{W}_i(1)$, $t \in [0,1]$, is a Brownian bridge. □

Proof of Theorem 3.2. Theorem 3.1 together with continuous mapping theorem yields $T_n \xrightarrow{d} T$, where

$$T = \sum_{i=2}^m \int_0^1 \big[\mathcal{B}_i(t) + \langle a_i, \overline{\psi}(t) - t\overline{\psi}(1)\rangle\big]^2 dt$$

with independent Brownian bridges $\mathcal{B}_2, \dots, \mathcal{B}_m$. Te distribution of T can be expressed using the Karhuen-Loève expansion. For this purpose, consider the Hilbert space $L^2 =$

$L^2([0,1])$ with inner product $\langle f,g\rangle_{L^2} = \int_0^1 f(t)g(t)dt,\ f,g \in L^2$. An orthonormal basis is given by $e_k(t) = \sqrt{2}\sin(k\pi t),\ t \in [0,1],\ k \in \mathbb{N}$. The well-known Karhuen-Loève expansion for Brownian bridges yields

$$\mathcal{B}_i \overset{d}{=} \sum_{k=1}^{\infty} \frac{Z_{k,i}}{k\pi} e_k,$$

where $Z_{k,i},\ k \in \mathbb{N},\ i = 1,\dots,m$, are independent standard normal distributed random variables. The map $t \mapsto \bar{\psi}(i,t),\ t \in [0,1]$, is continuous for each $i = 1,\dots,m$ by assumption, hence an element of L^2. The same is true for $f_i(t) = \langle a_i, \overline{\psi}(t) - t\overline{\psi}(1)\rangle,\ t \in [0,1]$, and hence $f_i = \sum_{k=1}^{\infty} c_{k,i}e_k,\ i = 1,\dots,m$, where

$$c_{k,i} = \langle f_i, e_k\rangle_{L^2} = \int_0^1 \langle a_i, \overline{\psi}(t) - t\overline{\psi}(1)\rangle \sqrt{2}\sin(k\pi t)dt,\ \ k \in \mathbb{N},\ \ i = 2,\dots,m,$$

it is

$$\mathcal{B}_i + f_i \overset{d}{=} \sum_{k=1}^{\infty} \Big[\frac{Z_{k,i}}{k\pi} + c_{k,i}\Big] e_k,\ \ i = 2,\dots,m,$$

and

$$T \overset{d}{=} \sum_{i=2}^{m}\sum_{k=1}^{\infty} \Big[\frac{Z_{k,i}}{k\pi} + c_{k,i}\Big]^2.$$

Partial integration and $\cos(k\pi) = (-1)^k$ yields $\int_0^1 t\sin(k\pi t)dt = (-1)^k\frac{1}{k\pi},\ k \in \mathbb{N}$. The first convergence stated follows using the linearity. If $\overline{\psi}$ is differentiable we apply partial integration to $\int_0^1 \langle a_i, \overline{\psi}(t)\rangle\sqrt{2}\sin(k\pi t)dt$ and use $\overline{\psi}(0) = 0$ to obtain the second convergence stated. In the case that $\rho^{(n)}$ approaches independence locally it holds $\overline{\psi}(t) = t\overline{\psi}(1)$, $t \in [0,1]$, and hence $\overline{\psi}'(t) = \overline{\psi}(1),\ t \in [0,1]$. Analogously to above we obtain

$$d_{k,i} = \int_0^1 \langle a_i, \overline{\psi}'(t)\rangle\sqrt{2}\cos(k\pi t)dt = \langle a_i, \overline{\psi}(1)\rangle\sqrt{2}\int_0^1 \cos(k\pi t)dt = 0,\ \ k \in \mathbb{N},\ \ i = 2,\dots,m.$$

It follows that $T \overset{d}{=} \sum_{k=1}^{\infty} \frac{1}{k^2\pi^2}\sum_{i=2}^{m} Z_{k,i}$ and from the fact that the sum of independent squared standard normal distributed random variables is χ-squared distributed the last convergence stated follows. □

Proof of Corollary 3.2. First we consider the case where $\psi^{(n)}$ approaches independence locally. Theorem 3.2 yields that $T_n \xrightarrow{d} T$, where $T = \sum_k W_k/(k^2\pi^2)$ with independent random variables $W_k,\ k \in \mathbb{N}$, each with the same χ-squared distribution with $m-1$ degrees of freedom. The distribution function of the a.s. positive random variable T is continuous and strictly increasing on $(0,\infty)$. For that reason, the $(1-\alpha)$-quantile c is uniquely determined and $P(T_n > c) \to \alpha$ follows. Now we consider local alternatives, where $\psi^{(n)}$ does not approach independence locally. The first convergence in Theorem 3.2 yields that $T_n \xrightarrow{d} T'$, say, where T' is the sum of independent random variables, each with a continuous distribution and support $(0,\infty)$. This yields the convergence of $P(T_n > c)$ to some $\beta \in [0,1]$. Note that the $c_{k,i}$ appearing in the representation of T' are zero for all $k \in \mathbb{N}$, $i = 2,\dots,m$, if and only if $\psi^{(n)}$ approaches independence locally; hence at least one $c_{k,i}$ is different from zero resulting in that T' is stochastically strictly larger then T which yields $\beta > \alpha$ in this case. These arguments prove (i) and (ii). The consistency of the test in (iii) for fixed alternatives follows with standard arguments. □

A.3. Proofs for Section 4

Proof of Lemma 4.1. By symmetry it is sufficient to show that $P(h(Y) < g(Y)) = 0$. It is $\{h(Y) < g(Y)\}$ valid if and only if there exists $y \in \mathbb{Q}$ such that $h(Y) \leqslant y$ and $g(Y) > y$. Union bound yields

$$P(h(Y) < g(Y)) \leqslant \sum_{y\in\mathbb{Q}} P(h(Y) \leqslant y, g(Y) > y).$$

For that reason, it is sufficient to show that

$$P(h(Y) \leqslant y, g(Y) > y) = 0 \text{ for each } y \in \mathbb{R}.$$

Let $y \in \mathbb{R}$ be fixed and let $A = \{h(Y) \leqslant y\}$ and $B = \{g(Y) \leqslant y\}$ such that $P(h(Y) \leqslant y, g(Y) > y) = P(A \cap B^c)$ holds. Because h, g are assumed to be non-decreasing it is $A \subseteq B$ or $B \subseteq A$. If $A \subseteq B$ it follows that $A \cap B^c = \varnothing$ and hence

$$P(h(Y) \leqslant y, g(Y) > y) = P(\varnothing) = 0.$$

If $B \subseteq A$ it follows that

$$P(h(Y) \leqslant y, g(Y) > y) = P(A) - P(B).$$

Since $h(Y)$ and $g(Y)$ are assumed to have the same distribution it is $P(A) = P(B)$, hence

$$P(h(Y) \leqslant y, g(Y) > y) = 0.$$

□

Proof of Lemma 4.2. For $k = 1,2$ let $F_k^{-1}(y) = \inf\{x \in \mathbb{R} : F_k(x) \geqslant y\}, y \in (0,1)$ be the inverse quantile function of F_k which is non-decreasing. Since F_k is continuous and Y_k has distribution F_k, $F_k(Y_k)$ follows the uniform distribution on $[0,1]$. First, we show that (ii) implies (i). Applying F_1^{-1} to the second components of both vectors yields

$$(X_1, F_1^{-1}(F_1(Y_1))) \stackrel{d}{=} (X_2, F_1^{-1}(F_2(Y_2)))$$

Since $F_1(Y_1)$ follows the uniform distribution on $[0,1]$ the random variable $F_1^{-1}(F_1(Y_1))$ has the same distribution as Y_1. The function $h = F_1^{-1} \circ F_1$ is non-decreasing and hence $F_1^{-1}(F_1(Y_1)) = Y_1$ almost surely by Lemma 4.1. This yields

$$(X_1, Y_1) \stackrel{d}{=} (X_2, F_1^{-1}(F_2(Y_2)))$$

and hence (i) follows by application of the non-decreasing function $g = F_1^{-1} \circ F_2$. To show that (i) implies (ii), let g be a non-decreasing function such that $(X_1, Y_1) \stackrel{d}{=} (X_2, g(Y_2))$. Applying F_1 to the second components on both sides yields

$$(X_1, F_1(Y_1)) \stackrel{d}{=} (X_2, F_1(g(Y_2))).$$

Since $F_1(Y_1)$ follows the uniform distribution on $[0,1]$, $F_1(g(Y_2))$ follows the uniform distribution on $[0,1]$ as well. But $F_2(Y_2)$ is also uniformly distributed on $[0,1]$, hence $F_1(g(Y_2))$ and $F_2(Y_2)$ have the same distribution. Since both $F_1 \circ g$ and F_2 are non-decreasing Lemma 4.1 applies and we obtain $F_1(g(Y_2)) = F_2(Y_2)$ almost surely and hence (ii). □

Proof of Corollary 4.1. Since $\rho^1 = \rho^2 = \rho$ we add a zero $0 = \rho - \rho$ and obtain

$$\mathcal{F}_{n_1,n_2} = \sqrt{\frac{n_2}{n_1+n_2}}\mathcal{G}_{1,n_1} - \sqrt{\frac{n_1}{n_1+n_2}}\mathcal{G}_{2,n_2}. \tag{30}$$

Theorem 2.1 yields $\mathcal{G}_{k,n_k} \xrightarrow{d} \mathrm{GP}(\rho)$ for $k = 1,2$. The independence of $\mathcal{G}_{1,n_1}, \mathcal{G}_{2,n_2}$ yields $\mathcal{F}_{n_1,n_2} \xrightarrow{d} \mathcal{G} = \sqrt{1-\gamma}\mathcal{G}' - \sqrt{\gamma}\mathcal{G}''$ where $\mathcal{G}', \mathcal{G}'' \sim \mathrm{GP}(\rho)$ are independent. Basic calculations yield $\mathcal{G} = \sqrt{1-\gamma}\mathcal{G}' - \sqrt{\gamma}\mathcal{G}'' \sim \mathrm{GP}(\rho)$. □

Proof of Lemma 4.3. Since we can adopt the argumentation in the proof of Corollary 4.1, it is sufficient to show that $\mathcal{G}_{1,n_1}$ and $\mathcal{G}_{2,n_2}$ converge in distribution towards $\mathrm{GP}(\rho)$. This is clear for $\mathcal{G}_{1,n_1}$ by Theorem 2.1 because Assumption 1 is fulfilled. For $\mathcal{G}_{2,n_2}$ note that

$$\rho^{(n)}(i,t) = \rho(i,t) + \frac{\sqrt{\frac{n_2}{n_1}} \cdot \psi^{(n)}(i,t)}{\sqrt{n_2}}, \ (i,t) \in \Sigma \times [0,1],$$

and because $\psi^{(n)}$ converges uniformly towards a continuous ψ by assumption, $\sqrt{\frac{n_2}{n_1}} \cdot \psi^{(n)}$ converges uniformly towards the continuous function $\sqrt{\frac{1-\gamma}{\gamma}} \cdot \psi$. Hence Assumption 1 is also satisfied for $\mathcal{G}_{2,n_2}$. □

Proof of Lemma 4.4. Let $\mathcal{C}(\Sigma)$ be the set of all functions $\rho : \Sigma \times [0,1] \to [0,1]$ that are of the form $\rho(i,t) = P(X = i, U \leqslant t)$, $i \in \Sigma, t \in [0,1]$ for some bivariate random vector (X,U) with $P(X \in \Sigma) = 1$ and $P(U \leqslant t) = t$ for all $t \in [0,1]$. For each $w \in \Sigma^n$ we define $N_w : \Sigma \times [0,1] \to [0,1]$ by $N_w(i,t) = \frac{1}{n}\sum_{j=1}^{\lfloor nt \rfloor} 1(w_j = i)$, $i \in \Sigma, t \in [0,1]$. Let

$$M = \{N_w; w \in \cup_{n \geqslant 1}\Sigma^n\} \cup \mathcal{C}(\Sigma).$$

For $\rho, \rho' \in M$ define $d(\rho,\rho') = \sup_{i \in \Sigma, t \in [0,1]} |\rho(i,t) - \rho'(i,t)|$. In Gerstenberg [12] it is shown that (M,d) is a compact metric space. Considering the map $(i,t) \mapsto \widehat{N}_{k,n_k}(i, \lfloor n_k t \rfloor)/n_k$, $(i,t) \in \Sigma \times [0,1]$, $k = 1,2$, as random variable with values in the compact metric space

(M, d), the convergences stated in (ii) hold in probability on any probability space. Moreover we have $\mathcal{G}_{1,n_1} \xrightarrow{d} \mathcal{G}' \sim \mathrm{GP}(\rho_1)$ and $\mathcal{G}_{2,n_2} \xrightarrow{d} \mathcal{G}'' \sim \mathrm{GP}(\rho_2)$. In all, since the limits ρ_1 and ρ_2 are constants and $\mathcal{G}_{1,n_1}, \mathcal{G}_{2,n_2}$ are assumed to be independent it holds the joint convergence in distribution with respect to the product space $M \times M \times D[0,1]^m \times D[0,1]^m$, which is as a product of polish spaces also a polish space, to $(\rho_1, \rho_2, \mathcal{G}', \mathcal{G}'')$ on any probability space. Using the Skohorod embedding theorem yields the existence of a probability space $(\Omega, \mathcal{A}, P)$ where all these random variables are defined and the convergence holds almost surely. In all, (i) and (ii) are proved. (iii) can be obtained easily by attaching uniform randomization to the probability space, i.e., by setting $(\Omega \times [0,1], \mathcal{A} \otimes \mathcal{B}_{[0,1]}, P \otimes \mathrm{unif}[0,1])$ and $U(\omega_1, \omega_2) = \omega_2$. □

Proof of Proposition 4.1. Let U be uniform distributed on $[0,1]$ independent from X_{n_1,n_2} and for each $k = 1, 2$ let $J_k = \lceil n_k U \rceil$. For $(i,t) \in \Sigma \times [0,1]$ define

$$\rho_{k,n_k}(i,t) = P\Big(X_{k,[J_k:n_k]} = i, U \leqslant t \Big| X_{n_1,n_2}\Big).$$

It holds that

$$\rho_{k,n_k}(i,t) = \frac{\widehat{N}_{k,n_k}(i, \lfloor n_k t \rfloor)}{n_k} + P\Big(\ X_{k,[J_k:n_k]} = i,\ \frac{\lfloor n_k t \rfloor}{n_k} < U \leqslant t \ \Big| X_{n_1,n_2}\Big),\ (i,t) \in \Sigma \times [0,1],$$

and hence

$$\sup_{(i,t) \in \Sigma \times [0,1]} \left| \rho_{k,n_k}(i,t) - \frac{\widehat{N}_{k,n_k}(i, \lfloor n_k t \rfloor)}{n_k} \right| \ \leqslant\ \frac{1}{n_k}. \tag{31}$$

By construction, the distribution of the bootstrap sample $\widehat{X}_{n_1,n_2}$ conditioned on X_{n_1,n_2} is determined by

$$\rho^{(n)} = \frac{n_1}{n_1 + n_2} \rho_{1,n_1} + \frac{n_2}{n_1 + n_2} \rho_{2,n_2}.$$

We aim to apply Lemma 4.3 and to show that the sequence $\rho^{(n)}$ fulfills the assumptions there almost surely. Because of (31), $\sup_{(i,t) \in \Sigma \times [0,1]} |\rho^{(n)}(i,t) - (\gamma \rho_1(i,t) + (1 - \gamma)\rho_2(i,t))| \to 0$ almost surely. Moreover, $t \mapsto \gamma \rho_1(i,t) + (1-\gamma)\rho_2(i,t)$ is continuously differentiable for all $i \in \Sigma$. It remains to show that

$$\sqrt{n_1}\Big(\rho^{(n)} - [\gamma \rho_1 + (1-\gamma)\rho_2]\Big) \tag{32}$$

converges almost surely uniformly towards a continuous limit. Plugging in definitions yields

$$
\begin{aligned}
&\sqrt{n_1}\Big(\rho^{(n)} - [\gamma\rho_1 + (1-\gamma)\rho_2]\Big) \\
&\qquad = \sqrt{n_1}\Big(\frac{n_1}{n_1+n_2}\rho_{1,n_1} - \gamma\rho_1\Big) - \sqrt{n_1}\Big(\frac{n_2}{n_1+n_2}\rho_{2,n_2} - (1-\gamma)\rho_2\Big) \\
&\qquad = \frac{n_1}{n_1+n_2}\left[\sqrt{n_1}\big(\rho_{1,n_1} - \rho_1\big) + \rho_1\frac{(1-\gamma)n_1 - \gamma n_2}{\sqrt{n_1}}\right] \\
&\qquad\qquad - \frac{n_2}{n_1+n_2}\sqrt{\frac{n_1}{n_2}}\left[\sqrt{n_2}\big(\rho_{2,n_2} - \rho_2\big) + \rho_2\frac{\gamma n_2 - (1-\gamma)n_1}{\sqrt{n_2}}\right].
\end{aligned}
$$

We consider the terms $\sqrt{n_k}\big(\rho_{k,n_k} - \rho_k\big), k = 1, 2$. From (31) it follows

$$
\sup_{(i,t)\in\Sigma\times[0,1]}\Big|\sqrt{n_k}\big(\rho_{k,n_k}(i,t) - \rho_k(i,t)\big) - \mathcal{G}_{k,n_k}(i,t)\Big| \leqslant \frac{1}{\sqrt{n_k}}.
$$

It converges $\mathcal{G}_{1,n_1} \to \mathcal{G}', \mathcal{G}_{2,n_2} \to \mathcal{G}''$ almost surely and since the upper bound in the latter formula goes to zero, both $\sqrt{n_1}\big(\rho_{1,n_1} - \rho_1\big) \to \mathcal{G}'$ and $\sqrt{n_2}\big(\rho_{2,n_2} - \rho_2\big) \to \mathcal{G}''$ almost surely. The terms $\rho_1\frac{(1-\gamma)n_1 - \gamma n_2}{\sqrt{n_1}}$ and $\rho_2\frac{\gamma n_2 - (1-\gamma)n_1}{\sqrt{n_2}}$ converge by assumption on the integer sequences, see (19). Finally we obtain the almost sure convergence of (32) against a continuous limit function. Hence for almost all ω the sequence $\rho^{(n)}(\omega)$ satisfies the assumption needed for applying Lemma 4.3. □

Proof of Lemma 4.5. Assume we are on the probability space in Lemma 4.4. We write $\mathcal{L}(\cdot)$ for the distribution of a random element and $\mathcal{L}(\cdot|\cdot)$ for conditional distributions. It is

$$
\mathcal{L}\big(\mathcal{F}_{n_1,n_2}, \widehat{\mathcal{F}}^{(1)}_{n_1,n_2}, \widehat{\mathcal{F}}^{(2)}_{n_1,n_2}\big) = E\Big(\mathcal{L}\Big(\mathcal{F}_{n_1,n_2}, \widehat{\mathcal{F}}^{(1)}_{n_1,n_2}, \widehat{\mathcal{F}}^{(2)}_{n_1,n_2}\Big| X_{n_1,n_2}\Big)\Big).
$$

Since $\mathcal{F}_{n_1,n_2}$ is measurable with respect to X_{n_1,n_2} it follows that

$$
\mathcal{L}\big(\mathcal{F}_{n_1,n_2}, \widehat{\mathcal{F}}^{(1)}_{n_1,n_2}, \widehat{\mathcal{F}}^{(2)}_{n_1,n_2}\big) = E\Big(\delta_{\mathcal{F}_{n_1,n_2}} \otimes \mathcal{L}\Big(\widehat{\mathcal{F}}^{(1)}_{n_1,n_2}, \widehat{\mathcal{F}}^{(2)}_{n_1,n_2}\Big| X_{n_1,n_2}\Big)\Big),
$$

where δ_x is the Dirac measure at x. Since $\widehat{\mathcal{F}}^{(1)}_{n_1,n_2}, \widehat{\mathcal{F}}^{(2)}_{n_1,n_2}$ are conditionally independent given X_{n_1,n_2} it is

$$
\mathcal{L}\big(\mathcal{F}_{n_1,n_2}, \widehat{\mathcal{F}}^{(1)}_{n_1,n_2}, \widehat{\mathcal{F}}^{(2)}_{n_1,n_2}\big) = E\Big(\delta_{\mathcal{F}_{n_1,n_2}} \otimes \mathcal{L}\Big(\widehat{\mathcal{F}}^{(1)}_{n_1,n_2}\Big| X_{n_1,n_2}\Big) \otimes \mathcal{L}\Big(\widehat{\mathcal{F}}^{(2)}_{n_1,n_2}\Big| X_{n_1,n_2}\Big)\Big).
$$

Now $\mathcal{F}_{n_1,n_2} \to \mathcal{G}$ almost surely by assumption, see Lemma 4.2, hence $\delta_{\mathcal{F}_{n_1,n_2}} \xrightarrow{d} \delta_{\mathcal{G}}$ almost surely. Moreover by Proposition 4.1 it is $\mathcal{L}\big(\widehat{\mathcal{F}}^{(b)}_{n_1,n_2}\big| X_{n_1,n_2}\big) \xrightarrow{d} \mathrm{GP}(\rho)$ almost surely for $b = 1, 2$. The map $(\mu, \nu) \mapsto \mu \otimes \nu$ is continuous and so

$$
\delta_{\mathcal{F}_{n_1,n_2}} \otimes \mathcal{L}\big(\widehat{\mathcal{F}}^{(1)}_{n_1,n_2}\big| X_{n_1,n_2}\big) \otimes \mathcal{L}\big(\widehat{\mathcal{F}}^{(2)}_{n_1,n_2}\big| X_{n_1,n_2}\big) \xrightarrow{d} \delta_{\mathcal{G}} \otimes \mathrm{GP}(\rho) \otimes \mathrm{GP}(\rho) \quad \text{almost surely.}
$$

In addition, it is

$$\mathcal{L}\big(\mathcal{F}_{n_1,n_2}, \widehat{\mathcal{F}}^{(1)}_{n_1,n_2}, \widehat{\mathcal{F}}^{(2)}_{n_1,n_2}\big) = E\Big(\delta_{\mathcal{F}_{n_1,n_2}} \otimes \mathcal{L}\Big(\widehat{\mathcal{F}}^{(1)}_{n_1,n_2}\Big| X_{n_1,n_2}\Big) \otimes \mathcal{L}\Big(\widehat{\mathcal{F}}^{(2)}_{n_1,n_2}\Big| X_{n_1,n_2}\Big)\Big)$$
$$\xrightarrow{d} E\Big(\delta_{\mathcal{G}} \otimes \mathrm{GP}(\rho) \otimes \mathrm{GP}(\rho)\Big).$$

and we get

$$\mathcal{L}\big(\mathcal{F}_{n_1,n_2}, \widehat{\mathcal{F}}^{(1)}_{n_1,n_2}, \widehat{\mathcal{F}}^{(2)}_{n_1,n_2}\big) \xrightarrow{d} E\big(\delta_{\mathcal{G}}\big) \otimes \mathrm{GP}(\rho) \otimes \mathrm{GP}(\rho) = \mathrm{GP}(\rho) \otimes \mathrm{GP}(\rho) \otimes \mathrm{GP}(\rho)\,.$$

□

Proof of Lemma 4.6. First note that $T = \sup_{i\in\Sigma, t\in[0,1]\cap\mathbb{Q}} |\mathcal{G}(i,t)|$ almost surely since $\mathcal{G} \sim \mathrm{GP}(\rho)$ has continuous paths. Noting that $\mathcal{G}$ is a centered Gaussian process and takes values in a separable space we want to apply Corollary 1.3 together with Remark 4.1 from Gänssler, Molnár and Rost [9]. It remains to show in that $\mathcal{G}$ is not identically to zero, i.e., that the variances $\mathrm{Var}(\mathcal{G}(i,t))$ differ from zero for some $(i,t) \in (i,t) \in \Sigma \times [0,1]$. By assumption it is $\rho(i,1) \in (0,1)$ for all $i \in \Sigma$ and hence

$$\mathrm{Var}(\mathcal{G}(i,1)) = \rho(i,1)(1-\rho(i,1)) > 0.$$

□

References

[1] Bahadur, R. (1966). A Note on Quantiles in Large Samples. *The Annals of Mathematical Statistics* 37: 577–580.

[2] Baringhaus, L., Gaigall, D. (2019). On an asymptotic relative efficiency concept based on expected volumes of confidence regions. *Statistics* 53: 1396–1436.

[3] Bücher, A., Kojadinovic, I. (2019). A Note on conditional versus joint unconditional weak convergence in bootstrap consistency results. *Journal of Theoretical Probability* 32, 1145–1165.

[4] Chatterjee, S. (2020). A new coefficient of correlation. *Journal of the American Statistical Association*, 1–21.

[5] David, H.A., Nagaraja, H.N. (1998). *Concomitants of Order Statistics.* In Handbook of Statistics, Vol. 16, Balakrishnan, N., Rao, C.R. (eds.),

[6] David, H.A., Nagaraja, H.N. (2003). *Order Statistics, Third Edition.* Wiley.

[7] Davydov, Y., Egorov, V. (2001). Functional CLT and LIL for induced order statistics. *Asymptotic Methods in Probability and Statistics with Applications (St. Petersburg, 1998). Stat. Ind. Technol.* 333–349. Boston, MA: Birkhäuser.

[8] Gaigall, D. (2019). Testing marginal homogeneity of a continuous bivariate distribution with possibly incomplete paired data. *Metrika.*

[9] Gänssler, P., Molnár, P., Rost, D. (2007). On Continuity and Strict Increase of the CDF for the Sup-Functional of a Gaussian Process with Applications to Statistics. *Result.Math.* 51 (2007), 51–60.

[10] Genest, C., Remillard, B. (2004). Tests of independence or randomness based on the empirical copula process. *Test* 13: 335–360.

[11] Genest, C., Nešlehová, J. G., Remillard, B., Murohy, O.A. (2019). Testing for independence in arbitrary distributions. *Biometrika* 106: 47–68.

[12] Gerstenberg, J. (2020). General erased-word processes: Product-type filtrations, ergodic laws and Martin boundaries. *Stochastic Processes and their Applications* 130: 3540–3573.

[13] Heller, R., Heller, Y., Kaufman, S., Brill, B., Gorfine, M. (2016). Consistent distribution-free K-sample and independence tests for univariate random variables. *J. Mach. Learn. Res.* 17: 1–54.

[14] Kamps, U. (1995). *A Concept of Generalized Order Statistics.* Springer.

[15] Kiefer, J. (1959). K-Sample Analogues of the Kolmogorov-Smirnov and Cramér-V. Mises Tests. *The Annals of Mathematical Statistics* 30: 420–447.

[16] Pitman, J., Tran, N. (2015). Size-biased permutation of a finite sequence with independent and identically distributed terms. *Bernoulli* 21: 2484-2512.

[17] Puri, M. L., Sen, P. K. (1971). *Nonmarametric Methods in Multivariate Analysis.* Wiley.

[18] Robbins, M.W., Lund, R.B., Gallagher C.M., Lu, Q. (2011). Changepoints in the North Atlantic Tropical Cyclone Record. *Journal of the American Statistical Association* 106:493, 89-99.

[19] Shorack, G.R., Wellner, J.A. (2009). *Empirical processes with applications to statistics.* Siam.

[20] Serfling, R. (1980). *Approximation theorems of mathematical statistics.* Wiley.

[21] Székely, G. J., Rizzo, M. L., Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* 35: 2769–2794.

[22] Ziegler, K. (1997). Functional Central Limit Theorems for Triangular Arrays of Function- Indexed Processes under Uniformly Integrable Entropy Conditions. *Journal of Multivariate Analysis* 62: 233–272.